Université de Montréal

Inférence robuste à la présence des valeurs aberrantes dans les enquêtes

par

Valéry Dongmo Jiongo

Département de mathématiques et de statistique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en statistique

14 janvier 2016

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Inférence robuste à la présence des valeurs aberrantes dans les enquêtes

présentée par

Valéry Dongmo Jiongo

a été évaluée par un jury composé des personnes suivantes :

Mylène Bédard

(président-rapporteur)

David Haziza

(directeur de recherche)

Pierre Duchesne

(codirecteur)

Alejandro Murua

(membre du jury)

Louis-Paul Rivest

(examinateur externe)

Simona Bignami

(représentant du doyen de la FAS)

Thèse acceptée le

09 décembre 2015

SOMMAIRE

Cette thèse comporte trois articles dont un est publié et deux en préparation. Le sujet central de la thèse porte sur le traitement des valeurs aberrantes représentatives dans deux aspects importants des enquêtes que sont : l'estimation des petits domaines et l'imputation en présence de non-réponse partielle.

En ce qui concerne l'estimation des petits domaines, les estimateurs robustes dans le cadre des modèles au niveau des unités ont été étudiés. Sinha & Rao (2009) proposent une version robuste du meilleur prédicteur linéaire sans biais empirique pour la moyenne des petits domaines. Leur estimateur robuste est de type «plug-in», et à la lumière des travaux de Chambers (1986), cet estimateur peut être biaisé dans certaines situations. Chambers *et al.* (2014) proposent un estimateur corrigé du biais. En outre, un estimateur de l'erreur quadratique moyenne a été associé à ces estimateurs ponctuels. Sinha & Rao (2009) proposent une procédure bootstrap paramétrique pour estimer l'erreur quadratique moyenne. Des méthodes analytiques sont proposées dans Chambers *et al.* (2014). Cependant, leur validité théorique n'a pas été établie et leurs performances empiriques ne sont pas pleinement satisfaisantes.

Ici, nous examinons deux nouvelles approches pour obtenir une version robuste du meilleur prédicteur linéaire sans biais empirique : la première est fondée sur les travaux de Chambers (1986), et la deuxième est basée sur le concept de biais conditionnel comme mesure de l'influence d'une unité de la population. Ces deux classes d'estimateurs robustes des petits domaines incluent également un terme de correction pour le biais. Cependant, ils utilisent tous les deux l'information disponible dans tous les domaines contrairement à celui de Chambers *et al.* (2014) qui utilise uniquement l'information disponible dans le domaine d'intérêt. Dans certaines situations, un biais non négligeable est possible pour l'estimateur de Sinha & Rao (2009), alors que les estimateurs proposés exhibent un faible biais pour un choix approprié de la fonction d'influence et de la constante de robustesse. Les simulations Monte Carlo sont effectuées, et les comparaisons sont faites entre les estimateurs proposés et ceux de Sinha & Rao (2009) et Chambers

et al. (2014). Les résultats montrent que les estimateurs de Sinha & Rao (2009) et de Chambers *et al.* (2014) peuvent avoir un biais important, alors que les estimateurs proposés ont une meilleure performance en termes de biais et d'erreur quadratique moyenne.

En outre, nous proposons une nouvelle procédure bootstrap pour l'estimation de l'erreur quadratique moyenne des estimateurs robustes des petits domaines. Contrairement aux procédures existantes, nous montrons formellement la validité asymptotique de la méthode bootstrap proposée. Par ailleurs, la méthode proposée est semi-paramétrique, c'est-à-dire, elle n'est pas assujettie à une hypothèse sur les distributions des erreurs ou des effets aléatoires. Ainsi, elle est particulièrement attrayante et plus largement applicable. Nous examinons les performances de notre procédure bootstrap avec les simulations Monte Carlo. Les résultats montrent que notre procédure performe bien et surtout performe mieux que tous les compétiteurs étudiés. Une application de la méthode proposée est illustrée en analysant les données réelles contenant des valeurs aberrantes de Battese, Harter & Fuller (1988).

S'agissant de l'imputation en présence de non-réponse partielle, certaines formes d'imputation simple ont été étudiées. L'imputation par la régression déterministe entre les classes, qui inclut l'imputation par le ratio et l'imputation par la moyenne sont souvent utilisées dans les enquêtes. Ces méthodes d'imputation peuvent conduire à des estimateurs imputés biaisés si le modèle d'imputation ou le modèle de non-réponse n'est pas correctement spécifié. Des estimateurs doublement robustes ont été développés dans les années récentes. Ces estimateurs sont sans biais si l'un au moins des modèles d'imputation ou de non-réponse est bien spécifié. Cependant, en présence des valeurs aberrantes, les estimateurs imputés doublement robustes peuvent être très instables. En utilisant le concept de biais conditionnel, nous proposons une version robuste aux valeurs aberrantes de l'estimateur doublement robuste. Les résultats des études par simulations montrent que l'estimateur proposé performe bien pour un choix approprié de la constante de robustesse.

Mots clés : Estimateur corrigé pour le biais ; biais conditionnel ; mesure d'influence ; valeurs aberrantes ; inférence basée sur le modèle ; échantillonnage ; estimation des petits domaines ; bootstrap ; modèle linéaire mixte ; robustesse ; imputation.

SUMMARY

This thesis focuses on the treatment of representative outliers in two importants aspects of surveys : small area estimation and imputation for item non-response.

Concerning small area estimation, robust estimators in unit-level models have been studied. Sinha & Rao (2009) proposed estimation procedures designed for small area means, based on robustified maximum likelihood estimators and robust empirical best linear unbiased predictors. Their robust methods for estimating area means are of the plug-in type, and in view of the results of Chambers (1986), the resulting robust estimators may be biased in some situations. Bias-corrected estimators have been proposed by Chambers *et al.* (2014). In addition, these robust small area estimators were associated with the estimation of the Mean Square Error (MSE). Sinha & Rao (2009) proposed a parametric bootstrap procedure based on the robust estimates of the parameters of the underlying linear mixed model to estimate the MSE. Analytical procedures for the estimation of the MSE have been proposed in Chambers *et al.*. (2014). However, their theoretical validity has not been formally established and their empirical performance are not fully satisfactorily.

Here, we investigate two new approaches for the robust version the best empirical unbiased estimator : the first one relies on the work of Chambers (1986), while the second proposal uses the concept of conditional bias as an influence measure to assess the impact of units in the population. These two classes of robust small area estimators also include a correction term for the bias. However, they are both fully bias-corrected, in the sense that the correction term takes into account the potential impact of the other domains on the small area of interest unlike the one of Chambers *et al.* (2014) which focuses only on the domain of interest. Under certain conditions, non-negligible bias is expected for the Sinha-Rao method, while the proposed methods exhibit significant bias reduction, controlled by appropriate choices of the influence function and tuning constants. Monte Carlo simulations are conducted, and comparisons are made between : the new robust

estimators, the Sinha-Rao estimator, and the bias-corrected estimator. Empirical results suggest that the Sinha-Rao method and the bias-adjusted estimator of Chambers *et al* (2014) may exhibit a large bias, while the new procedures offer often better performances in terms of bias and mean squared error.

In addition, we propose a new bootstrap procedure for MSE estimation of robust small area predictors. Unlike existing approaches, we formally prove the asymptotic validity of the proposed bootstrap method. Moreover, the proposed method is semiparametric, i.e., it does not rely on specific distributional assumptions about the errors and random effects of the unit-level model underlying the small-area estimation, thus it is particularly attractive and more widely applicable. We assess the finite sample performance of our bootstrap estimator through Monte Carlo simulations. The results show that our procedure performs satisfactorily well and outperforms existing ones. Application of the proposed method is illustrated by analyzing a well-known outlier-contaminated small county crops area data from North-Central Iowa farms and Landsat satellite images.

Concerning imputation in the presence of item non-response some single imputation methods have been studied. The deterministic regression imputation, which includes the ratio imputation and mean imputation are often used in surveys. These imputation methods may lead to biased imputed estimators if the imputation model or the non-response model is not properly specified. Recently, doubly robust imputed estimators have been developed. However, in the presence of outliers, the doubly robust imputed estimators can be very unstable. Using the concept of conditional bias as a measure of influence (Beaumont, Haziza and Ruiz-Gazen, 2013), we propose an outlier robust version of the doubly robust imputed estimator. Thus this estimator is denoted as a triple robust imputed estimator. The results of simulation studies show that the proposed estimator performs satisfactorily well for an appropriate choice of the tuning constant.

Key words : Corrected-bias estimator ; conditional bias ; influence measure ; outliers ; model-based inference ; survey sampling ; small-area estimation ; bootstrap ; Linear mixed model ; Robustness ; Imputation.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

À ma douce moitié Ornella, mon fils Martin et ma fille Mylène qui s'en vient !
À ma mère et à mon père.

REMERCIEMENTS

Je remercie mon directeur de recherche David Haziza pour avoir orienté cette thèse dans le domaine de la robustesse en échantillonnage, pour sa disponibilité, pour toutes les discussions intéressantes à ce propos et pour les ressources financières qu'il a trouvées. Je remercie également mon co-directeur de recherche Pierre Duchesne pour toute son aide sur les aspects théoriques, pour les ressources financières et pour le soutien continu qu'il m'a accordé pendant la thèse. Je remercie les membres du jury pour leurs commentaires encourageants et constructifs sur la première version du manuscrit. Les révisions apportées ont grandement contribué à améliorer la qualité de la thèse.

Plusieurs personnes ont contribué à ce travail sur le plan scientifique. Je remercie en l'occurence Pierre Nguimkeu qui est co-auteur du deuxième article de cette thèse. Je remercie J.N.K. Rao, Mike Hidiroglou et Jean-François Beaumont qui ont lu les manuscrits de certains des articles de la thèse et donné des commentaires très constructifs. Je remercie Christian Léger qui a formulé des commentaires très constructifs sur le projet de recherche de la thèse. Je tiens également à remercier Nicola Salvati qui a eu la bienveillance de me fournir le code R pour les estimateurs analytiques de l'erreur quadratique moyenne des estimateurs des petits domaines.

Qu'il me soit permis de remercier mon chef Chris Mohl à Statistique Canada qui a mis à ma disposition une machine supplémentaire avec de très grandes capacités afin que je puisse mener à bien mes simulations. Je pense également à mon superviseur Nelson Émond qui m'a beaucoup soutenu et encouragé pour que j'aille jusqu'au bout de la thèse.

Les remerciements vont aussi à tous mes camarades du département de mathématiques et statistique qui m'ont soutenu durant toutes mes études. Je pense par exemple à Romuald Momeya, Thierry Chekouo, Herbert Nkwimi, Blache Paul Akpoué, Christian Nambeu, Zeinab Mashreghi, Béné Fabiola et Annick Nembot.

Je tiens également à remercier les amis qui m'ont accordé une excellente assistance en recherche et encouragé lors des passages à vide. Il s'agit entre autres

de Achille Pegoué, Édouard Tsagué, Guy Tchuenté, Maximilien Kaffo, Vincent Bellinga, Arthur Goussanou, Jacques Éwoudou, Hilaire Lekeufack, Bruno Sontsa, Théophile Bougna, Marc Pandi et Alain Bignom.

Enfin, je dis toute ma gratitude à ma famille et belle famille, vos prières ont contribué à l'aboutissement de ce projet.

Je rends grâce au *Seigneur* de m'avoir donné la force d'accomplir ce projet.

INTRODUCTION

Motivation

Les valeurs aberrantes sont généralement présentes dans les enquêtes économiques. Elles peuvent être des erreurs ou alors des valeurs légitimes prises par certaines unités de la population. Les valeurs légitimes sont de vraies valeurs qui peuvent être uniques ou représenter d'autres unités appartenant à l'univers non échantillonné. Chambers (1986) désigne cette dernière catégorie par le terme «valeurs aberrantes représentatives» par opposition aux «valeurs aberrantes non représentatives» qui sont des erreurs ou des valeurs légitimes uniques. Dans cette thèse, nous traitons uniquement le cas des valeurs aberrantes représentatives. Les valeurs aberrantes peuvent avoir un impact considérable sur les estimations. En pratique, l'influence des valeurs aberrantes peut être réduite en adoptant par exemple un plan stratifié où les valeurs aberrantes vont appartenir à une strate à tirage complet. Cependant, même avec un bon plan d'échantillonnage, il est difficile d'éliminer complètement le problème de valeurs aberrantes. En effet, dans les enquêtes entreprises par exemple les strates sont généralement déterminées selon la géographie, le type d'entreprise et des variables de taille. Si la variable d'intérêt n'est pas liée aux variables de taille, alors cela pourrait conduire à la présence des valeurs aberrantes dans l'échantillon. Il y a deux domaines particuliers de l'échantillonnage où la présence des valeurs aberrantes représentatives dans l'échantillon peut conduire à des inférences erronées à cause d'estimations ponctuelles et d'intervalles de confiance inexacts. Ce sont les petits domaines et l'imputation en présence de non-réponse partielle.

L'estimation des petits domaines

Les domaines sont des sous-ensembles de la population qui peuvent être des aires géographiques, des catégories socio-professionnelles ou des secteurs d'activités économiques. En échantillonnage, un estimateur de domaine est dit direct

s'il utilise uniquement l'information tirée du domaine. Un domaine sera dit petit lorsque les estimateurs traditionnels directs de domaines basés sur le plan et/ou sur le modèle ne sont pas appropriés. C'est généralement le cas lorsque la taille de l'échantillon du domaine est faible ou même égale à zéro. Les méthodes d'estimation indirectes des petits domaines ont reçu une attention considérable ces dernières années en raison d'une demande sans cesse croissante des statistiques fiables pour les petites régions. Ces méthodes d'estimation indirectes vont chercher l'information disponible dans les autres domaines à travers des modèles explicites et des variables auxiliaires, provenant par exemple du recensement ou des données administratives (voir par exemple Rao, 2003,2005).

Deux catégories de modèles sont généralement utilisées : les modèles au niveau des domaines et les modèles au niveau des unités. Les modèles au niveau du domaine lient la variable d'intérêt à des variables auxiliaires au niveau du domaine et les modèles au niveau des unités lient la variable auxiliaire et la variable d'intérêt pour chaque unité de la population. Ces deux classes de modèles sont des cas particuliers du modèle linéaire mixte. Dans cette thèse, seul le cas des modèles au niveau des unités avec une matrice de variance covariance ayant une structure bloc diagonale est traité. Un estimateur bien connu dans ce cadre est le meilleur prédicteur linéaire sans biais empirique (MPLSBE). Il est asymptotiquement sans biais et efficace lorsque le modèle est correctement spécifié, mais est très sensible à la présence des valeurs aberrantes. L'inclusion ou l'exclusion des valeurs aberrantes dans le calcul du meilleur prédicteur linaire sans biais empirique peut avoir un grand impact sur les résultats, comme mentionné par Fellner (1986) et Stahel & Welsh (1997). Sinha & Rao (2009) ont proposé une version robuste du meilleur prédicteur linéaire sans biais. Chambers *et al.* (2014) notent que l'estimateur de Sinha & Rao (2009) est basé sur une version de type plug-in du meilleur prédicteur linéaire sans biais. De ce fait, les travaux de Chambers (1986) suggèrent que cet estimateur peut avoir un biais non négligeable. Pour se protéger contre le biais, Chambers *et al.* (2014) proposent des estimateurs robustes qui corrigent pour le biais et donc peuvent être moins biaisés que celui de Sinha & Rao (2009). Cependant, les estimateurs proposés par Chambers *et al.* (2014) n'obéissent pas à la stratégie suggérée par les travaux de Chambers (1986) selon laquelle le processus de correction du biais devrait correspondre à une stratégie qui consiste à faire un arbitrage entre le biais et la variance. Ainsi l'estimateur de Chambers *et al.* (2014) pourrait avoir un faible biais mais une grande variance.

L'imputation en présence de non-réponse partielle

La non-réponse partielle apparaît pratiquement dans toutes les enquêtes. Elle est généralement traitée par une forme d'imputation qui consiste à remplacer les valeurs manquantes par des «valeurs artificielles». L'on distingue généralement l'imputation déterministe de l'imputation aléatoire. Dans le cas de des méthodes déterministes, la valeur imputée est unique et souvent obtenue au moyen d'un modèle de régression. On parle alors d'imputation par la régression déterministe. Dans le cas de l'imputation aléatoire, la valeur imputée change si le processus d'imputation est repris. S'agissant du cas particulier de l'imputation par la régression aléatoire, la valeur imputée peut être vue comme la somme de la valeur obtenue par la l'imputation déterministe et d'un résidu aléatoire tiré parmi les résidus du modèle de régression.

Il est à noter que si le modèle d'imputation est mal spécifié, alors l'imputation par la régression peut conduire à un biais de non réponse important. Pour se protéger contre ce problème, il est coutume en pratique de construire les classes d'imputation et d'imputer directement à l'intérieur de ces classes à partir d'un modèle de régression. La construction des classes nécessite l'utilisation d'une variable auxiliaire appropriée disponible pour toutes les unités de l'échantillon. Cette variable auxiliaire peut être la même que celle du modèle de régression. L'objectif de la construction des classes est de former les groupes qui sont homogènes par rapport aux probabilités de réponse et/ou par rapport à la variable d'intérêt.

L'imputation par la régression déterministe entre les classes, qui inclut le ratio ou la moyenne entre les classes, est largement utilisée dans les enquêtes, Haziza (2009) constitue une excellente revue de la littérature sur les méthodes d'imputation utilisées dans les enquêtes. Pour toutes ces méthodes, l'utilisation des modèles est incontournable. Cependant, si le modèle est mal spécifié, l'estimateur imputé peut être biaisé si les caractéristiques des répondants sont différentes des caractéristiques des non répondants. Pour améliorer la protection contre le biais de non-réponse, des méthodes qui nécessitent la modélisation de la variable d'intérêt et des probabilités de réponses ont été développées. Ces méthodes peuvent être justifiées au moyen du modèle de non-réponse et du modèle d'imputation. Elles jouissent ainsi d'une double protection en cas d'invalidité de l'un ou l'autre des modèles. Ces méthodes dites doublement robustes, ont connu un intérêt particulier dans la littérature sur l'imputation en présence de non-réponses partielles dans les années récentes, par exemple, Robins *et al.* (1994), Scharfstein *et al.* (1999), Bang & Robins (2005), Cao *et al.* (2009) en population infinie. En échantillonnage, les méthodes doublement robustes ont été étudiées par Kott (1994),

Kim & Park (2006), Haziza & Rao (2006), Haziza & Picard (2012), et Kim & Haziza (2014). En présence de valeurs aberrantes, les estimateurs doublement robustes sont sans biais mais instables. Traiter les valeurs aberrantes revient à faire un arbitrage entre le biais et la variance.

Contribution et structure de la thèse

L'objectif principal de cette thèse est de contribuer à l'amélioration du traitement des valeurs aberrantes représentatives dans le cas des petits domaines et l'imputation en présence de non-réponse partielle. À cet effet, le concept de biais conditionnel sera utilisé. Une extension de ce concept sera effectuée pour les petits domaines, qui sont un cas non standard de l'approche basée sur le modèle car ici les unités ne sont pas indépendantes. En outre, dans le cas de l'imputation en présence de non-réponse partielle, le biais conditionnel va s'avérer être un excellent outil du traitement unifié de la mauvaise spécification du modèle d'imputation ou du modèle de non-réponse, et ainsi que de la présence des valeurs aberrantes dans les enquêtes. De manière explicite, la contribution de cette thèse est structurée autour de quatre chapitres. Le premier chapitre est une revue de la littérature sur la robustesse en échantillonnage. Les notions et concepts utiles à la compréhension de la thèse sont présentées dans ce chapitre. Les trois chapitres suivants discutent de l'élaboration des estimateurs robustes qui possèdent de bonnes propriétés en termes de biais et de variance.

Le deuxième chapitre est basé sur l'article intitulé «Controlling the bias of robust small-area estimators». Cet article est écrit en collaboration avec mon directeur de recherche David Haziza et mon co-directeur de recherche Pierre Duchesne. Cet article a été publié dans le journal *Biometrika* (2013 Vol. 100 pp. 843-858). Il a pour objectif principal de proposer de nouveaux estimateurs robustes pour la moyenne des petits domaines qui possèdent un terme de correction pour le biais. Deux approches y sont étudiées. Dans la première, les arguments de Chambers (1986) sont adaptés au cas du modèle au niveau des unités. Dans la deuxième approche, le concept de biais conditionnel est utilisé. Ces estimateurs ont pour objectifs d'avoir des propriétés désirées en termes de biais et de variance.

Le troisième chapitre est basé sur l'article intitulé «Bootstrapping mean squared errors of robust small-area estimators». Cet article est écrit en collaboration avec Pierre Nguimkeu. L'objectif de cet article est de proposer une nouvelle procédure bootstrap semi-paramétrique pour l'estimation de l'erreur quadratique moyenne des estimateurs robustes des petits domaines. Nous démontrons formellement que lorsque les hypothèses du modèle sont vérifiées, les échantillons bootstrap ont une distribution similaire à celle de l'échantillon originel. De ce fait, la

procédure proposée conduit à des estimateurs convergents de l'erreur quadratique moyenne.

Le quatrième chapitre est basé sur l'article intitulé «Triple robustesse en présence des données imputées dans les enquêtes». Cet article est écrit en collaboration avec mon directeur de recherche David Haziza et mon co-directeur de recherche Pierre Duchesne. L'objectif de cet article est de proposer des estimateurs imputés robustes à la présence des valeurs aberrantes et qui en plus possèdent les propriétés de double robustesse. De ce fait, ces estimateurs sont dits triplement robustes.

Bibliographie

[1] Bang, H. & Robins, J. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, **61**, 962–73.

[2] Cao, W., Tsiatis, A. A. & Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, **96**, 723–34.

[3] Chambers, R. L. (1986). Outliers robust finite population estimation. *Journal of the American Statistical Association*, **81**, 1063–1069.

[4] Chambers, R. L., Chandra, H., Salvati, N. & Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society. Series B*, **76**, 47–69.

[5] Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics*, **28**, 51–60.

[6] Haziza, D. (2009). Imputation and Inference in the presence of missing data. In C. R. Rao and D. Pfefferman (Editors), *Handbook of Statistics, Sample Surveys, Design Methods and Applications*, **29A**, 215–246.

[7] Haziza, D. & Picard, F. (2012). On doubly robust point and variance estimation in the presence of imputed data. *The Canadian Journal of Statistics*, **40**, 259–281.

[8] Haziza, D. et Rao, J. N. K. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodology*, **32**, 53–64.

[9] Jiongo, V., D., Haziza, D. & Duchesne, P. (2013). Controlling the bias of robust small-area estimators. *Biometrika*, **100**, 843–858.

[10] Kim, J.K. & Haziza, D. (2014). Doubly robust inference with missing data. *Statistica Sinica*, **24**, 375–394

[11] Kim, J.K. & Park, H. (2006). Imputation using response probability. *The Canadian Journal of Statistics*, **34**, 171–182.

[12] Kott, P.S. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association*, **89**, 693–696.

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

[14] Rao, J. N. K. (2005). Inferential Issues in Small Area Estimation : Some New Developments. *Statistics in Transition*, **7**, 513–526.

[15] Robins, J.M., Rotnitzky, A. & Zhao, L.P. (1994). Estimation of regression coefficient when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846–66.

[16] Scharfstein, D.O., Rotnitzky, A. & Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion and rejoinder). *Journal of the American Statistical Association*, **94**, 1096–146.

[17] Sinha, S. K. & Rao, J. N. K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, **37**, 381–399.

[18] Stahel, W. A. & Welsh, A. (1997). Approaches to robust estimation in the simplest variance components model. *Journal of Statistical Planning and Inference*, **57**, 295–319.

Chapitre 1

REVUE DE LITTÉRATURE

1.1. Robustesse en échantillonnage

L'inférence peut être erronée en présence de valeurs aberrantes à cause de la très grande variabilité des estimations ponctuelles. Pour se protéger contre ces effets indésirables, plusieurs travaux ont été effectués sur le traitement des valeurs aberrantes à l'étape de l'estimation. Ces travaux peuvent être classés en fonction de l'approche pour l'inférence. En échantillonnage, on distingue deux types d'approches pour l'inférence : l'approche basée sur le plan et celle basée sur le modèle. S'agissant de l'approche basée sur le plan, les propriétés de l'analyse, notamment le biais et la variance, s'obtiennent en considérant tous les échantillons possibles. Par contre, pour l'approche fondée sur le modèle, les propriétés découlent du modèle décrivant la structure de la population.

1.1.1. Approche basée sur le plan

Soit $\mathcal{U} = \{1, 2, \ldots, N\}$ une population finie de taille N et soit y la variable d'intérêt que l'on veut mesurer. En échantillonnage, on s'intéresse généralement à l'estimation du total $Y = \sum_{i=1}^{N} y_i$, ou de la moyenne $\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i$. Soit s un échantillon de taille n sélectionné selon un plan $p(s)$. Dans l'approche basée sur le plan, les valeurs $y_i, i = 1, \ldots, N$ sont supposées fixes. Un estimateur classique du total est l'estimateur de Horvitz-Thompson (1952) donné par :

$$\hat{Y}_{HT} = \sum_{i \in s} d_i y_i, \tag{1.1.1}$$

où $d_i = 1/\pi_i, i = 1, \ldots, N$ sont les poids d'échantillonnage et π_i les probabilités d'inclusion. En présence d'un vecteur de p variables auxiliaires $\mathbf{x}_i, i = 1, \ldots, n$, disponibles pour toutes les unités de l'échantillon et dont le total de la population $\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i$ est connu, on peut construire d'autres estimateurs de Y en postulant

un modèle liant la variable d'intérêt y au vecteur de variables auxiliaires $\mathbf{x}$:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{1.1.2}$$

où les erreurs ϵ_i sont non corrélées de moyenne nulle et de variance $\sigma_i^2 = \sigma^2 c_i$, $c_i > 0$. Les valeurs $c_i = c_i(\mathbf{x}_i)$ permettent de décrire la structure de la variance et sont connues. Un exemple de structure de variance est de choisir une combinaison linéaire des $\mathbf{x}_i$: $c_i = \mathbf{a}^\top \mathbf{x}_i$, où le vecteur $\mathbf{a}$ est connu. Un estimateur alternatif du total fondé sur le plan est l'estimateur par la régression généralisée (GREG) défini par Särndal *et al.* (1992), p. 232 :

$$\hat{Y}_{GREG} = \hat{Y}_{HT} + \left(\mathbf{X} - \hat{\mathbf{X}}_{HT}\right)^\top \hat{\mathbf{B}}, \tag{1.1.3}$$

où

$$\hat{\mathbf{B}} = \left(\sum_{i\in s} d_i \mathbf{x}_i \mathbf{x}_i^\top / c_i\right)^{-1} \sum_{i\in s} d_i \mathbf{x}_i y_i / c_i. \tag{1.1.4}$$

Une autre expression du GREG est donnée par :

$$\hat{Y}_{GREG} = \sum_{i\in s} d_i g_i y_i, \tag{1.1.5}$$

où

$$g_i = 1 + \left(\mathbf{X} - \hat{\mathbf{X}}_{HT}\right)^\top \left(\sum_{i\in s} d_i \mathbf{x}_i \mathbf{x}_i^\top / c_i\right)^{-1} \mathbf{x}_i / c_i.$$

Il est à noter que le GREG est un estimateur de calage (Deville & Särndal 1992). En effet, soit $\hat{T}_y$ un estimateur du total Y et soit $\hat{T}_{\mathbf{x}}$ l'estimateur du total de $\mathbf{X} = \sum_{i\in\mathcal{U}} \mathbf{x}_i$ obtenu en remplaçant y_i par $\mathbf{x}_i$ dans l'expression de $\hat{T}_y$. On dit que $\hat{T}_y$ est un estimateur de calage si $\hat{\mathbf{T}}_{\mathbf{x}} = \mathbf{X}$.

L'estimateur de Horvitz-Thompson et le GREG sont respectivement sans biais et asymptotiquement sans biais. Cependant, en présence de valeurs aberrantes ils sont très variables. Pour réduire l'influence des valeurs aberrantes, trois méthodes ont généralement été utilisées dans la littérature. La première consiste à changer les valeurs par winsorisation ou par «troncation», la deuxième réduit le poids des valeurs aberrantes, et la troisième utilise les techniques d'estimation robustes telles que les M-estimateurs.

Une définition de la winsorisation est fournie par Tukey & McLaughlin (1963). Ils définissent la moyenne winsorisée d'ordre g (symétrique) comme étant la moyenne arithmétique de n valeurs tirées de l'échantillon en remplaçant : (i) chacune des g-plus petites valeurs de y par la plus proches des autres valeurs de l'échantillon dénotée $y_{[g]}$ et (ii) chacune des g-plus grandes valeurs de y par la plus proches des autres valeurs dénotées $y_{[n-g]}$, où $y_{[1]} \leq y_{[2]} \leq \ldots \leq y_{[n]}$ désignent les valeurs ordonnées des y_i, $i = 1, \ldots, n$. Ils définissent également la moyenne

tronquée qui comme étant la moyenne arithmétique des $n - 2g$ valeurs obtenues en supprimant les g-plus petites et les g-plus grandes valeurs de l'échantillon.

Les M-estimateurs sont introduits par Huber (1964) pour le problème de l'estimation robuste d'un paramètre de position. Huber (1973) fait une extension de ses résultats pour un paramètre de position au modèle de régression linéaire. Schématiquement, il propose de construire un M-estimateur $\hat{\mathbf{B}}_R$ solution du problème de minimisation en $\boldsymbol{\beta}$ de l'expression (1.1.6) :

$$\sum_{i=1}^{n} \rho\left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right), \tag{1.1.6}$$

où ρ est une fonction de $\rho : \mathbb{R} \longrightarrow \mathbb{R}^+$. Si ρ a une dérivée $\psi(r) = \frac{\partial \rho(r)}{\partial r}$, alors $\hat{\mathbf{B}}_R$ satisfait le système d'équation (1.1.7)

$$\sum_{i=1}^{n} \psi\left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right) \mathbf{x}_i = \mathbf{0}. \tag{1.1.7}$$

Krasker (1978) examine une classe plus générale d'estimateurs définit par (1.1.8)

$$\sum_{i=1}^{n} \boldsymbol{\phi}\left(y_i, \mathbf{x}_i, \boldsymbol{\beta}\right) = \mathbf{0}, \tag{1.1.8}$$

où $\boldsymbol{\phi}$ est une fonction définie de $\mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p$ vers $\mathbb{R}^p$. Il y a plusieurs propositions du choix de $\boldsymbol{\phi}$ dans la littérature et une classe de fonctions $\boldsymbol{\phi}$ généralement utilisée est de la forme (voir Hampel *et al.* 1986, p. 315) :

$$\boldsymbol{\phi}\left(y_i, \mathbf{x}_i, \boldsymbol{\beta}\right) = w(\mathbf{x}_i)\psi\left\{\left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right) v(\mathbf{x}_i)\right\} \mathbf{x}_i, \tag{1.1.9}$$

où $w(\cdot)$ et $v(\cdot)$ sont des fonctions positives définies de $\mathbb{R}^p$ vers $\mathbb{R}^+$. Huber (1973) utilise $w(\mathbf{x}_i) = 1$ et $v(\mathbf{x}_i) = 1$. Les estimateurs robustes de $\boldsymbol{\beta}$ obtenus en utilisant la classe des fonctions (1.1.9) et les équations (1.1.8) sont appelés M-estimateurs généralisés et dénotés GM-estimateurs.

Lee (1995) discute plus en détails les trois approches pour réduire l'influence des valeurs aberrantes en échantillonnage ; Beaumont & Rivest (2009) fournissent les développements récents.

1.1.1.1. *Réduction du poids des valeurs aberrantes*

Hidiroglou & Srinat (1981) étudient le traitement des valeurs aberrantes dans le cadre d'un plan aléatoire simple sans remise. Ils définissent une sous-population des valeurs aberrantes contenant T valeurs aberrantes. La sous-population des valeurs aberrantes est définie par : $\mathcal{U}_A = \{i, \textit{tel que} \quad y_i > \gamma\}$, où γ est connu. Le reste de la population est défini par : $\mathcal{U}_B = \{i, \textit{tel que} \quad y_i \leq \gamma\}$. Le total de la population $Y = \sum_{i=1}^{n} y_i$ est estimé en utilisant un échantillon s de taille n.

Ils suggèrent trois estimateurs qui réduisent les poids des valeurs aberrantes et ajustent ceux des valeurs non-aberrantes de manière à ce que la somme des poids soit égale à la taille de la population N. Ces trois estimateurs sont donnés par :

$$\hat{Y}_1 = \frac{n}{N} \sum_{i \in s \cap \mathcal{U}_A} d_i y_i + \frac{1 - \frac{t}{N}}{1 - \frac{t}{n}} \sum_{i \in s \cap \mathcal{U}_B} d_i y_i, \tag{1.1.10}$$

$$\hat{Y}_2 = \frac{n+t}{2n} \sum_{i \in s \cap \mathcal{U}_A} d_i y_i + \frac{2n+t}{2n} \sum_{i \in s \cap \mathcal{U}_B} d_i y_i, \tag{1.1.11}$$

$$\hat{Y}_3 = \frac{rn}{N} \sum_{i \in s \cap \mathcal{U}_A} d_i y_i + \frac{1 - \frac{rt}{N}}{1 - \frac{t}{n}} \sum_{i \in s \cap \mathcal{U}_B} d_i y_i, \tag{1.1.12}$$

où $d_i = \frac{N}{n}, \quad i = 1, \ldots, N$, t est le nombre de valeurs aberrantes dans l'échantillon et r est un poids optimal choisi de manière à minimiser l'erreur quadratique moyenne.

L'estimateur (1.1.10) réduit le poids des valeurs aberrantes par un facteur égal à $\frac{n}{N}$. Le facteur de réduction (coefficient du premier terme à droite de l'expression (1.1.11)) est de $\frac{n+t}{2n}$ pour l'estimateur (1.1.11). L'estimateur (1.1.12) généralise l'estimateur (1.1.10) en introduisant un paramètre r qui minimise l'erreur quadratique moyenne (EQM).

1.1.1.2. *Alternative robuste à l'estimateur par le ratio*

L'estimateur par le ratio de la moyenne de la population est donné par : $\hat{Y}_R = N^{-1} \sum_{i \in U} x_i \hat{B} = \bar{X} \hat{B}$, où $\bar{X} = N^{-1} \sum_{i \in U} x_i$ et $\hat{B} = \frac{\sum_{i \in s} y_i}{\sum_{i \in s} x_i}$. Gwet & Rivest (1992) introduisent une version robuste de l'estimateur par le ratio dans le cas d'un plan aléatoire simple sans remise. Il est construit avec les techniques des GM-estimateurs. L'estimateur proposé par Gwet & Rivest s'écrit :

$$\hat{Y}_{GR} = \bar{X} \hat{B}_g, \tag{1.1.13}$$

$\hat{B}_g$ est solution de l'équation suivante :

$$\sum_{i \in s} \frac{\sqrt{x}_i}{h(x_i)} \psi \left\{ \frac{\left(y_i - x_i^\top B\right) h(x_i)}{c\sigma\sqrt{x}_i} \right\} = 0.$$

Ici, c est une constante qui permet de faire un arbitrage entre le biais et la variance et $h(x_i)$ est une fonction qui réduit l'influence des valeurs aberrantes dans les x. L'estimateur $\hat{B}_g$ est robuste aux valeurs aberrantes dans les résidus et dans les variables auxiliaires. De ce fait, l'estimateur (1.1.13) est également robuste aux valeurs aberrantes dans les résidus et les variables auxiliaires. Gwet & Rivest (1992) proposent également un estimateur de l'erreur quadratique moyenne de

(1.1.13). Cet estimateur est donné par :

$$\widehat{EQM}\left(\hat{Y}_{GR}\right) = b_2\left(\hat{Y}_{GR}\right) + v\left(\hat{Y}_{GR}\right), \tag{1.1.14}$$

où $v\left(\hat{Y}_{GR}\right)$ un estimateur de la variance de (1.1.13) et

$$b_2\left(\hat{Y}_{GR}\right) = \max\left\{0, \left(\hat{Y}_{GR} - \hat{Y}_R\right)^2 - \bar{X}^2 \frac{1 - nN^{-1}}{n} \sum_{i \in s} \left(\frac{y_i - \hat{Y}_R x_i}{n^{-1}\sum_{i \in s} x_i} - \widehat{IC}_i\right)^2\right\}$$

est un estimateur tronqué (positif) du carré du biais ; il est à noter que $\widehat{IC}_i$ est donné par

$$\widehat{IC}_i = \frac{\frac{c\sigma\sqrt{x_i}}{h(x_i)}\psi\left\{\frac{\left(y_i - \hat{B}_g x_i\right)h(x_i)}{c\sigma\sqrt{x_i}}\right\}}{n^{-1}\sum_{i \in s} x_i \psi'\left\{\frac{\left(y_i - \hat{B}_g x_i\right)h(x_i)}{c\sigma\sqrt{x_i}}\right\}}.$$

L'estimateur de l'EQM (1.1.14) dépend d'une constante de robustesse c. En statistique classique, le choix de c est dicté par des considérations d'efficacité de $\hat{B}_g$ dans l'hypothèse où les erreurs sont normalement distribuées et les variables auxiliaires suivent une loi du Khi-deux. Gwet & Rivest (1992) utilisent les valeurs de c prédéterminées en statistique classique (voir Hampel *et al.* 1986, p. 333). Ceci peut être problématique si les distributions des données n'obéissent pas aux hypothèses postulées en statistique classique. Une autre approche est proposée par Kokic & Bell (1994) et consiste à déterminer la constante de robustesse qui minimise l'erreur quadratique moyenne de l'estimateur robuste.

1.1.1.3. *Constante de robustesse optimale*

Le problème de la détermination de la constante de robustesse en échantillonnage est présent dans la plupart des approches. Pour des distributions asymétriques, établir des seuils pour identifier les valeurs aberrantes est une procédure arbitraire. Kokic & Bell (1994) étudient les procédures qui intègrent la détermination de la constante de robustesse dans le processus d'estimation. Le plan d'échantillonnage considéré est stratifié aléatoire simple sans remise (PSASSR). La population est partitionnée en H strates de tailles respectives $N_h, h = 1 \ldots, H$. Ils considèrent une variable positive $y_{hi}, h = 1, \ldots, H; i = 1, \ldots, N_h$. Le total de la population est donné par : $Y = \sum_{h=1}^{H} \sum_{i=1}^{N_h} y_{hi}$. L'estimateur winsorisé est donné par :

$$\hat{Y}_{KB} = \sum_{h=1}^{H} \left(\frac{N_h}{n_h}\right) \sum_{i=1}^{n_h} y_{hi}(K_h), \tag{1.1.15}$$

les constantes K_h sont des nombres positifs et $y_{hi}(K_h)$ est donné :

$$y_{hi}(K_h) = \begin{cases} y_{hi}, & \text{si } y_{hi} \leq K_h, \\ f_h y_{hi} + (1 - f_h)K_h, & \text{sinon,} \end{cases}$$

avec $0 \leq f_h \leq 1$. Les valeurs de K_h sont obtenues en minimisant l'EQM de l'estimateur winsorisé du total (1.1.15) donnée par : $E_p E_m (Y - \hat{Y})^2$, où E_m est l'espérance par rapport à la distribution des y_{hi} et E_p est l'espérance par rapport au plan de sondage.

Kokic & Bell (1994) mentionnent qu'il est possible de généraliser leur méthode lorsqu'une variable auxiliaire est disponible pour toutes les unités de la population. Mais, il est à noter que dans un tel cadre, les versions robustes des estimateurs de calage, que nous définirons à la sous-section 1.1.1.4, tels que le GREG perdent leurs propriétés de calage.

1.1.1.4. *Estimateurs de calage robustes*

Duchesne (1999) développe une classe d'estimateurs robustes du total qui possède les propriétés de calage. Cet estimateur est de la forme $\sum_{i \in s} w_i y_i$ où les poids $w_i \in [L, U], 0 < L < U$ sont robustes et calés sur des totaux connus. En outre, ces poids de calage robustes doivent satisfaire aux contraintes désirées par les méthodologistes d'enquêtes. Ce sont par exemple des contraintes positives, car il est difficile d'expliquer aux utilisateurs les poids négatifs présents dans les fichiers fournis. L'estimateur proposé s'écrit :

$$\hat{Y}_g = \sum_{i \in s} d_i g_{iR} u_i y_i, \tag{1.1.16}$$

où

$$g_{iR} = 1 + \left(\sum_{i \in \mathcal{U}} \mathbf{x}_i - \sum_{i \in s} d_i u_i \mathbf{x}_i \right)^\top \left(\sum_{i \in s} d_i h_i^{1-\alpha} u_i \mathbf{x}_i \mathbf{x}_i^\top / (\sigma^2 c_i) \right)^{-1} h_i^{1-\alpha} \mathbf{x}_i^\top / (\sigma^2 c_i),$$

$$u_i = \frac{\psi \left\{ \left(y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_g \right) / \left(\sigma h_i^\alpha \sqrt{c_i} \right) \right\}}{\left(y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_g \right) / \left(\sigma h_i^\alpha \sqrt{c_i} \right)}$$

et $\hat{\mathbf{B}}_g$ est un GM-estimateur solution de

$$\sum_{i \in s} d_i \frac{h_i x_i}{\sqrt{c_i}} \psi \left\{ \frac{y_i - \mathbf{x}_i^\top \mathbf{B}}{\sigma h_i^\alpha \sqrt{c_i}} \right\} = 0.$$

Ici, σ et $c_i, i \in s$ sont connus ; et la fonction h est choisie de manière à réduire l'influence des valeurs aberrantes dans les variables auxiliaires. La constante α prend les valeurs 0 ou 1. La valeur de $\alpha = 0$ conduit au choix de Mallows qui réduit l'impact des valeurs aberrantes au niveau des variables auxiliaires quel que

soit la valeur des résidus. La valeur $\alpha = 1$ permet quant à elle de réduire l'impact des valeurs aberrantes au niveau des auxiliaires seulement lorsque les résidus correspondants sont larges, (voir Hampel *et al.* 1986, page 322). Une écriture alternative de $\hat{Y}_g$ dans (1.1.16) est donnée par :

$$\hat{Y}_g = \sum_{i\in\mathcal{U}} \mathbf{x}_i^\top \hat{\mathbf{B}}_g + \sum_{i\in s} d_i u_i (y_i - \mathbf{x}_i^\top \hat{\mathbf{B}}_g).$$

1.1.2. Approche basée sur le modèle

Sous l'approche basée sur le modèle, l'hypothèse courante stipule que la variable d'intérêt $y_i, i = 1, \ldots, N$ est une réalisation indépendante d'un échantillon de taille N d'une super-population ξ_0. Le vecteur $(y_1, \ldots, y_N)^\top$ n'est pas fixé. Par contre, l'échantillon $\mathbf{s} = (I_1(s), \ldots, I_N(s))^\top$ est fixé. Les propriétés de l'analyse découlent de la structure du modèle décrivant la population. Le biais et la variance sont calculés sous le modèle postulé et conditionnellement à l'échantillon obtenu.

1.1.2.1. *Version robuste du meilleur prédicteur linéaire sans biais*

Le modèle de régression linéaire (1.1.2) est généralement postulé. Sous ce modèle, le meilleur estimateur linéaire sans biais du total est donné par Royall (1970) :

$$\hat{Y}_{MPLSB} = \sum_{i\in s} y_i + \sum_{i\in\mathcal{U}/s} \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}, \tag{1.1.17}$$

où $\hat{\boldsymbol{\beta}} = \left\{\sum_{i\in s} \mathbf{x}_i \mathbf{x}_i^\top / c_i\right\}^{-1} \left\{\sum_{i\in s} \mathbf{x}_i y_i / c_i\right\}$.

Chambers (1986) note que $\hat{\boldsymbol{\beta}}$ est sensible aux valeurs aberrantes. Pour réduire leur influence, une option est de remplacer $\hat{\boldsymbol{\beta}}$ par un estimateur robuste $\hat{\beta}_R$ de $\boldsymbol{\beta}$ qui peut être déterminé en utilisant les techniques de M-estimateur. Une version robuste, $\hat{Y}_R = \sum_{i\in s} y_i + \sum_{i\in\mathcal{U}/s} \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_R$, du meilleur prédicteur linéaire sans biais du total est alors obtenue. Mais, bien que cet estimateur stabilise la variance, il ne résout pas le problème d'estimation robuste du total de la population finie car il peut être considérablement biaisé. Chambers (1986) propose alors une alternative qui permet de faire un arbitrage entre le biais et la variance de manière à avoir les propriétés asymptotiques désirées. Cette alternative est motivée par la décomposition suivante :

$$\hat{Y}_{MPLSB} = \sum_{i\in s} y_i + \sum_{i\in\mathcal{U}/s} \mathbf{x}_i^\top \boldsymbol{\theta} + \sum_{i\in s} u_i \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\theta}}{\sigma\sqrt{c_i}}, \tag{1.1.18}$$

où $\boldsymbol{\theta}$ est arbitraire et $u_i = \left(\sum_{i\in\mathcal{U}/s} \mathbf{x}_i^\top\right) \left(\sum_{i\in s} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\sigma^2 c_i}\right)^{-1} \frac{\mathbf{x}_i}{\sigma\sqrt{c_i}}$. La version robuste du meilleur estimateur linéaire sans biais proposé par Chambers est obtenue en

réduisant les résidus importants de (1.1.18) et en prenant pour $\boldsymbol{\theta}$ un estimateur robuste $\hat{\boldsymbol{\beta}}_R$. Cet estimateur $\hat{Y}_C$, est donné par :

$$\hat{Y}_C = \sum_{i\in s} y_i + \sum_{i\in\mathcal{U}/s} x_i^\top \hat{\beta}_R + \sum_{i\in s} u_i \psi\left(\frac{y_i - x_i^\top \hat{\beta}_R}{\sigma\sqrt{c_i}}\right). \tag{1.1.19}$$

Un choix approprié de la fonction ψ permet de faire une discussion entre le biais et la variance. Les cas extrêmes sont $\psi(t) = t$ et $\psi(t) = 0$. Chambers (1986) montre empiriquement que l'estimateur proposé performe mieux que les compétiteurs.

1.1.3. Approche unifiée de la robustesse en échantillonnage

Le traitement des valeurs aberrantes représentatives au niveau de l'estimation nécessite une mesure de leur influence sur les estimateurs. Plusieurs études ont été menées pour essayer d'adapter la fonction d'influence en statistique classique à l'échantillonnage. Par exemple Gwet & Rivest (1992) utilisent une approximation basée sur la fonction d'influence. Mais Beaumont *et al.* (2013) relèvent que cette approximation est utile pour l'estimation de la variance mais elle conduit à une mesure d'influence qui ne tient pas compte du plan d'échantillonnage. Beaumont *et al.* (2013) proposent alors une approche basée sur le concept de biais conditionnel pour identifier et traiter les valeurs aberrantes au niveau de l'estimation.

Le biais conditionnel est introduit par Muñoz-Pichardo *et al.* (1995) pour mesurer l'influence dans le cadre d'un modèle linéaire général. Leurs travaux reposent sur le lemme de décomposition de Efron & Stein (1981) de l'erreur de prédiction en plusieurs termes dont le premier est la somme des biais conditionnels définis par Muñoz-Picardo *et al.* (1995). Moreno-Rebello *et al.* (1999, 2002) ajustent la définition du biais conditionnel dans le but d'obtenir une mesure d'influence en échantillonnage. Beaumont *et al.* (2013) montrent que le biais conditionnel est approximativement proportionnel à la fonction d'influence en statistique classique et constitue de ce fait une mesure d'influence. Ils montrent également que les autres termes de la décomposition de Efron & Stein (1981) s'annulent pour certaines statistiques. Cette propriété sera utile pour dériver leur estimateur robuste.

S'agissant de l'inférence basée sur le plan, le biais conditionnel de l'unité i de l'estimateur de Horvitz-Thompson est par exemple donné par :

$$B_i^{HT} = E_p\left\{\hat{Y}_\pi - Y | I_i(s)\right\} = \begin{cases} \sum_{j\in\mathcal{U}} \left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right) y_j, & \text{si } i \in s, \\ -\frac{1}{d_i - 1} \sum_{j\in\mathcal{U}} \left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right) y_j, & \text{si } i \in \mathcal{U}/s, \end{cases} \tag{1.1.20}$$

où $I_i(s) = 1$ si $i \in s$ et $I_i(s) = 0$ sinon. Partant de la décomposition de l'erreur d'estimation comme une somme de biais conditionnels, l'estimateur robuste du

biais conditionnel s'écrit :

$$\hat{Y}_{RHT} = \hat{Y}_{HT} - \sum_{i \in s} \hat{B}_i^{RHT} + \sum_{i \in s} \psi_c \left(\hat{B}_i^{RHT} \right), \tag{1.1.21}$$

où $\hat{B}_i^{RHT}$ est un estimateur robuste du biais conditionnel (1.1.20) pour les unités de l'échantillon, c est la constante de robustesse, et ψ_c est une fonction de type Huber $\psi_c(x) = max\{-c, min(x, c)\}$. Pour un choix approprié de la fonction ψ_c, l'estimateur (1.1.21) se réduit à celui de Kokic & Bell (1994) dans le cas stratifié aléatoire simple.

Dans l'approche basée sur le modèle, le biais conditionnel de l'unité i est donné par :

$$B_i(y_i, \boldsymbol{\beta}) = E_p \left\{ \hat{Y}_{MPLSB} - Y | y_i, s \right\} = \begin{cases} (w_i - 1) \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right), & \text{si } i \in s, \\ - \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right), & \text{si } i \in \mathcal{U}/s, \end{cases} \tag{1.1.22}$$

où

$$w_i = 1 + \mathbf{x}_i^\top \left(\sum_{i \in s} \mathbf{x}_i \mathbf{x}_i^\top / c_i \right)^{-1} \left(\sum_{i \in \mathcal{U}} \mathbf{x}_i \right) / c_i.$$

De même que pour l'estimation basée sur le plan, l'estimateur robuste du biais conditionnel basé sur le modèle est donné par :

$$\hat{Y}_R = \hat{Y}_{MPLSB} - \sum_{i \in s} \hat{B}_i^R + \sum_{i \in s} \psi_c \left(\hat{B}_i^R \right), \tag{1.1.23}$$

où $\hat{B}_i^R$ est un estimateur robuste du biais conditionnel (1.1.22) pour les unités de l'échantillon. L'estimateur (1.1.23) s'écrit encore : $\hat{Y}_R = \sum_{i \in s} y_i + \sum_{i \in \mathcal{U}/s} \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_R + \sum_{i \in s} \psi_c \left\{ (w_i - 1) \left(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_R \right) \right\}$. Cette dernière expression montre que l'estimateur basé sur le concept de biais conditionnel est similaire à celui de Chambers (1986) avec comme différence le terme $w_i - 1$ qui se situe à l'intérieur de la fonction ψ_c.

Par ailleurs, Beaumont *et al.* (2013) proposent une méthode pour déterminer la constante de robustesse c. Elle consiste à choisir celle qui minimise le maximum du biais conditionnel en valeur absolue de l'estimateur robuste. Ils montrent en outre que ce choix de la constante de robustesse conduit à une expression réduite des estimateurs (1.1.21) et (1.1.23) :

$$\hat{Y}_R = \hat{Y} - \frac{1}{2} \left(\hat{B}_i^{Min} + \hat{B}_i^{Max} \right). \tag{1.1.24}$$

Le biais conditionnel apparaît donc comme un outil unifié du traitement des valeurs aberrantes car il s'applique aussi bien pour l'approche basée sur le plan que celle basée sur le modèle. De ce fait, il trouve une application dans deux domaines

particuliers de l'échantillonnage où les modèles sont abondamment utilisés : il s'agit des petits domaines et de l'imputation en présence de non-réponse partielle.

1.2. L'estimation des petits domaines

En pratique, les enquêtes ne sont pas seulement utilisées pour fournir des estimations pour le total de la population d'intérêt $\mathcal{U}$. Elles sont également utilisées pour générer des estimations des sous-populations appelées domaines. Un exemple de domaine géographique inclut les provinces, les municipalités, les divisions de recensement, les subdivisions de recensements, et les régions agricoles de recensement. Les domaines peuvent également être des caractéristiques socio-démographiques telles que les classes d'âge, le sexe, la catégorie socio-professionnelle. D'autres exemples de domaines concernent les regroupement des entreprises suivant le groupe d'industrie en utilisant par exemple des classifications telles que le Système de Classification des Industries de l'Amérique du Nord (SCIAN), voir SCIAN Canada (2012) pour les détails.

La population peut alors être partitionnée en k domaines $\mathcal{U} = \bigcup_{i=1}^{k} \mathcal{U}_i$ de taille N_i. Soit s un échantillon tiré de $\mathcal{U}$, alors $s = \cup_{i=1}^{k} s_i$ où $s_i = s \cap \mathcal{U}_i$. Lorsque la taille n_i du domaine $\mathcal{U}$ est faible, le domaine est dit petit. Soit y la variable d'intérêt. Le total du domaine dénoté Y_i est donné par $Y_i = \sum_{j \in \mathcal{U}_i} y_{ij}$. On distingue trois catégories d'estimateurs du total du domaine Y_i : les estimateurs directs, les estimateurs indirects avec modélisation implicite, et les estimateurs indirects avec modélisation explicite. Les deux premiers estimateurs sont construits en utilisant l'approche traditionnelle en échantillonnage qui préconise l'inférence basée sur le plan. Quant au troisième estimateur, il utilise l'approche de l'inférence basée sur le modèle.

1.2.1. Méthodes traditionnelles

1.2.1.1. *Estimations directes des petits domaines*

Les estimateurs directs de domaines utilisent uniquement les données de l'échantillon s_i propre au domaine. L'estimateur classique direct élémentaire du total du domaine est donné par :

$$\hat{Y}_{iHT} = \sum_{j \in s_i} d_j y_{ij}, \tag{1.2.1}$$

L'estimateur (1.2.1) est sans biais et possède la propriété d'additivité, c'est-à-dire : $\hat{Y}_{HT} = \sum_{i=1}^{k} \hat{Y}_{iHT}$. La variance de (1.2.1) est donnée par :

$$V\left(\hat{Y}_{iHT}\right) = \sum_{j\in\mathcal{U}_i}\sum_{l\in\mathcal{U}_i}\left(\pi_{jl} - \pi_j\pi_l\right)\frac{y_{ij}}{\pi_j}\frac{y_{il}}{\pi_l}. \tag{1.2.2}$$

Dans le cas particulier d'un plan aléatoire simple sans remise (PASSR), la variance (1.2.2) s'écrit (voir Särndal *et al.* 1992, p. 392) :

$$V\left(\hat{Y}_{iHT}\right) = N^2\left(1 - \frac{n}{N}\right)\frac{P_i}{E_p(n_i)}\frac{(N_i-1)S_{iy}^2 + N_i(1-P_i)\bar{Y}_i^2}{N-1},$$

où $S_{iy}^2 = \sum_{j\in\mathcal{U}_i}\left(y_{ij} - \bar{Y}_i\right)^2/(N_i - 1)$, $\bar{Y}_i = Y_i/N_i$, $P_i = N_i/N$ et $E_p(n_i)$ est l'espérance par rapport au plan de la taille du domaine : $E_p(n_i) = nP_i$. La variabilité de $\hat{Y}_{iHT}$ est très grande si l'espérance de la taille du domaine, $E_p(n_i)$, est faible. En présence de variables auxiliaires $\mathbf{x}_{ij}, j \in \mathcal{U}_i$, une option est d'utiliser l'estimateur direct par la régression généralisée :

$$\hat{Y}_{iGREG} = \hat{Y}_{iHT} + \left(\mathbf{X}_i - \hat{\mathbf{X}}_{iHT}\right)^\top \hat{\mathbf{B}}_i, \tag{1.2.3}$$

où $\mathbf{X}_i = \sum_{j\in\mathcal{U}_i}\mathbf{x}_{ij}$ et

$$\hat{\mathbf{B}}_i = \left(\sum_{j\in s_i} d_j\mathbf{x}_{ij}\mathbf{x}_{ij}^\top/c_j\right)^{-1}\sum_{j\in s_i} d_j\mathbf{x}_{ij}y_{ij}/c_j.$$

L'expression (1.2.3) s'écrit encore :

$$\hat{Y}_{iGREG} = \sum_{j\in s_i} d_j g_{ij} y_{ij}, \tag{1.2.4}$$

où

$$g_{ij} = 1 + \left(\mathbf{X}_i - \hat{\mathbf{X}}_{iHT}\right)^\top\left(\sum_{j\in s_i} d_i\mathbf{x}_{ij}\mathbf{x}_{ij}^\top/c_j\right)^{-1}\mathbf{x}_{ij}/c_j.$$

Il est à noter que la variance du GREG direct (1.2.3) peut être plus faible que celle de (1.2.1) si les résidus $y_{ij} - \mathbf{x}_{ij}^\top\hat{\mathbf{B}}_i$ ne sont pas très dispersés. Seulement l'estimateur GREG direct (1.2.3) perd sa propriété d'additivité, c'est-à-dire $\hat{Y}_{GREG} \neq \sum_{i=1}^{k}\hat{Y}_{iGREG}$; de plus, il n'est pas approximativement sans biais sous le plan si l'espérance de la taille de l'échantillon du domaine est faible, voir Rao (2003) p. 18. Ceci conduit à considérer comme alternative les estimateurs indirects avec modélisation implicite.

1.2.1.2. *Estimatieurs indirects avec modélisation implicite*

Les estimateurs indirects avec modélisation implicites s'appuient sur un modèle en y reliant le domaine d'intérêt $\mathcal{U}_i$ au reste de la population $\mathcal{U}$. L'objectif de

ces méthodes est de prendre avantage de l'information disponible hors du domaine pour gagner en précision. Les estimateurs indirects comprennent entre autres les estimateurs synthétiques et les estimateurs composites.

Un exemple d'estimateur synthétique généralement utilisé est donné par :

$$\hat{Y}_{iSYN} = \mathbf{X}_i^\top \hat{\mathbf{B}}, \tag{1.2.5}$$

où $\hat{\mathbf{B}}$ est donné par (1.1.4). La variance sous le plan de l'estimateur (1.2.5) sera en général plus petite que la variance sous le plan de l'estimateur direct (1.2.3) car la précision de (1.2.5) ne dépend pas de la taille de l'échantillon s_i du domaine d'intérêt mais plus tôt de celle de l'échantillon total s. En outre, les estimateurs (1.2.5), $i = 1 \ldots, k$, s'adidtionnent à l'estimateur GREG (1.1.3) lorsque $Var(y_{ij}) = c_{ij} = \boldsymbol{\lambda}^\top \mathbf{x}_{ij}$, ou $\boldsymbol{\lambda}$ est un vecteur colonne constant, (voir Rao 2003. p. 47). Cependant, il est biaisé et son biais est approximativement égal à $\mathbf{X}_i^\top \mathbf{B} - Y_i$, où

$$\mathbf{B} = \left(\sum_{i \in \mathcal{U}} \mathbf{x}_i \mathbf{x}_i^\top / c_i \right)^{-1} \sum_{i \in \mathcal{U}} d_i \mathbf{x}_i y_i / c_i$$

est le coefficient de régression au niveau de la population.

Une manière naturelle d'arbitrer entre le biais potentiel de l'estimateur synthétique, dénoté $\hat{Y}_{iSYN}$, et l'estimateur direct $\hat{Y}_{iGREG}$, est de considérer une moyenne pondérée de $\hat{Y}_{iGREG}$ et $\hat{Y}_{iSYN}$. Un tel estimateur, dénoté $\hat{Y}_{iCOMP}$, s'écrit :

$$\hat{Y}_{iCOMP} = \phi_i \hat{Y}_{iGREG} + (1 - \phi_i) \hat{Y}_{iSYN}, \tag{1.2.6}$$

où $0 \leq \phi_i \leq 1$. L'estimateur (1.2.6) est complètement défini lorsqu'une valeur du poids ϕ_i est donnée. Une option est de considérer le choix optimal, ϕ_i^*, qui est déterminé en minimisant l'EQM de (1.2.6). Rao (2003) fournit une revue détaillée sur les différents estimateurs possibles de ϕ_i^*. Une alternative est généralement de déterminer le poids ϕ_i en fonction de la taille n_i de l'échantillon du domaine d'intérêt $\mathcal{U}_i$. De manière générale, l'on donne plus de poids à l'estimateur direct à mesure que la taille de l'échantillon n_i du domaine d'intérêt $\mathcal{U}_i$ augmente. L'estimateur composite (1.2.6) sera généralement moins variable que l'estimateur direct et aura un biais plus faible que celui du synthétique. Cependant, il n'est pas possible d'éliminer complètement le biais.

Nous venons de voir que les estimateurs traditionnels basés sur le plan de sondage présentent certains désavantages. Par exemple, les estimateurs directs classiques sont généralement sans biais mais ont une grande variance. À l'inverse, les estimateurs indirects avec modélisation implicite sont moins variables mais biaisés. Un autre désavantage est qu'il arrive parfois des demandes pour des estimations dont la taille de l'échantillon du domaine est égale à zéro, Pfeffermann

(2013). Dans ce cas spécifique, il est impossible de faire de l'inférence basée sur le plan.

1.2.2. Méthodes indirectes basées sur des modèles explicites

On distingue deux types de modèles : les modèles au niveau des domaines et les modèles au niveau des unités. Les modèles au niveau des domaines relient les estimateurs directs des petits domaines aux variables auxiliaires au niveau des domaines. Ils sont utilisés pour la première fois par Fay-Herriot (1979). Les modèles au niveau des unités relient la variable d'intérêt et les variables auxiliaires au niveau des unités. Ils sont introduits dans la littérature par Battese *et al.* (1988). Nous nous consacrerons uniquement aux travaux effectués dans le cadre des modèles au niveau des unités.

Une hypothèse nécessaire dans le cas des modèles au niveau des unités est que le plan d'échantillonnage est non-informatif ; c'est-à-dire que le modèle est valide aussi bien au niveau de l'échantillon que de la population. Autrement dit, le biais de sélection est absent. Nous faisons également l'hypothèse qu'il existe au niveau de la population un vecteur de p variables auxiliaires $\mathbf{x}_{ij}$. La variable d'intérêt y est liée aux variables auxiliaires par le modèle suivant :

$$y_{ij} = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + v_i + e_{ij} \quad (i = 1, \ldots, k \quad \text{et} \quad j = 1, \ldots, N_i), \tag{1.2.7}$$

où v_i dénote une variable aléatoire associée au petit domaine $\mathcal{U}_i$ et e_{ij} désigne le terme d'erreur. Les erreurs e_{ij} et les effets aléatoires v_i sont supposés indépendants et de loi normales $\mathcal{N}(0, \sigma_e^2)$ et $\mathcal{N}(0, \sigma_v^2)$ respectivement. Pour les unités de l'échantillon, le modèle (1.2.7) s'écrit encore :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \tag{1.2.8}$$

où $\mathbf{y}$ est un vecteur de taille $n = \sum_{i=1}^{k} n_i$; $\mathbf{X}$ et $\mathbf{Z}$ sont des matrices connues de plein rang et de taille respectives $n \times p$ et $n \times k$; et $\mathbf{v}$ et $\mathbf{e}$ sont indépendants de moyenne $\mathbf{0}$ et de matrice de variance covariance respectives $\mathbf{G}$ et $\mathbf{R}$ dépendant d'un paramètre de variance $\boldsymbol{\delta} = (\sigma_e^2, \sigma_v^2)$. Le paramètre $\boldsymbol{\delta}$ est également supposé appartenir à un sous-ensemble ouvert de $\mathbb{R}^2$ tel que la matrice de variance covariance $\mathbf{V} = \mathbf{V}(\boldsymbol{\delta}) = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^{\top}$ soit inversible pour tout $\boldsymbol{\delta}$ appartenant à ce sous ensemble.

L'on s'intéresse par exemple à l'estimation de la moyenne $\bar{Y}_i = N_i^{-1} \sum_{j \in \mathcal{U}_i} y_{ij}$. Lorsque le paramètre de variance $\boldsymbol{\delta}$ est connu, le meilleur prédicteur linéaire sans biais de Y_i (MPLSB) s'écrit :

$$\tilde{\bar{Y}}_{iMPLSB} = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in \mathcal{U}_i / s_i} \mathbf{x}_{ij}^{\top}\tilde{\boldsymbol{\beta}} + (N_i - n_i)\tilde{v}_i \right\}, \tag{1.2.9}$$

où

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}, \tag{1.2.10}$$

est le meilleur estimateur linéaire sans biais (MELSB) de $\boldsymbol{\beta}$ et

$$\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\boldsymbol{\delta}) = \mathbf{G}\mathbf{Z}^\top \mathbf{V}^{-1} \left(\mathbf{y} - \mathbf{X}^\top \tilde{\boldsymbol{\beta}}\right), \tag{1.2.11}$$

$\tilde{\mathbf{v}}_i$ est le MPLSB de $\mathbf{v}_i$.

Le MPLSB (1.2.9) dépend du vecteur inconnu des paramètres de variance $\boldsymbol{\delta}$. En remplaçant $\boldsymbol{\delta}$ par un estimateur $\hat{\boldsymbol{\delta}}$, on obtient le meilleur prédicteur linéaire sans biais empirique (MPLSBE) :

$$\hat{\bar{Y}}_{iMPLSB} = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in \mathcal{U}_i / s_i} \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + (N_i - n_i) \hat{v}_i \right\}. \tag{1.2.12}$$

Les procédures standard d'estimation de $\boldsymbol{\delta}$ requièrent la normalité de $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$ et de $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. De ce fait, les équations normales pour l'estimation de $\boldsymbol{\beta}$ et $\boldsymbol{\delta}$ par le maximum de vraisemblance (MV) sont respectivement données par :

$$\mathbf{M}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}, \tag{1.2.13}$$

$$\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\delta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{(j)} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + tr\left(\mathbf{V}^{-1} \mathbf{V}_{(j)}\right) = \mathbf{0}, \tag{1.2.14}$$

où $\mathbf{V}_{(j)} = \partial \mathbf{V} / \partial \delta_j$ et $\mathbf{V}^{(j)} = \partial \mathbf{V}^{-1} / \partial \delta_j = -\mathbf{V}^{-1} \mathbf{V}_{(j)} \mathbf{V}^{(j)}$, $j = 1, \ldots, q$. Les estimateurs du maximum de vraisemblance sont obtenus numériquement en utilisant l'algorithme du score. L'évaluation à la $a^{\text{ième}}$ itération est donnée par :

$$\boldsymbol{\beta}^{(a)} \left\{\boldsymbol{\delta}^{(a)}\right\} = \left[\mathbf{X}^\top \mathbf{V}^{-1} \left\{\boldsymbol{\delta}^{(a)}\right\} \mathbf{X}\right]^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \left\{\boldsymbol{\delta}^{(a)}\right\} \mathbf{y}, \tag{1.2.15}$$

$$\boldsymbol{\delta}^{(a+1)} = \left[\mathcal{I}\left\{\boldsymbol{\delta}^{(a)}\right\}\right]^{-1} \mathbf{S}\left\{\boldsymbol{\beta}^{(a)}, \boldsymbol{\delta}^{(a)}\right\}, \tag{1.2.16}$$

où $\mathcal{I}(\boldsymbol{\delta}) = E_m\left[-\frac{\partial}{\partial \boldsymbol{\delta}} \{\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\delta})\}\right]$ dont le terme d'ordre (l, m) est donné par $\mathcal{I}_{lm}(\boldsymbol{\delta}) = \text{tr}\left\{\mathbf{V}^{-1} \mathbf{V}_{(l)} \mathbf{V}^{-1} \mathbf{V}_{(m)}\right\}$. Voir Harville (1977) et Rao (2003) pour plus de détails sur l'estimation des paramètres du modèle et les algorithmes associés.

1.2.3. Estimation de l'erreur quadratique moyenne

1.2.3.1. *Estimateurs analytiques de EQM du MPLSB*

L'erreur quadratique moyenne du meilleur prédicteur linéaire sans biais (1.2.12) est donnée par :

$$EQM\left(\hat{\bar{Y}}_{iMPLSB}\right) = E_m\left(\hat{\bar{Y}}_{iMPLSB} - \bar{Y}_i\right)^2.$$

Kackar et Harville (1984) proposent une approximation d'ordre $O_p(k^{-1})$ de l'EQM de toute combinaison linéaire $\mu = \mathbf{l}^\top \boldsymbol{\beta} + \mathbf{m}^\top \mathbf{v}$, où $\mathbf{l}$ et $\mathbf{m}$ sont connus et constants.

Prasad et Rao (1990) font une extension de ces travaux et proposent une approximation d'ordre $o_p(k^{-1})$ l'erreur quadratique moyenne de $\mu_i = \bar{\mathbf{X}}_i^\top \boldsymbol{\beta} + v_i$. L'approximation de Prasad et Rao (1990) est donnée par :

$$EQM\left(\hat{\bar{Y}}_{iMPLSB}\right) \approx g_{1i}(\boldsymbol{\delta}) + g_{2i}(\boldsymbol{\delta}) + g_{3i}(\boldsymbol{\delta}),$$

où

$$g_{1i}(\boldsymbol{\delta}) = (1-\rho_i)\,\sigma_v^2,$$
$$g_{2i}(\boldsymbol{\delta}) = \left(\bar{\mathbf{X}}_i - \rho_i \bar{\mathbf{x}}_i\right)^\top \left(\bar{\mathbf{X}}^\top \mathbf{V}^{-1} \bar{\mathbf{X}}\right)^{-1} \left(\bar{\mathbf{X}}_i - \rho_i \bar{\mathbf{x}}_i\right),$$
$$g_{3i}(\boldsymbol{\delta}) = n_i^{-2} \mathrm{tr}\left\{ \left(\partial \rho_i / \partial \boldsymbol{\delta}\right) \mathbf{V}_i \left(\partial \rho_i / \partial \boldsymbol{\delta}\right)^\top \bar{\mathbf{V}}_i(\hat{\boldsymbol{\delta}}) \right\},$$

avec $\rho_i = \frac{n_i \sigma_v^2}{\sigma_\epsilon^2 + n_i \sigma_v^2}$, $\bar{\mathbf{X}} = N^{-1} \sum_{i=1}^k \sum_{j \in \mathcal{U}_i} \mathbf{x}_{ij}$ est le vecteur des moyennes de la population de tous les domaines, $\bar{\mathbf{X}}_i = N_i^{-1} \sum_{j \in \mathcal{U}_i} \mathbf{x}_{ij}$ est le vecteur des moyennes de la population du domaine, $\bar{\mathbf{x}}_i = n_i^{-1} \sum_{j \in s_i} \mathbf{x}_{ij}$ est le vecteur des moyennes des variables auxiliaires des unités de l'échantillon propre au domaine $\mathcal{U}_i$, et enfin, $\bar{\mathbf{V}}_i(\hat{\boldsymbol{\delta}})$ est la matrice de variance covariance asymptotique de $\hat{\boldsymbol{\delta}}$. Les termes $g_{2i}(\boldsymbol{\delta})$ et $g_{3i}(\boldsymbol{\delta})$ sont dus à l'estimation de $\boldsymbol{\beta}$ et $\boldsymbol{\delta}$ et sont d'ordre plus faible que $g_{1i}(\boldsymbol{\delta})$. Prasad et Rao (1990) proposent également un estimateur de l'EQM donné par :

$$\widehat{EQM}\left(\hat{\bar{Y}}_{iMPLSB}\right) \approx g_{1i}(\hat{\boldsymbol{\delta}}) + g_{2i}(\hat{\boldsymbol{\delta}}) + 2g_{3i}(\hat{\boldsymbol{\delta}}).$$

Chambers *et al.* (2011) proposent un estimateur du MPLSBE pour des estimateurs pseudo-linéaires de la moyenne $\bar{Y}_i$ du domaine $\mathcal{U}_i$. L'estimateur est dit pseudo-linéaire car il peut s'écrire comme une somme pondérée des unités de tout l'échantillon s avec des poids w_{ij} qui dépendent des valeurs de y. Pour le MPLSBE (1.2.12), on a concrètement :

$$\hat{\bar{Y}}_{iMPLSBE} = \left\{ N_i^{-1} \mathbf{w}_{is}^{MPLSBE}(\hat{\boldsymbol{\delta}}) \right\}^\top \mathbf{y},$$

où

$$\mathbf{w}_{is}^{MPLSBE}(\hat{\boldsymbol{\delta}}) = \mathbf{1}_i + \mathbf{V}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{X} \mathbf{A}_i^\top(\hat{\boldsymbol{\delta}}) + (N_i - n_i) \hat{\sigma}_v^2 \mathbf{V}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{1}_i,$$

ici, $\mathbf{A}_i(\hat{\boldsymbol{\delta}}) = \left\{ \sum_{j \in \mathcal{U}_i / s_i} \mathbf{x}_{ij}^\top - \hat{\sigma}_v^2 (N_i - n_i) \mathbf{1}_{n_i}^\top \mathbf{V}_i^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{X}_i \right\} \left\{ \mathbf{X}^\top \mathbf{V}^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{X} \right\}^{-1}$. De plus, $\mathbf{1}_i$ est un vecteur de taille n où toutes les composantes sont égales à 1 pour toutes les unités de l'échantillon appartenant au domaine d'intérêt $\mathcal{U}_i$ et à 0 sinon. Étant donné cette forme pseudo-linéaire du MPLSBE, Chambers *et al.* (2011) développent une approximation de l'EQM sous l'hypothèse d'une version conditionnelle du modèle (1.2.8), c'est-à-dire que l'effet aléatoire $\mathbf{v}$ est supposé fixe mais inconnu. L'estimateur de l'EQM proposé est donné par :

$$\widehat{EQM}(\hat{\bar{Y}}_{iMPLSBE}) = \hat{V}(\hat{\bar{Y}}_{iMPLSBE}) + \hat{B}^2(\hat{\bar{Y}}_{iMPLSBE}), \tag{1.2.17}$$

où

$$\hat{V}(\hat{\bar{Y}}_{iMPLSBE}) = N_i^{-2} \sum_{h=1}^{k} \sum_{j \in s_h} \left\{ a_{ihj}^2 + (N_i - n_i) n^{-1} \right\} \hat{\lambda}_{hj}^{-1} (y_{hj} - \hat{\mu}_{hj})^2$$

est un estimateur de la variance conditionnelle du MPLSBE avec

$$a_{ihj} = w_{ihj}^{MPLSBE} - I(j \in \mathcal{U}_i)$$

et

$$\hat{B}(\hat{\bar{Y}}_{iMPLSBE}) = N_i^{-1} \left(\sum_{h=1}^{k} \sum_{j \in s_h} w_{ihj}^{MPLSBE} \hat{\mu}_{hj} - \sum_{j \in \mathcal{U}_i} \hat{\mu}_{ij} \right)$$

est un estimateur du biais sous le modèle conditionnel. Il est à noter que $I(\cdot)$ est une fonction indicatrice. L'estimateur de l'EQM (1.2.17) est entièrement défini lorsque les valeurs de $\hat{\mu}_{hj}$ et $\hat{\lambda}_{hj}$ sont calculées. Les valeurs de $\hat{\mu}_{hj}$ sont données par :

$$\hat{\mu}_{hj} = \mathbf{x}_{hj}^\top \hat{\boldsymbol{\beta}} + \tilde{u}_i,$$

où $\tilde{u}_i = n_i^{-1} \sum_{i \in s_i} \left(y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} \right)$. Le terme $\hat{\mu}_{hj}$ s'écrit encore $\hat{\mu}_{hj} = \boldsymbol{\phi}_{ihj}^\top \mathbf{y}$ avec

$$\boldsymbol{\phi}_{ihj} = \begin{cases} \mathbf{C}_{iij} + n_i^{-1} \mathbf{1}_i, & \text{si } h = i, \\ \mathbf{C}_{ihj}, & \text{sinon,} \end{cases}$$

et

$$\mathbf{C}_{ihj} = \mathbf{V}_i^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{X}_i \left\{ \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\hat{\boldsymbol{\delta}}) \mathbf{X}_i \right\}^{-1} (\mathbf{x}_{hj} - \bar{\mathbf{x}}_i).$$

Finalement, le terme $\hat{\lambda}_{hj}$ est donné par

$$\hat{\lambda}_{hj} = \sum_{g=1}^{k} \sum_{l \in s_g} \left\{ (1 - \phi_{ihjl})^2 I(j = l) + \phi_{ihjl}^2 I(j \neq l) \right\}.$$

1.2.3.2. *Estimateurs Bootstrap de l'EQM du MPLSBE*

Hall & Maiti (2006a) proposent un double bootstrap paramétrique pour l'estimation de l'EQM. La procédure se déroule en deux étapes. À la première étape, un estimateur bootstrap de l'EQM est construit selon le processus décrit par les cinq phases ci-dessous :

(i) À la phase, 1, ils génèrent e_{ij}^* et v_i^*, $i = 1, \ldots, k$, $j = 1, \ldots, N_i$, de moyenne nulle, et de même loi que e_{ij} et v_{ij} et de variance respective $\hat{\sigma}_e^2$ et $\hat{\sigma}_v^2$.

(ii) Dans la deuxième phase, ils génèrent les valeurs de la variable d'intérêt pour la population bootstrap d'après le modèle (1.2.7) $y_{ij}^* = \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + v_i^* + e_{ij}^*$.

(iii) La troisième phase consiste à calculer le total bootstrap de la population finie $Y_i^* = \sum_{j \in \mathcal{U}_i} y_{ij}^*$.

(iv) Dans la quatrième phase, k échantillons s_i^* de taille respectives n_i sont indépendamment tirés dans chaque domaine $\mathcal{U}_i$. Par la suite, les estimateurs

bootstrap $\hat{\boldsymbol{\beta}}^*$ et $\hat{\boldsymbol{\delta}}^*$ des paramètres sont obtenus en utilisant les mêmes procédures que celles de l'échantillon originel.

(v) Finalement, ils estiment l'EQM par

$$\widehat{EQM}_1\left(\hat{Y}_{iMPLSBE}\right) = \frac{1}{B_1}\sum_{b_1=1}^{B_1}\left(\hat{Y}_{iMPLSBE}^{(b_1)} - Y_i^{(b_1)}\right)^2,$$

où b_1 désigne la $b_1^{\text{ième}}$ population bootstrap et

$$\hat{Y}_{iMPLSB}^{(b_1)} = \sum_{j\in s_i^{(b_1)}} y_{ij}^{(b_1)} + \sum_{j\in\mathcal{U}_i/s_i^{(b_1)}} \mathbf{x}_{ij}^\top\hat{\boldsymbol{\beta}}^{(b_1)} + (N_i - n_i)\hat{v}_i^{(b_1)}$$

avec $\hat{\mathbf{v}}^{(b_1)} = \hat{\mathbf{v}}^{(b_1)}\left\{\boldsymbol{\delta}^{(b_1)}\right\} = \mathbf{G}^{(b_1)}\mathbf{Z}^\top\mathbf{V}^{-1(b_1)}\left\{\mathbf{y}^{(b_1)} - \mathbf{X}^\top\hat{\boldsymbol{\beta}}^{(b_1)}\right\}$.

Cet estimateur a un biais d'ordre $O_p(k^{-1})$. Pour réduire ce biais, Hall & Maiti (2006a) vont à la deuxième étape répéter la procédure décrite ci-dessus, disons B_2 fois, pour chaque échantillon bootstrap généré. Par la suite, un estimateur de l'EQM à la deuxième étape est calculé :

$$\widehat{EQM}_2\left(\hat{Y}_{iMPLSBE}\right) = \frac{1}{B_1}\sum_{b_1=1}^{B_1}\frac{1}{B_2}\sum_{b_2=1}^{B_2}\left[\hat{Y}_{iMPLSBE}^{\{b_1(b_2)\}} - Y_i^{\{b_1(b_2)\}}\right]^2.$$

L'estimateur double bootstrap de l'EQM d'ordre $o_p(k^{-1})$ est alors obtenu en corrigeant le biais par les procédures classiques telles que :

$$\widehat{EQM}_{DB}\left(\hat{Y}_{iMPLSB}^{(b)}\right) = \begin{cases} 2\widehat{EQM}_1 - \widehat{EQM}_2, & \text{si } \widehat{EQM}_1 \geq \widehat{EQM}_2, \\ \widehat{EQM}_1\exp\left\{\left(\widehat{EQM}_1 - \widehat{EQM}_2\right)/\widehat{EQM}_1\right\}, & \text{sinon.} \end{cases}$$

Un désavantage du bootstrap paramétrique est qu'il peut être biaisé si les hypothèses sur les erreurs et les effets aléatoires ne sont pas valides (Opsomer *et al.* 2008). Hall & Maiti (2006b) proposent une estimation non-paramétrique de l'EQM du MPLSBE en utilisant le bootstrap. Ils font une approximation de l'EQM qui s'exprime uniquement en fonction des moments d'ordre 2 et 4 des erreurs et des effets aléatoires. Puis, ils montrent qu'il suffit de construire une procédure bootstrap, appelée «moment-matching» ou encore «wild» bootstrap, qui est valide pour ces moments d'ordre 2 et 4.

Soit γ_e et γ_v les moments d'ordre 4 de e_{ij} et v_i respectivement. Les moments d'ordre 2 et 4 de e_{ij} et v_i doivent satisfaire les conditions suivantes : $\sigma_e^4 \leq \gamma_e$ et $\sigma_v^4 \leq \gamma_v$ respectivement. Sous ces hypothèses, Hall & Maiti (2006b) proposent l'algorithme décrit ci-dessous :

(i) Premièrement, ils considèrent deux réels positifs $z_2, z_4 > 0$ avec $z_2^2 \leq z_4$. Soit $D(z_2, z_4)$ la distribution d'une variable aléatoire Z pour laquelle on a $E(Z) = 0$, $E(Z^2) = z_2$ et $E(Z^4) = z_4$. Soit $\mathcal{D}$ la classe de ces distributions

avec exactement un seul représentant $D(z_2, z_4)$ pour chaque paire (z_2, z_4). La classe des distributions $D(z_2, z_4)$ est très vaste. Un exemple simple est la classe $D(1, p^{-1})$, $0 < p < 1$, de la distribution Z prenant trois valeurs et définie par $P(Z = 0) = 1 - p$, $P(Z = -p^{-1/2}) = \frac{1}{2}p$, $P(Z = p^{-1/2}) = \frac{1}{2}p$. Ainsi, un choix de la distribution $D(z_2, z_4)$ serait $z_2^{1/2} Z$, où $p = z_2^2/z_4$.

(ii) Ensuite, ils considèrent les estimateurs $\hat{\sigma}_e^2$, $\hat{\sigma}_v^2$, de σ_e^2, σ_v^2 respectivement ; ils considèrent également les estimateurs respectifs $\hat{\gamma}_e$ et $\hat{\gamma}_v$ de γ_e et γ_v vérifiant $\hat{\sigma}_e^2 \leq \hat{\gamma}_e$, et $\hat{\sigma}_v^2 \leq \hat{\gamma}_v$ respectivement.

(iii) Puis, ils génèrent e_{ij}^*, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ et v_i^*, $i = 1, \ldots, k$ à partir des lois $D(\hat{\sigma}_e^2, \hat{\gamma}_e)$ et $D(\hat{\sigma}_v^2, \hat{\gamma}_v)$ respectivement. Le vecteur des unités de l'échantillon bootstrap est construit à partir d'un modèle similaire à (1.2.8) $\mathbf{y}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{v}}^* + \mathbf{e}^*$.

(iv) Le reste de l'algortihme procède exactement comme dans Hall & Maiti (2006a) décrit ci-dessus en faisant un double bootstrap pour obtenir un estimateur de l'EQM d'ordre $o_p(k^{-1})$.

1.2.4. Estimation robuste des petits domaines

Dans le cas des modèles au niveau des unités, Chambers & Tzavidis (2006) utilisent les quantiles robustes pour proposer une alternative robuste à l'estimation de la moyenne des petits domaines. Ils définissent le quantile d'ordre q de y conditionnellement à $\mathbf{x}$ comme une fonction linéaire de $\boldsymbol{\beta}$: $Q_q(x) = \mathbf{x}^\top \boldsymbol{\beta}_q$. Ils font l'hypothèse que les quantiles robustes obéissent au modèle linéaire mixte et proposent un pseudo estimateur de la moyenne donné par :

$$\mu_i = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in \mathcal{U}_i / s_i} \mathbf{x}_{ij}^\top \boldsymbol{\beta}_{q_{ij}} \right\}.$$

Par la suite, Chambers & Tzavidis (2006) vont faire une approximation heuristique par un développement de Taylor d'ordre 1 de cette dernière expression pour obtenir l'équation suivante :

$$\mu_i \approx N_i^{-1} \left[\sum_{j \in s_i} y_{ij} + \sum_{j \in \mathcal{U}_i / s_i} \mathbf{x}_{ij}^\top \boldsymbol{\beta}_{q_i} + \sum_{j \in \mathcal{U}_i / s_i} \mathbf{x}_{ij}^\top \left\{ \frac{\partial \boldsymbol{\beta}_q(q_i)}{\partial q} \right\} (q_{ij} - q_i) \right],$$

où $q_i = N_i^{-1} \sum_{j \in \mathcal{U}_i}$ est la moyenne des quantiles du domaine d'intérêt. Ils affirment que le dernier terme à droite de l'approximation ci-dessus est négligeable par rapport au deuxième. De ce fait, ils proposent d'estimer de manière robuste la moyenne du petit domaine, dénoté $\hat{Y}_{iCT}$, par :

$$\hat{Y}_{iCT} = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in \mathcal{U}_i / s_i} \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{iqR} \right\}, \tag{1.2.18}$$

où le vecteur $\hat{\boldsymbol{\beta}}_{iqR}$ est obtenu en résolvant l'équation

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\psi\left(r_{ij}\right)\mathbf{x}_{ij}=\mathbf{0},$$

avec $r_{ij}=y_{ij}-\mathbf{x}^{\top}\boldsymbol{\beta}_{q_i}$, $\psi\left(r_{ij}\right)=2\psi\left(\sigma^{-1}r_{ij}\right)\left\{q_{ij}I\left(r_{ij}>0\right)+(1-q_{ij})I\left(r_{ij}\leq 0\right)\right\}$, ψ est une fonction utilisée pour réduire l'influence des valeurs aberrantes et $\sigma^2=\sigma_v^2+\sigma_v^2$ désigne la variance de y_{ij} et peut être estimé de manière robuste en utilisant par exemple la médiane des écarts absolus à la médiane des résidus r_{ij}.

Tzavidis *et al.* (2010) notent que l'estimateur (1.2.18) est de type «plug-in» et donc est potentiellement biaisé. Ils proposent alors une version corrigée pour le biais basée sur l'estimation des fonctions de distribution. Pour cela, ils utilisent l'approche de Welsh & Ronchetti (1998) qui conduit à un estimateur robuste corrigé pour le biais de la fonction de distribution ($\hat{F}_{iWR}$) des unités du domaine $\mathcal{U}_i$ de la forme :

$$\begin{aligned}\hat{F}_{iWR}(t) &= N_i^{-1}\sum_{j\in s_i}I(y_{ij}\leq t) \qquad (1.2.19)\\ &+ n_i^{-1}N_i^{-1}\sum_{j\in s_i}\sum_{l\in\mathcal{U}_i/s_i}I\left[\mathbf{x}_{il}^{\top}\hat{\boldsymbol{\beta}}_{iqR}+w_i^{MQ}\psi_c\left\{(y_{ij}-\mathbf{x}_i^{\top}\hat{\boldsymbol{\beta}}_{iqR})/w_{ij}^{MQ}\right\}\leq t\right].\end{aligned}$$

Ici, w_i^{MQ} est un estimateur robuste de l'écart type des résidus $y_{ij}-\mathbf{x}_i^{\top}\hat{\boldsymbol{\beta}}_{iqR}$, et ϕ_c est une fonction bornée de type Huber. Tzavidis *et al.* (2010) obtiennent alors un estimateur robuste de la moyenne du petit domaine, dénoté $\hat{Y}_{iTMC}$, en prenant l'espérance de la fonctionnelle définie par (1.2.19). Ceci conduit à un estimateur corrigé pour le biais de la forme

$$\begin{aligned}\hat{Y}_{iTMC} &= f_i\bar{y}_i+(1-f_i)\bar{\mathbf{x}}_{ri}^{\top}\hat{\boldsymbol{\beta}}_{iqR}\\ &+ (1-f_i)n_i^{-1}\sum_{j\in s_i}w_i^{MQ}\psi_c\left\{(y_{ij}-\mathbf{x}_{ij}^{\top}\hat{\boldsymbol{\beta}}_{iqR})/w_i^{MQ}\right\}, \qquad (1.2.20)\end{aligned}$$

où $f_i=n_i/N_i$, $\bar{\mathbf{x}}_{ri}^{\top}=\sum_{j\in\mathcal{U}_i/s_i}\mathbf{x}_{ij}^{\top}/(N_i-n_i)$.

Sinha & Rao (2009) proposent une alternative qui est une version robuste du meilleur prédicteur linéaire sans biais empirique. Leur approche s'effectue en deux étapes. À la première étape des estimateurs robustes $\hat{\boldsymbol{\beta}}_R$, $\hat{\boldsymbol{\delta}}_R$ des paramètres du modèle (1.2.8) sont obtenus en résolvant les équations (1.2.21) et (1.2.22) ci-dessous

$$\mathbf{X}^{\top}\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}) = \mathbf{0}, \qquad (1.2.21)$$

et

$$\boldsymbol{\Psi}^{\top}(\mathbf{r})\mathbf{U}^{1/2}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\delta_l}\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r})-tr\left(\mathbf{K}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\delta_l}\right) = \mathbf{0}, \qquad (1.2.22)$$

où $l = 1, \ldots, q$, $\mathbf{r} = \mathbf{U}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, $\mathbf{U} = diag(\mathbf{V})$, $\mathbf{K} = E\{\psi_c^2(r)\}\mathbf{I}$ avec $r \sim N(0,1)$ et $\mathbf{I}$ dénote une matrice du même ordre que $\mathbf{V}$. Le vecteur $\boldsymbol{\Psi}$ est d'ordre $n \times 1$ et est défini par $\boldsymbol{\Psi}(\mathbf{u}) = \{\psi_c(u_1), \psi_c(u_2), \ldots, \psi_c(u_{n_i})\}^\top$, avec ψ_c la fonction de Huber définie par

$$\psi_c(x) = \max\{-c, \min(x, c)\} \tag{1.2.23}$$

et la constante de robustesse $c = 1.345$. À la deuxième étape, les estimateurs robustes $\hat{\boldsymbol{\beta}}_R$ et $\hat{\boldsymbol{\delta}}_R$ sont utilisés pour obtenir un estimateur robuste de $\mathbf{v}$, dénoté $\hat{\mathbf{v}}_R$, obtenu en résolvant les équations de Fellner (1986) :

$$\mathbf{Z}^\top \mathbf{R}^{-1/2} \boldsymbol{\Psi}\left\{\mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v})\right\} - \mathbf{G}^{-1/2}\psi_c\left(\mathbf{G}^{-1/2}\mathbf{v}\right) = 0.$$

La version robuste du meilleur prédicteur linéaire sans biais de $\bar{Y}_{iMPLSB}$ est donc obtenue en substituant $\hat{\boldsymbol{\beta}}_R$ et $\hat{\mathbf{v}}_{iR}$ dans (1.2.12), ce qui conduit à :

$$\hat{\bar{Y}}_i^{SR} = f_i \bar{y}_i + (1 - f_i)\left(\bar{\mathbf{x}}_{ri}^\top \hat{\boldsymbol{\beta}}_R + \bar{\mathbf{z}}_{ri}^\top \hat{\mathbf{v}}_{iR}\right). \tag{1.2.24}$$

Chambers *et al.* (2014) notent que l'estimateur de Sinha & Rao (1.2.24) est de type «plug-in» et de ce fait, est potentiellement biaisé. Ils proposent une version corrigée pour le biais qui est une application directe de l'approche de Tzavidis *et al.* (2010). L'estimateur corrigé pour le biais est notamment obtenu en remplaçant dans (1.2.20) les formes projectives $\bar{\mathbf{x}}_{ri}^\top \hat{\boldsymbol{\beta}}_{iqR}$ et $\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_{iqR}$ par $\bar{\mathbf{x}}_{ri}^\top \hat{\boldsymbol{\beta}}_R + \bar{\mathbf{z}}_{ri}^\top \hat{\mathbf{v}}_{iR}$ et $\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R + \mathbf{z}_{ij}^\top \hat{\mathbf{v}}_{iR}$ respectivement. Ceci conduit à un estimateur corrigé pour le biais de la forme :

$$\begin{aligned}
\hat{Y}_{iCCST} &= f_i \bar{y}_i + (1 - f_i)\left(\bar{\mathbf{x}}_{ri}^\top \hat{\boldsymbol{\beta}}_R + \bar{\mathbf{z}}_{ri}^\top \hat{\mathbf{v}}_{iR}\right) \\
&+ (1 - f_i) n_i^{-1} \sum_{j \in s_i} w_i^R \psi_c\left\{\frac{y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \mathbf{z}_{ij}^\top \hat{\mathbf{v}}_{iR}}{w_i^R}\right\} \\
&= \hat{\bar{Y}}_i^{SR} + (1 - f_i) n_i^{-1} \sum_{j \in s_i} w_i^R \psi_c\left\{\frac{y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \mathbf{z}_{ij}^\top \hat{\mathbf{v}}_{iR}}{w_i^R}\right\}, \qquad (1.2.25)
\end{aligned}$$

où w_i^R est la médiane des écarts à la médiane (MAD) des résidus $y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \mathbf{z}_{ij}^\top \hat{\mathbf{v}}_{iR}$. Il est à noter que l'estimateur (1.2.25) s'obtient en deux étapes. À la première étape, la fonction de Huber ψ_c avec une constante $c = 1.345$ est utilisée pour calculer $\hat{\boldsymbol{\beta}}_R$ et $\hat{\mathbf{v}}_{iR}$. À la deuxième étape, la fonction d'influence ψ_c est également celle de Huber mais avec une constante $c = 3$.

Bibliographie

[1] Battese, G. E., Harter R. M. & Fuller W. A. (1988). An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.

[2] Beaumont, J.-F., Haziza, D. & Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, **100**, 555–569.

[3] Beaumont, J.-F., & Rivest L.-P. (2009). Dealing with outliers in survey data. In C. R. Rao and D. Pfefferman (Editors), *Handbook of Statistics, Sample Surveys, Design Methods and Applications*, **29A**, 247–279.

[4] Chambers, R. L. (1986). Outliers robust finite population estimation. *Journal of the American Statistical Association*, **81**, 1063–1069.

[5] Chambers, R. L., Chandra, H., Salvati, N. & Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society. Series B*, **76**, 47–69.

[6] Chambers, R. L., Chandra, H., & Tzavidis, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methodology*, **37**, 153–170.

[7] Chambers, R. L. & Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, **93**, 255–268.

[8] Deville, J-C., Särndal, C-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**, 376–382.

[9] Duchesne, P. (1999). Robust calibration estimators. *Survey Methodology*, **25**, 43–56.

[10] Efron B. & Stein C. (1981). The Jackknife estimate of variance. *The Annals of Statistics*, **9**, 586–596

[11] Fay, R. E. & Herriot, R. A. (1979). Estimates of income for small places : An application of James–Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

[12] Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics*, **28**, 51–60.

[13] Gwet, J.-P & Rivest, L.-P. (1992). Outlier resistant alternatives to the ratio estimator. *Journal of the American Statistical Association*, **87**, 736–739.

[14] Hall, P. & Maiti, T. (2006a). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society. Series B*, **68**, 221–238.

[15] Hall, P. & Maiti, T. (2006b). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *The Annals of Statistics*, **34**, 1733–1750.

[16] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. E. (1986), *Robust statistics : the approach based on influence functions*, New York : John Wiley.

[17] Harville, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320–338.

[18] Hidiroglou, M.A. & Srinath, K.P. (1981). Some estimators of population total category large units. *Journal of the American Statistical Association*, **78**, 690–695.

[19] Horvitz, D. G. & Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663–685.

[20] Huber, P. J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.

[21] Huber, P. J. (1973) Robust regression : asymptotics, conjectures, and Monte Carlo. *Annals of Statistics*, **1**, 799–821.

[22] Kackar, R.N. & Harville D.A. (1984) Approximations for standard errors of estimators of fixed and random effect in mixed linear models. *Journal of the American Statistical Association*, **79**, 853–862.

[23] Kokic, P.N. & Bell, P.A. (1994). Optimal winzorizing cutoffs for a stratified finite population estimator. *Journal of Official Statistics*, **10**, 419–435.

[24] Krasker W.S. (1980) Estimation in linear regression models with disparate data points. *Econometrica*, **48**, 1333–1346.

[25] Lee, H. (1995). *Outliers in business surveys*. In *Business Survey Methods, Wiley Series in Probability and Mathematical Statistics*, Ed. B. G Cox, D. A. Binder, B. N. Chinnappa, A. Christianson, M. J. Colledge & P. S. Kott, pp. 503–526, New York : John Wiley.

[26] Moreno-Rebollo, J.L., Muñoz-Reyes, A. M., & Muñoz-Pichardo, J. (1999). Influence diagnostic in survey sampling : conditional bias. *Biometrika*, **86**, 923–928.

[27] Moreno-Rebollo, J. L., Muñoz-Reyes, A. M., Jimenez-Gamero, M. D. & Muñoz-Pichardo, J. (2002). Influence diagnostics in survey sampling : estimating the conditional bias. *Metrika*, **55**, 209–214.

[28] Muñoz-Pichardo, J., Muñoz-Garcia, J., Moreno-Rebollo, J. L. & Pino-Mejias, R. (1995). A new approach to influence analysis in linear models. *Sankhya A*, **57**, 393–409.

[29] Opsomer, J. P., Claeskens, G., Ranalli, M. G., Kauemann, G. & Breidt, F. J. (2008). Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society. Series B*, **70**, 265-286.

[30] Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, **28**, 1, 40-68.

[31] Prasad, N. G. N. & 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.

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

[33] Rao, J. N. K. (2005). Inferential issues in small area estimation : Some new developments. *Statistics in Transition*, **7**, 513–526.

[34] Särndal, C. E., Swensson, B. & Wretman, J. H. (1992) *Model Assisted Survey Sampling*. New York : Springer-Verlag.

[35] Sinha, S. K. & Rao, J. N. K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, **37**, 381–399.

[36] Stahel, W. A. & Welsh, A. (1997). Approaches to robust estimation in the simplest variance components model. *Journal of Statistical Planning and Inference*, **57**, 295–319.

[37] Statistique Canada (2012). Système de classification des industries de l'Amérique du Nord (SCIAN) Canada. *Ministre de l'Industrie du Canada, Ottawa,* ISBN 978-1-100-98303-5, No 12-501-X au catalogue.

[38] Tukey, J. W, & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample :Trimming/Winsorization 1. *Sankhyā : The Indian Journal of Statistics. Series A*, **25**, 331–352.

[39] Tzavidis, N., Marchetti, S. & Chambers, R. L. (2010). Robust estimation of small-area means and quantiles. *Australian & New Zealand Journal of Statistics*, **52**, 167–186.

[40] Welsh, A. H. & Ronchetti, E. (1998). Bias-calibrated estimation from sample surveys containing outliers. *Journal of the Royal Statistical Society. Series B*, **60**, 413–428.

Chapitre 2

CONTROLLING THE BIAS OF ROBUST SMALL AREA ESTIMATORS

Abstract

Sinha & Rao (2009) proposed estimation procedures designed for small area means, based on robustified maximum likelihood estimators and robust empirical best linear unbiased predictors. Their methods are of the plug-in type, and in view of the results of Chambers (1986), the robust estimators may be biased. Bias-corrected estimators have been proposed by Chambers *et al.* (2014). Here, we investigate two new approaches : one relying on the work of Chambers (1986), and the second using the concept of conditional bias to measure the influence of units in the population. These two classes of robust small area estimators also include correction terms for the bias but are both fully bias-corrected, in the sense that the corrections account for the potential impact of the other domains on the small area of interest. We demonstrate that the fully bias-corrected estimators differ substantially from the proposals of Chambers *et al.* (2014). Non-negligible biases may arise for the Sinha–Rao method, while the proposed methods are less biased, controlled by appropriate choices of ψ-functions and tuning constants. Monte Carlo simulations suggest that the Sinha–Rao method and the bias-adjusted estimator of Chambers *et al.* (2014) may exhibit a large bias, while the new procedures often offer lower bias and mean squared error. A parametric bootstrap procedure is considered for constructing confidence intervals.

Key words : Corrected-bias estimator ; conditional bias ; influence measure ; outliers ; model-based inference ; survey sampling ; small area estimation.

2.1. Introduction

In the last two decades, the demand for small area estimators has been growing in most countries. This has led survey statisticians to develop theoretically sound

and yet practical estimation procedures, providing reliable estimators for the variables of interest at the small area level (e.g., Rao, 2003). Small areas are defined as domains whose sample sizes are not large enough to justify the use of direct estimators. Typically, the information in the domain-specific sample data does not suffice to obtain efficient estimators. For example, traditionally direct estimators, such as the classical Horvitz–Thompson estimator, are not appropriate in such circumstances. To overcome these small sample size problems, small area estimation techniques have been developed, based on model-based methods and the use of auxiliary information. Consequently, so-called indirect estimators have been proposed, typically relying on linear models, and observed values from related areas are used in order to predict the small area characteristics. The construction of indirect estimators requires auxiliary information at the area and/or unit levels, and efficiency requires a correct specification of the model linking the study variable y to the auxiliary information. See Rao (2003, 2005) for comprehensive accounts on small area estimation.

Let U be a finite population of size N, which is partitioned into k subpopulations $U_1, \ldots, U_k$, of sizes $N_1, \ldots, N_k$, respectively. Thus $U = \bigcup_{i=1}^{k} U_i$ such that $U_i \cap U_l = \emptyset$, $i \neq l$, and $N = \sum_{i=1}^{k} N_i$. The domain sizes N_i are assumed to be known. For a variable of interest y, let y_{ij} be the value of y attached to the jth element of the ith area. The parameters of interest are the small area means $\theta_i = N_i^{-1} \sum_{j \in U_i} y_{ij}, (i = 1, \ldots, k)$.

A sample s of size n is selected from U according to a given sampling plan $p(s)$. Let $s_i = s \cap U_i$ denote the ith area specific sample of size n_i. Thus, $s = \bigcup_{i=1}^{k} s_i$ and $n = \sum_{i=1}^{k} n_i$. Let $\mathbf{x}_{ij}$ be a deterministic vector of dimension p, containing the auxiliary variables corresponding to unit j of the area i. The ith area-specific totals, $\sum_{j \in U_i} \mathbf{x}_{ij}$, are also supposed to be available.

In this paper, we focus on the basic unit-level model that can be expressed as

$$y_{ij} = \mathbf{x}_{ij}^{\top} \boldsymbol{\beta} + v_i + e_{ij} \quad (j = 1, \ldots, n_i), \tag{2.1.1}$$

for the ith area, where v_i denotes a random variable associated with the small area i. The use of a random effects model allows us to account for the between-area variation which is not explained by the auxiliary information. The error term is given by e_{ij}, and it is assumed that e_{ij} and v_i are independent random variables for all i and j. Classical distributional assumptions for the error terms and the random effects rely on normal theory, assuming that e_{ij} are independent $\mathcal{N}(0, \sigma_e^2)$, and v_i are independent and identically distributed $\mathcal{N}(0, \sigma_v^2)$ random variables, respectively. We collect the variance components in the vector $\boldsymbol{\delta} = (\sigma_e^2, \sigma_v^2)^{\top}$. Model (2.1.1) is a special case of the more general linear mixed model. In matrix

notation, the unit-level model (2.1.1) can be compactly rewritten as

$$y_i = \mathbf{X}_i\beta + v_i\mathbf{1}_{n_i} + \mathbf{e}_i \quad (i = 1, \ldots, k), \tag{2.1.2}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^\top$ is a matrix of dimension $n_i \times p$ and $\mathbf{1}_{n_i}$ corresponds to a vector of ones of dimension $n_i \times 1$. The classical normality assumption for the error term is $\mathbf{e}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_e^2\mathbf{I}_{n_i})$, with $\mathbf{e}_i = (e_{i1}, \ldots, e_{in_i})^\top$, where $\mathbf{I}_{n_i}$ represents the identity matrix of order n_i. The variance-covariance matrix of $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is thus $\mathbf{V}_i(\delta) = \sigma_e^2\mathbf{I}_{n_i} + \sigma_v^2\mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top$. Popular estimation methods for $\boldsymbol{\delta}$ are maximum likelihood or restricted maximum likelihood estimators which in turn can be used to obtain empirical best linear unbiased estimator of $\boldsymbol{\beta}$ and empirical best linear unbiased predictors of v_i. See Section 2.2.

Outliers occur frequently in surveys, especially in business surveys, since economic variables whose distributions are highly skewed are typically studied. Usually, the distributions of such variables display heavy right tails, and outlying units present unusual large values for such variables of interest. In an ideal survey sampling set-up, the survey design accounts such large units. However, in applications, imperfect auxiliary information when stratifying the population may yield outliers in the sample. In the terminology of Chambers (1986), these outliers are representative, in the sense that they are representative of the non-sampled part of the population. See Lee (1995), Duchesne (1999) and Beaumont & Rivest (2009) for comprehensive reviews of outliers in survey sampling. In a small area framework, empirical best linear unbiased predictors are efficient under correct model specification and distributional assumptions, but they may be highly sensitive to the presence of outliers. Including or excluding outlying units in the calculation of the empirical best linear unbiased predictor may have a large impact on its magnitude. See, e.g., Fellner (1986), Stahel & Welsh (1997) and Sinha & Rao (2009).

Robust small area estimation has received considerable attention in recent years. Chambers & Tzavidis (2006) proposed robust estimators based on the so-called M-quantile modelling approach of the conditional distribution of the variable of interest given the auxiliary information. Ghosh *et al.* (2008) studied robust procedures using Bayesian methods. Tzavidis *et al.* (2010) studied robust prediction of small area means and distributions. They proposed a general framework for robust small area estimation, based on representing the small area estimator as a functional of a predictor of this small area cumulative distribution function. Sinha & Rao (2009) studied the robustness of the empirical best linear unbiased predictor and proposed resistant methods for small area estimation. They estimated the mean squared errors of the robust estimators of the small

area means, using a parametric bootstrap procedure. Chambers *et al.* (2014) studied robust small area estimation, and noted that the methods of Sinha & Rao (2009) are based on plug-in robust versions of the original empirical best linear unbiased predictors. Work of Chambers (1986) suggests that these approaches may introduce non-negligible prediction biases, and it appears necessary to include additional terms to correct for the bias. Chambers *et al.* (2014) proposed bias-corrected predictive estimators which can be less biased than the estimators of Sinha & Rao (2009). Chambers *et al.* (2011) also derived analytical mean squared error estimators for outlier robust predictors of small area means.

The main objective here is to propose new robust estimators for small area means with correction terms for the bias. Two approaches are studied. Under the first approach, the arguments of Chambers (1986) are adapted to the unit-level model. In the second approach, the concept of conditional bias is used. Muñoz-Pichardo *et al.* (1995) proposed to study the influence of a given observation in general linear models using the conditional bias. In Moreno-Rebollo *et al.* (1999), the concept of conditional bias is proposed as a measure of influence in the context of design-based inference. Beaumont *et al.* (2013) advocated its use in a model-based framework, proposed robust estimators for population means and totals and established links with the robust methods of Chambers (1986) in classical linear models. Interestingly, both classes of predictors can be interpreted as a compromise between the Sinha–Rao method and the non-robust empirical best linear unbiased predictors, through ψ-functions and tuning constants.

2.2. Preliminaries

Several robust estimation techniques have been developed for estimating the unit-level model (2.1.1). First, the empirical best linear unbiased predictor method finds explicit expressions for the best linear unbiased estimator and best linear unbiased predictors of $\boldsymbol{\beta}$ and $v_1, \ldots, v_k$, respectively, assuming $\boldsymbol{\delta}$ known. They are given by

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = \left(\sum_{h=1}^{k} \mathbf{X}_h^\top \mathbf{V}_h^{-1} \mathbf{X}_h \right)^{-1} \sum_{h=1}^{k} \mathbf{X}_h^\top \mathbf{V}_h^{-1} \mathbf{y}_h,$$

and $\tilde{v}_h(\boldsymbol{\delta}) = \sigma_v^2 \mathbf{1}_{n_h}^\top \mathbf{V}_h^{-1} (\mathbf{y}_h - \mathbf{X}_h \tilde{\boldsymbol{\beta}})$ with $\mathbf{V}_h \equiv \mathbf{V}_h(\boldsymbol{\delta})$ and $\tilde{\boldsymbol{\beta}} \equiv \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta})$. These expressions can be justified using a penalized-likelihood criterion, and they are derived in Rao (2003, Chapter 6). The best linear unbiased predictor of θ_i can be written as

$$\hat{\theta}_i(\boldsymbol{\delta}) = N_i^{-1} \left[\sum_{j \in s_i} y_{ij} + \sum_{j \in U_i - s_i} \{ \mathbf{x}_{ij}^\top \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) + \tilde{v}_i(\boldsymbol{\delta}) \} \right]. \tag{2.2.1}$$

It can be shown that the predictor (2.2.1) minimizes the model mean squared error under model (2.1.1) in the class of linear model unbiased predictors of θ_i, see Rao (2003, p. 98). Using maximum likelihood or restricted maximum likelihood techniques for $\boldsymbol{\delta}$, a suitable estimator $\hat{\boldsymbol{\delta}}$ is plugged in the best linear unbiased estimator and best linear unbiased predictors. This two-stage method leads to the empirical best linear unbiased estimator of $\boldsymbol{\beta}$ and the empirical best linear unbiased predictors of v_h, that is $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\delta}})$ and $\hat{v}_h = \tilde{v}_h(\hat{\boldsymbol{\delta}})$. The empirical best linear unbiased predictor of θ_i is thus given by

$$\hat{\theta}_{i\text{EBLUP}} = \hat{\theta}_i(\hat{\boldsymbol{\delta}}). \tag{2.2.2}$$

Since the best linear unbiased predictor and empirical best linear unbiased predictor methods are sensitive to the presence of outliers, robust empirical best linear unbiased predictors have been proposed by Fellner (1986). Robust estimators and predictors of $\boldsymbol{\beta}$ and $v_1, \ldots, v_k$ are developed in robustifying the model equations leading to the best linear unbiased estimator and best linear unbiased predictors, for known $\boldsymbol{\delta}$. The so-called Fellner equations are

$$\sigma_e^{-1} \sum_{h=1}^{k} \mathbf{X}_h^\top \boldsymbol{\Psi}\{\sigma_e^{-1}(\mathbf{y}_h - \mathbf{X}_h\boldsymbol{\beta} - v_h \mathbf{1}_{n_h})\} = 0, \tag{2.2.3}$$

$$\sigma_e^{-1} \mathbf{1}_{n_i}^\top \boldsymbol{\Psi}\{\sigma_e^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - v_i \mathbf{1}_{n_i})\} - \sigma_v^{-1}\psi_b(\sigma_v^{-1} v_i) = 0, \tag{2.2.4}$$

where $i = 1, \ldots, k$, the $\boldsymbol{\Psi}$ function is an $n_i \times 1$ vector defined as $\boldsymbol{\Psi}(u) = \{\psi_b(u_1), \psi_b(u_2), \ldots, \psi_b(u_{n_i})\}^\top$, with

$$\psi_b(u) = \min\{|b|, \max(-|b|, u)\}. \tag{2.2.5}$$

In a classical robust estimation framework, the choice of the tuning constant b is often dictated by efficiency considerations in a perfectly observed normal model. For example, a popular choice is $b = 1.345$. Then, using robustified Henderson equations (Rao, 2003 and Sinha & Rao, 2009), robust estimators of the variance components $\boldsymbol{\delta}$ are obtained.

Sinha & Rao (2009) used an alternative two-step approach for constructing robust estimators. Their approach is based on robustified score equations of the Gaussian maximum likelihood estimators. Assuming normality, a Gaussian multivariate model can be formulated as a function of $(\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top)^\top$. The maximum likelihood estimators are then solutions of the score equations

$$\mathbf{X}^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}, \tag{2.2.6}$$

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \boldsymbol{\delta}_l} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \operatorname{tr}\left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \boldsymbol{\delta}_l}\right) = \mathbf{0}, \tag{2.2.7}$$

where $l = 1, 2$. Since the maximum likelihood estimators are non robust, (2.2.6) and (2.2.7) need to be robustified. One proposal has been considered in Huggins (1993). See also Richardson & Welsh (1995). Sinha & Rao (2009) proposed the robustified maximum likelihood equations

$$\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{U}^{1/2} \boldsymbol{\Psi}(\mathbf{r}) = \mathbf{0}, \qquad (2.2.8)$$

$$\boldsymbol{\Psi}^\top(\mathbf{r}) \mathbf{U}^{1/2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_l} \mathbf{V}^{-1} \mathbf{U}^{1/2} \boldsymbol{\Psi}(\mathbf{r}) - \operatorname{tr}\left(\mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \boldsymbol{\delta}_l} \right) = \mathbf{0}, \qquad (2.2.9)$$

where $l = 1, \ldots, k$, $\mathbf{r} = \mathbf{U}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, $\mathbf{U} = \operatorname{diag}(\mathbf{V})$, $\mathbf{K} = E\{\psi_b^2(r)\}\mathbf{I}_n$, and r is a standard normal random variable. Solving equations (2.2.8) and (2.2.9) usually requires iterative algorithms such as Newton–Raphson algorithm. Let $\hat{\boldsymbol{\beta}}_R$ and $\hat{\boldsymbol{\delta}}_R = (\hat{\sigma}^2_{eR}, \hat{\sigma}^2_{vR})^\top$ be the robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, obtained by solving (2.2.8) and (2.2.9). In the second step, robust predictors of v_i are derived, obtained by solving (2.2.4), conditionally given $(\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\delta}}_R^\top)^\top$. We denote these predictors by $\hat{v}_{iR} \equiv \hat{v}_{iR}(\hat{\boldsymbol{\delta}}_R)$. Having estimated robustly all the parameters, the Sinha–Rao robust empirical best linear unbiased predictor of the plug-in type for θ_i can be written as

$$\hat{\theta}_{i\mathrm{SR}} = N_i^{-1} \sum_{j \in s_i} y_{ij} + N_i^{-1} \sum_{j \in U_i / s_i} \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R + \left(1 - n_i N_i^{-1}\right) \hat{v}_{iR}. \qquad (2.2.10)$$

Chambers *et al.* (2014) argued that $\hat{\theta}_{i\mathrm{SR}}$ in (2.2.10) may involve a large prediction bias when the population data are drawn from a mixture distribution, for which the means of the outliers and non-outliers are different. They proposed a bias-corrected version of the Sinha–Rao robust empirical best linear unbiased predictor, using an approach similar to that advocated by Welsh & Ronchetti (1998) for robust prediction of the empirical distribution function of y values in area i. An estimator of the area mean θ_i is then defined by the value of the mean functional

$$\hat{\theta}_{i\mathrm{CCST}} = \hat{\theta}_{i\mathrm{SR}} + \left(n_i^{-1} - N_i^{-1}\right) \sum_{j \in s_i} \phi_i \psi_c \left(\frac{y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR}}{\phi_i} \right), \qquad (2.2.11)$$

where the weights ϕ_i are the median absolute deviation of the ith area residuals, $y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR}$, and $\psi_c(t)$ is given by (2.2.5). When the tuning constant $c = 0$, the estimator $\hat{\theta}_{i\mathrm{CCST}}$ defined by (2.2.11) reduces to the robust estimator $\hat{\theta}_{i\mathrm{SR}}$. Thus, the bias-corrected estimator includes an additional term, whose role consists of controlling the potential bias. The correction factor given by the last term on the right side of (2.2.11) is based only on $j \in s_i$. Thus, the bias-corrected robust empirical best linear unbiased predictor of Chambers *et al.* (2014) neglects the information associated with the units which are not in the area i, and which may

still influence the estimators. On the other hand, $\hat{\theta}_{i\text{CCST}}$ should be less biased when $c = \infty$, at the price of being less robust. The estimator (2.2.11) with $c = \infty$ does not reduce to the non-robust predictor of the area mean $\hat{\theta}_{i\text{EBLUP}}$ defined by (2.2.2). This is due to the fact that $\hat{\theta}_{i\text{EBLUP}} \neq \hat{\theta}_{i\text{SR}} + \left(n_i^{-1} - N_i^{-1}\right) \sum_{j \in s_i} (y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR})$, in general. This issue is also discussed in Section 2.3.1

2.3. Fully bias-corrected robust predictors

2.3.1. Robust predictor of the area mean based on Chambers' approach

The predictor $\hat{\theta}_{i\text{EBLUP}}$ defined by (2.2.2) can alternatively be written as

$$\hat{\theta}_{i\text{EBLUP}} = N_i^{-1} \sum_{h=1}^{k} \sum_{j \in s_h} w_{ihj}(\hat{\boldsymbol{\delta}}) y_{hj}, \tag{2.3.1}$$

where the weights satisfy the relations

$$w_{ihj}(\boldsymbol{\delta}) = \begin{cases} k^{-1} \mathbf{a}_i \mathbf{X}_h^\top \mathbf{C}_h^{(j)} \quad (j \in s_h), \\ 1 + k^{-1} \mathbf{a}_i \mathbf{X}_i^\top \mathbf{C}_i^{(j)} + (N_i - n_i) \sigma_v^2 \mathbf{1}_{n_i}^\top \mathbf{C}_i^{(j)} \quad (j \in s_i), \end{cases} \tag{2.3.2}$$

with $\mathbf{a}_i = \{\sum_{j \in U_i/s_i} \mathbf{x}_{ij}^\top - \sigma_v^2 (N_i - n_i) \mathbf{1}_{n_i}^\top \mathbf{V}_i^{-1}(\boldsymbol{\delta}) \mathbf{X}_i\} \{k^{-1} \sum_{i=1}^{k} \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\delta}) \mathbf{X}_i\}^{-1}$, and $\mathbf{C}_i(\boldsymbol{\delta}) = \mathbf{V}_i^{-1}(\boldsymbol{\delta})$ is a matrix satisfying $\mathbf{C}_i(\boldsymbol{\delta}) \equiv \mathbf{C}_i = (\mathbf{C}_i^{(1)}, \ldots, \mathbf{C}_i^{(n_i)})$, with $\mathbf{C}_i^{(j)}$ corresponding to the jth column of $\mathbf{C}_i$. The weights of sampled units are indexed by ihj, and the form of the weight is different depending if $j \in s_i$ or $j \in s_h$, $h \neq i$. Clearly the observations in the area i of interest should have more weight than the observations in s_h, $h \neq i$. It is easily seen that the weights $w_{ihj}(\hat{\boldsymbol{\delta}})$, $j \in s_h$, satisfy the calibration constraints

$$\sum_{h=1}^{k} \sum_{j \in s_h} w_{ihj}(\hat{\boldsymbol{\delta}}) \mathbf{x}_{hj}^\top = \sum_{j \in U_i} \mathbf{x}_{ij}^\top. \tag{2.3.3}$$

Let $\alpha \in \mathcal{R}^p$ be an arbitrary vector of dimension p and let $u_1, \ldots, u_k$ be arbitrary random variables. Using arguments similar to those of Chambers (1986) and (2.3.3), the empirical best linear unbiased predictor defined by (2.2.2) can be written as

$$\begin{aligned} \hat{\theta}_{i\text{EBLUP}} &= N_i^{-1} \sum_{j \in s_i} y_{ij} + N_i^{-1} \sum_{j \in U_i/s_i} (\mathbf{x}_{ij}^\top \boldsymbol{\alpha} + u_i) \\ &+ N_i^{-1} \sum_{j \in s_i} \{w_{ij}(\hat{\boldsymbol{\delta}}) - 1\} (y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\alpha} - u_i) \\ &+ N_i^{-1} \sum_{\substack{h=1 \\ h \neq i}}^{k} \sum_{j \in s_h} w_{ihj}(\hat{\boldsymbol{\delta}}) (y_{hj} - \mathbf{x}_{hj}^\top \boldsymbol{\alpha} - u_h) + N_i^{-1} \sum_{h=1}^{k} W_{ih}(\hat{\boldsymbol{\delta}}) u_h, \end{aligned} \tag{2.3.4}$$

where the weights $W_{ih}(\hat{\boldsymbol{\delta}})$ are

$$W_{ih}(\hat{\boldsymbol{\delta}}) = \begin{cases} \sum_{j\in s_i} w_{iij}(\hat{\boldsymbol{\delta}}) - N_i & h = i, \\ \sum_{j\in s_h} w_{ihj}(\hat{\boldsymbol{\delta}}) & h \neq i. \end{cases} \tag{2.3.5}$$

In (2.3.4), the vector $\boldsymbol{\alpha}$ is arbitrary. A natural candidate is $\hat{\boldsymbol{\beta}}_R$, solution of equations (2.2.8) and (2.2.9). Similarly, a suitable predictor for u_h is $\hat{v}_{hR}$, see Section 2.2. The empirical best linear unbiased predictor $\hat{\theta}_{i\text{EBLUP}}$ can be written as

$$\begin{aligned} \hat{\theta}_{i\text{EBLUP}} &= \hat{\theta}_{i\text{SR}} + N_i^{-1} \sum_{j\in s_i} \{w_{ij}(\hat{\boldsymbol{\delta}}) - 1\}(y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR}) \\ &\quad + N_i^{-1} \sum_{\substack{h=1 \\ h\neq i}}^{k} \sum_{j\in s_h} w_{ihj}(\hat{\boldsymbol{\delta}})(y_{hj} - \mathbf{x}_{hj}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{hR}) \\ &\quad + N_i^{-1} \sum_{h=1}^{k} W_{ih}(\hat{\boldsymbol{\delta}}) \hat{v}_{hR}, \end{aligned} \tag{2.3.6}$$

where $\hat{\theta}_{i\text{SR}}$ is given by (2.2.10). The last three terms of (2.3.6) can be viewed as correction terms for the bias, which may be affected by large residuals. A robust predictor of θ_i is obtained, robustifying the representation (2.3.6)

$$\begin{aligned} \hat{\theta}_{i\text{C}} &= \hat{\theta}_{i\text{SR}} + N_i^{-1} \sum_{j\in s_i} \psi_{c_1}\{(w_{ij}(\hat{\boldsymbol{\delta}}) - 1)(y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR})\} + \\ &\quad N_i^{-1} \sum_{\substack{h=1 \\ h\neq i}}^{k} \sum_{j\in s_h} \psi_{c_1}\{w_{ihj}(\hat{\boldsymbol{\delta}})(y_{hj} - \mathbf{x}_{hj}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{hR})\} \\ &\quad + N_i^{-1} \sum_{h=1}^{k} \psi_{c_2}\{W_{ih}(\hat{\boldsymbol{\delta}}) \hat{v}_{hR}\}. \end{aligned} \tag{2.3.7}$$

The robust predictor (2.3.7) is based on ψ functions, and we propose to use the Huber function given by (2.2.5), with appropriate choices of the tuning constants c_1 and c_2. When c_1 and c_2 converge toward zero, (2.3.7) tends towards the Sinha–Rao predictor, whereas it tends towards the empirical best linear unbiased predictor when the tuning constants converge to infinity. Unlike the robust predictor (2.2.11), the correction terms of $\hat{\theta}_{i\text{C}}$ depend on s, not only s_i, which explains why the proposed predictor is called fully bias-corrected. The tuning constants need to be specified. They should be relatively large in order to lead to small biases, but small enough to ensure robustness and small mean square errors. The second and third terms of (2.3.7) are based on weighted residuals. A natural choice for c_1 is based on a robust estimator of the scale σ_e, and a measure of the weight. Thus, $c_1 = q \times \text{med}(w_{ihj}) \times \hat{\sigma}_{eR}$ with some constant q seems reasonable, where

med$(\cdot)$ denotes the median. If the outliers are non-representative, small values of q would be appropriate, for example $q = 3$ or even smaller values. In our framework, the outliers are representative, and thus a larger q is expected. From the simulation experiments in Section 2.5, a value of q as large as $q = 9$ seemed to control the biases well, and the corresponding empirical mean square errors were small. Similarly, $c_2 = q \times \text{med}(W_{ih}) \times \hat{\sigma}_{vR}$ seems natural for some value of q. See Section 2.5.

2.3.2. Robust predictor of the area mean based on the conditional bias

In predicting a population mean or total using auxiliary information and linear models, Beaumont *et al.* (2013) developed a model-based predictor using the conditional bias of a unit and showed that their proposal is closely related to the approach of Chambers (1986). This suggests that the conditional bias may be useful in the present framework.

Complications occur in mixed linear models, because the observations in a given area are correlated, due to area-specific random effects. Following Beaumont *et al.* (2013), we define the conditional bias of unit j in area h as

$$B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta}) = E_m\{\hat{\theta}_i(\boldsymbol{\delta}) - \theta_i \mid s, y_{hj}, v_h\}. \tag{2.3.8}$$

Thus, in linear mixed models, we calculate the conditional expectations, keeping fixed a given unit and also the local effect associated with the area which contains that particular unit. Consequently, the conditional bias measures the average joint effect of unit y_{hj} and local area effect v_h, j in area h, on the predictor $\hat{\theta}_i$. We adopt Definition (2.3.8) because of its simplicity, and in view of the relations with Chambers' approach discussed in Section 2.3.1, as discussed below.

Several situations need to be considered, since a unit j may or may not be in the area of interest. Using the definition of the weights $w_{ihj}(\boldsymbol{\delta})$ in (2.3.2), we obtain

$$B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta}) = \begin{cases} N_i^{-1}(w_{iij} - 1)(y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - v_i) + N_i^{-1} W_{ii} v_i \quad (j \in s_i), \\ N_i^{-1} w_{ihj}(y_{hj} - \mathbf{x}_{hj}^\top \boldsymbol{\beta} - v_h) + N_i^{-1} W_{ih} v_h \quad (j \in s_h, h \neq i), \\ -N_i^{-1}(y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - v_i) + N_i^{-1} W_{ii} v_i \quad (j \in U_i/s_i), \\ N_i^{-1} W_{ih} v_h \quad (j \in U_h/s_h, h \neq i), \end{cases} \tag{2.3.9}$$

where $w_{ihj} \equiv w_{ihj}(\boldsymbol{\delta})$ and $W_{ih} \equiv W_{ih}(\boldsymbol{\delta})$. The result (2.3.9) suggests that a unit outside the area of interest may have a significant impact. In fact, a unit $j \in s_h$ may have a large influence if its weight $w_{ihj}(\boldsymbol{\delta})$ is large. Naturally, a large model

residual of a given unit is expected to have a large influence, and a large residual $y_{hj} - \mathbf{x}_{hj}^\top \boldsymbol{\beta} - v_h$ is associated to a large conditional bias for unit j. Finally, a large random effect v_h increases the conditional bias, and the effect is more pronounced if the associated weight W_{ih} is large. Non-sampled units may have large influences, but it is not possible to reduce their impact at the estimation stage.

From Expressions (2.2.1) and (2.3.9), the prediction error of the best linear unbiased predictor can be written as

$$\begin{aligned}\hat{\theta}_i(\boldsymbol{\delta}) - \theta_i &= \sum_{h=1}^{k} \sum_{j \in U_h} B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta}) \\ &\quad - N_i^{-1} \sum_{h=1}^{k} (N_h - 1) W_{ih}(\boldsymbol{\delta}) v_h. \end{aligned} \tag{2.3.10}$$

Expression (2.3.10) suggests that the conditional bias $B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta})$ attached to unit j and the random effect v_h can be interpreted as their contribution to the prediction error of $\hat{\theta}_i(\boldsymbol{\delta})$. Following the approach of Beaumont *et al.* (2013), we define the following robust pseudo-best linear unbiased predictor of θ_i :

$$\begin{aligned}\tilde{\theta}_i(\boldsymbol{\beta}, \boldsymbol{\delta}, v) &= \theta_i + \sum_{h=1}^{k} \sum_{j \in s_h} \Phi_{d_1,d_2}\{B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta})\} \\ &\quad + \sum_{h=1}^{k} \sum_{j \in U_h / s_h} B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta}) \\ &\quad - N_i^{-1} \sum_{h=1}^{k} (N_h - 1) \psi_{d_2}\{W_{ih}(\boldsymbol{\delta}) v_h\}, \end{aligned} \tag{2.3.11}$$

where $v = (v_1, \ldots, v_k)^\top$ corresponds to the vector of random effects, and we define robustified conditional biases as

$$\Phi_{d_1,d_2}\{B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta})\} = \begin{cases} N_i^{-1} \psi_{d_1}\{(w_{iij}(\boldsymbol{\delta}) - 1)(y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - v_i)\} \\ \qquad + N_i^{-1} \psi_{d_2}\{W_{ii}(\boldsymbol{\delta}) v_i\} \quad (j \in s_i), \\ N_i^{-1} \psi_{d_1}\{w_{ihj}(\boldsymbol{\delta})(y_{hj} - \mathbf{x}_{hj}^\top \boldsymbol{\beta} - v_h)\} \\ \qquad + N_i^{-1} \psi_{d_2}\{W_{ih}(\boldsymbol{\delta}) v_h\} \quad (j \in s_h, h \neq i), \end{cases} \tag{2.3.12}$$

with the ψ functions in the Huber class, with some choices of tuning constants d_1 and d_2. Let $B_{ihj} \equiv B_{ihj}(y_{hj}, v_h; \boldsymbol{\beta}, \boldsymbol{\delta})$. Using (2.3.10) and (2.3.11), it follows that

$$\begin{aligned}\tilde{\theta}_i(\boldsymbol{\beta}, \boldsymbol{\delta}, v) &= \{\hat{\theta}_i(\boldsymbol{\delta}) - \sum_{h=1}^{k} \sum_{j \in U_h} B_{ihj} + N_i^{-1} \sum_{h=1}^{k} (N_h - 1) W_{ih}(\boldsymbol{\delta}) v_h\} \\ &\quad + \sum_{h=1}^{k} \sum_{j \in s_h} \Phi_{d_1,d_2}(B_{ihj}) + \sum_{h=1}^{k} \sum_{j \in U_h / s_h} B_{ihj} \end{aligned}$$

$$-N_i^{-1}\sum_{h=1}^{k}(N_h-1)\psi_{d_2}\{W_{ih}(\boldsymbol{\delta})v_h\}. \tag{2.3.13}$$

The first term on the right-side of (2.3.13) equals θ_i ; thus, it is not affected by outliers. Interestingly, if $d_2=\infty$, simplifications occur, since the weighted means of random effects cancel :

$$\tilde{\theta}_i(\boldsymbol{\beta},\boldsymbol{\delta},v)=\hat{\theta}_i(\boldsymbol{\delta})-\sum_{h=1}^{k}\sum_{j\in s_h}B_{ihj}\left(y_{hj},v_h;\boldsymbol{\beta},\boldsymbol{\delta}\right)+\sum_{h=1}^{k}\sum_{j\in s_h}\Phi_{d_1,\infty}\left\{B_{ihj}\left(y_{hj},v_h;\boldsymbol{\beta},\boldsymbol{\delta}\right)\right\}. \tag{2.3.14}$$

The form of the pseudo-predictor (2.3.14) is identical to that proposed in Beaumont *et al.* (2013). In the following, we adopt $d_2=\infty$. From (2.3.14), when a sample unit j has a small conditional bias, we have $\Phi_{d_1,\infty}\{B_{ihj}(y_{hj},v_h;\boldsymbol{\beta},\boldsymbol{\delta})\}\approx B_{ihj}(y_{hj},v_h;\boldsymbol{\beta},\boldsymbol{\delta})$ and the contribution of the second and third terms on the right hand side of (2.3.14) is expected to be small. Thus, if the influences are small, the predictor (2.3.14) is close to the best linear unbiased predictor, the constant d_1 controlling the influences when they are significantly large.

The conditional biases $B_{ihj}(y_{hj},v_h;\boldsymbol{\beta},\boldsymbol{\delta})$ are unknown and they depend on the model parameters $(\boldsymbol{\beta}^\top,\boldsymbol{\delta}^\top)^\top$ and also on the random small area effects v_h. The vector of parameters for the fixed effects $\boldsymbol{\beta}$ and the random effects can be estimated with the methods of Sinha & Rao (2009), as described in Section 2.2. Concerning the estimation of the variance components, a possible choice is to estimate $\boldsymbol{\delta}$ by maximum likelihood ; recall that in our framework the outliers are legitimate observations, and it appears desirable to have a predictor not too far from the empirical best linear unbiased predictor. Let $\hat{B}_{ihj}=B_{ihj}(y_{hj},\hat{v}_{hR};\hat{\boldsymbol{\beta}}_R,\hat{\boldsymbol{\delta}})$ denote the estimator of $B_{ihj}(y_{hj},v_h;\boldsymbol{\beta},\boldsymbol{\delta})$ based on $\hat{\boldsymbol{\delta}}$ and the robust estimators $\hat{\boldsymbol{\beta}}_R$ and $\hat{v}_R=(\hat{v}_{1R},\ldots,\hat{v}_{kR})^\top$. The following predictor represents a compromise between the empirical best linear unbiased predictor and the Sinha–Rao predictor :

$$\tilde{\theta}_i(\hat{\boldsymbol{\beta}}_R,\hat{\boldsymbol{\delta}},\hat{v}_R)=\hat{\theta}_{i\mathrm{EBLUP}}-\sum_{h=1}^{k}\sum_{j\in s_h}\hat{B}_{ihj}+\sum_{h=1}^{k}\sum_{j\in s_h}\Phi_{d_1,\infty}(\hat{B}_{ihj}), \tag{2.3.15}$$

where $\hat{\theta}_i(\hat{\boldsymbol{\delta}})=\hat{\theta}_{i\mathrm{EBLUP}}$ in (2.3.15) is given by (2.2.2). When $d_1=\infty$, the robust predictor (2.3.15) reduces to the empirical best linear unbiased predictor, which is asymptotically unbiased under model (2.1.1) but suffers from a potentially large variance in the presence of outliers. On the other hand, when d_1 converges toward zero, then (2.3.15) is expected to be highly robust but substantially biased. Using the definitions of $\Phi_{d_1,\infty}(\hat{B}_{ihj})$ given by (2.3.9) and (2.3.12) respectively, the robust

predictor denoted by $\hat{\theta}_{i\text{CB}}$, simplifies to

$$\begin{aligned}\hat{\theta}_{i\text{CB}} &= \hat{\theta}_{i\text{SR}} + N_i^{-1}\sum_{j\in s_i}\psi_{d_1}\{(w_{iij}(\hat{\boldsymbol{\delta}})-1)(y_{ij}-\mathbf{x}_{ij}^{\top}\hat{\boldsymbol{\beta}}_R-\hat{v}_{iR})\}\\ &\quad +N_i^{-1}\sum_{\substack{h=1\\h\neq i}}^{k}\sum_{j\in s_h}\psi_{d_1}\{w_{ihj}(\hat{\boldsymbol{\delta}})(y_{hj}-\mathbf{x}_{hj}^{\top}\hat{\boldsymbol{\beta}}_R-\hat{v}_{hR})\}\\ &\quad +N_i^{-1}\sum_{h=1}^{k}W_{ih}(\hat{\boldsymbol{\delta}})\hat{v}_{hR}. \qquad (2.3.16)\end{aligned}$$

From (2.3.16), the robust predictor $\hat{\theta}_{i\text{CB}}$ is closely related to the Chambers' approach described in Section 2.3.1. In fact, as $\hat{\theta}_{i\text{C}}$ defined by (2.3.7), the robust predictor $\hat{\theta}_{i\text{CB}}$ is fully bias-corrected, with correction terms based on robustified residuals for $j \in \cup_{h=1}^{k} s_h$ and random effects predictors $\hat{v}_{hR}$. When the tuning constant $d_1 \to \infty$, then (2.3.16) converges to the empirical best linear unbiased predictor $\hat{\theta}_{i\text{EBLUP}}$, and when $d_1 \to 0$ the predictor $\hat{\theta}_{i\text{CB}}$ tends to the Sinha–Rao predictor plus an additional term corresponding to a weighted mean of robust random effects predictors. In fact, (2.3.7) and (2.3.16) differ in the treatment of $N_i^{-1}\sum_{h=1}^{k} W_{ih}(\hat{\boldsymbol{\delta}})\hat{v}_{hR}$, and large values of the weighted random effect $W_{ih}(\hat{\boldsymbol{\delta}})\hat{v}_{hR}$, can be downweighted under Chambers' approach. Note that the extra term comes from the empirical best linear unbiased predictor representation as the Sinha–Rao predictor plus correction terms, where Expression (2.3.6) is used to obtain an explicit expression of (2.3.15) in terms of robust residuals and random effects. The simulation results presented in Section 2.5 suggest that $\hat{\theta}_{i\text{C}}$ and $\hat{\theta}_{i\text{CB}}$ perform well, with $\hat{\theta}_{i\text{CB}}$ offering less variability across the domains than $\hat{\theta}_{i\text{C}}$, at least in our simulation experiments.

2.4. Asymptotic biases and estimation of the mean square prediction error

2.4.1. Asymptotic biases under the mixture of two linear mixed models

An asymptotic approach is adopted to find the limiting expected values of the prediction errors $\hat{\theta}_i - \theta_i$, when $\hat{\theta}_i$ is one of the robust predictors $\hat{\theta}_{i\text{SR}}$, $\hat{\theta}_{i\text{C}}$ and $\hat{\theta}_{i\text{CB}}$. Asymptotic properties of the robust estimators $\hat{\boldsymbol{\beta}}_R$ and $\hat{\boldsymbol{\delta}}_R$ have been studied in Sinha & Rao (2009), assuming that the populations are generated according to (2.1.1). The asymptotic arguments assumed that $k \to \infty$, with fixed values of n_i, which can be relatively small.

Let ζ_γ be a population from which the model for observations is given by

$$\zeta_\gamma : \ \mathbf{y}_{\gamma i} = \mathbf{X}_i \boldsymbol{\beta}_\gamma + \mathbf{v}_{\gamma i} \mathbf{1}_{n_i} + e_{\gamma i} \quad (i = 1, \ldots, k).$$

The underlying model for the values of Y, denoted by ζ_m, is assumed to be a mixture of ζ_0 and ζ_1, in the sense that the population values for Y in domain i are

$$y_{ij} = (1 - A_{ij}) y_{0ij} + A_{ij} y_{1ij} \quad (j \in U_i; i = 1, \ldots, k), \tag{2.4.1}$$

where A_{ij} are independent realizations of Bernoulli random variables with parameter $p = \text{pr}(A_{ij} = 1)$. The model (2.1.1) is fitted. Let $\hat{\lambda}_m = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\delta}}^\top)^\top$ be the estimator under the working model. For example, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ can be the empirical best linear unbiased estimator of $\boldsymbol{\beta}$ and the maximum likelihood estimator of $\boldsymbol{\delta}$ assuming normality. The robust estimator $\hat{\boldsymbol{\lambda}}_R = (\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\delta}}_R^\top)^\top$, solution of the robustified maximum likelihood equations (2.2.8) and (2.2.9), is calculated. Under standard regularity conditions, a first-order Taylor series expansion gives the following asymptotic expression for the bias under model ζ_m :

$$\begin{aligned}
\text{AE}_m(\hat{\theta}_{i\text{C}} - \theta_i) \ = \ & -N_i^{-1} \sum_{j \in U_i / s_i} \left[\mathbf{x}_{ij}^\top (\boldsymbol{\beta}_m - \boldsymbol{\beta}_R) - E_m\{\hat{v}_{iR}(\boldsymbol{\lambda}_R)\} \right] \\
& + N_i^{-1} \sum_{j \in s_i} E_m \left(\psi_{c_1} \left[\{ w_{iij}(\boldsymbol{\delta}_m) - 1 \} \{ y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta}_R - \hat{v}_{iR}(\boldsymbol{\lambda}_R) \} \right] \right) \\
& + N_i^{-1} \lim_{k \to \infty} \sum_{h \neq i} \sum_{j \in s_h} E_m \left(\psi_{c_1} \left[w_{ihj}(\boldsymbol{\delta}_m) \{ y_{hj} - \mathbf{x}_{hj}^\top \boldsymbol{\beta}_R - \hat{v}_{hR}(\boldsymbol{\lambda}_R) \} \right] \right) \\
& + N_i^{-1} \lim_{k \to \infty} \sum_{h=1}^{k} E_m [\psi_{c_2} \{ W_{ih}(\boldsymbol{\delta}_m) \hat{v}_{hR}(\boldsymbol{\lambda}_R) \}],
\end{aligned} \tag{2.4.2}$$

where AE_m stands for the asymptotic expectation as $k \to \infty$ under the model ζ_m, the vector $\boldsymbol{\lambda}_R = (\boldsymbol{\beta}_R^\top, \boldsymbol{\delta}_R^\top)^\top$ is the probability limit of $\hat{\boldsymbol{\lambda}}_R$, and $\boldsymbol{\lambda}_m = (\boldsymbol{\beta}_m^\top, \boldsymbol{\delta}_m^\top)^\top$ is the probability limit of $\hat{\boldsymbol{\lambda}}_m$, with $\boldsymbol{\beta}_m = (1-p)\boldsymbol{\beta}_0 + p\boldsymbol{\beta}_1$. The main lines of the proof are given in the Supplementary material. When the tuning constant $c_1 = d_1$ and if c_2 converges to infinity, we obtain the asymptotic bias of the predictor (2.3.16). The asymptotic bias of the Sinha–Rao predictor is obtained by letting c_1 and c_2 converge to zero in (2.4.2) :

$$\text{AE}_m \left(\hat{\theta}_{i\text{SR}} - \theta_i \right) = \left(1 - n_i N_i^{-1} \right) \left[\bar{\mathbf{X}}_{\bar{s}_i}^\top \left(\boldsymbol{\beta}_R - \boldsymbol{\beta}_m \right) + E_m\{\hat{v}_{iR}(\boldsymbol{\lambda}_R)\} \right], \tag{2.4.3}$$

where $\bar{\mathbf{X}}_{i\bar{s}_i} = (N_i - n_i)^{-1} \sum_{j \in U_i / s_i} \mathbf{x}_{ij}$. The additional terms in (2.4.2) represent the correction terms for the bias. When $c_1, c_2 \to \infty$ in (2.4.2) the calibration constraints (2.3.3) show that the asymptotic biases are zero, as excepted. Choosing small values of c_1 and c_2 introduces some bias, but the gains in mean squared

error are typically significant, see Section 2.5. The asymptotic bias may be appreciable for the Sinha–Rao method if the limit of the robust estimator $\boldsymbol{\beta}_R$ is different from the parameter $\boldsymbol{\beta}_m$ of the mixture model. In model (2.4.1) with different slopes and no contamination in the random effects, the main term in (2.4.3) is $\boldsymbol{\beta}_m - \boldsymbol{\beta}_R$. In the limit case $\sigma_v^2 = 0$, the bias depends only on $\boldsymbol{\beta}_m - \boldsymbol{\beta}_R$. On the other hand, when the fixed effects in the mixture model have the same slopes, that is $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1$, and if the random effects v_h represent mixtures such as the contaminated normal distributions $(1-p)\mathcal{N}(0, \sigma_{v0}^2) + p\mathcal{N}(0, \sigma_{v1}^2)$, the biases of the Sinha–Rao predictor are also expected to be small. In the latter situation, the contaminations affect the variance, not the random effects means, and both terms in (2.4.3) will be small. Incidently, this scenario has been studied in Sinha & Rao (2009), who found good behavior for their robust predictor. Expression (2.4.3) suggests that the mixture model with $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_1$ or contamination in the random effects with non-null mean may create non-negligible biases for the Sinha–Rao predictor.

2.4.2. Estimation of the mean square prediction error

Constructing confidence interval for the robust predictors is important but challenging. Sinha & Rao (2009) proposed a parametric bootstrap procedure to estimate the mean square prediction error but it gave poor coverage rates in our empirical results for mixture models, because the populations were generated using the robust estimators, which do not reflect all the units. In our framework, all the observations are legitimate and using robust estimators reflected only the main component of the mixtures resulting in large biases. Let $\hat{\boldsymbol{\beta}}_R$ be the robust estimator of $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\delta}}$ be the non-robust estimator of $\boldsymbol{\delta}$. For generating the bootstrap populations, a reasonable solution relies on using the robust estimators for the slope parameters, and $\hat{\boldsymbol{\delta}}$ for estimating $\boldsymbol{\delta}$, which rely on all the legitimate observations defining the population. This comes from the form of the predictors (2.3.6) and (2.3.15), which are functions of the estimator $\hat{\boldsymbol{\delta}}$ and also of the robust residuals. We present the method for $\hat{\theta}_{iC}$, but it can be easily adapted for other predictors.

1. The random variables v_i^* et e_{ij}^* are generated according to $\mathcal{N}(0, \hat{\sigma}_v^2)$ and $\mathcal{N}(0, \hat{\sigma}_e^2)$, respectively, to create bootstrap samples from the model

$$y_{ij}^* = \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R + v_i^* + e_{ij}^*. \tag{2.4.4}$$

2. The bootstrap «parameter» θ_i is calculated, and is denoted as θ_i^*. Let $\mathcal{E}_i^*$ be the mean error of the non-sampled units from normal distributions

$\mathcal{N}\left(0,(N_i-n_i)^{-1}\hat{\sigma}_e^2\right)$. The bootstrap mean θ_i^* is defined as

$$\theta_i^* = N_i^{-1}\sum_{j\in s_i} y_{ij}^* + N_i^{-1}\sum_{j\in U_i/s_i}\mathbf{x}_{ij}^\top\hat{\boldsymbol{\beta}}_R + \left(1-n_iN_i^{-1}\right)v_i^* + \left(1-n_iN_i^{-1}\right)\mathcal{E}_i^*. \quad (2.4.5)$$

3. From the bootstrap samples, robust bootstrap estimators $\hat{\boldsymbol{\beta}}_R^*$, $\hat{\boldsymbol{\delta}}_R^*$, $\hat{v}_{hR}^*$ and estimators $\hat{\boldsymbol{\beta}}^*$, $\hat{\boldsymbol{\delta}}^*$, $\hat{v}_h^*$ of $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, v_h are calculated. The parametric bootstrap estimator of the small area mean is

$$\begin{aligned}\hat{\theta}_{i\mathrm{C}}^* &= \hat{\theta}_{i\mathrm{SR}}^* + N_i^{-1}\sum_{j\in s_i}\psi_{c_1}\{(w_{ij}(\hat{\boldsymbol{\delta}}^*)-1)(y_{ij}-\mathbf{x}_{ij}^\top\hat{\boldsymbol{\beta}}_R^*-\hat{v}_{iR}^*)\}+\\ &\quad N_i^{-1}\sum_{\substack{h=1\\h\neq i}}^{k}\sum_{j\in s_h}\psi_{c_1}\{w_{ihj}(\hat{\boldsymbol{\delta}}^*)(y_{hj}-\mathbf{x}_{hj}^\top\hat{\boldsymbol{\beta}}_R^*-\hat{v}_{hR}^*)\}\\ &\quad +N_i^{-1}\sum_{h=1}^{k}\psi_{c_2}\{W_{ih}(\hat{\boldsymbol{\delta}}^*)\hat{v}_{hR}^*\}, \end{aligned} \quad (2.4.6)$$

where $\hat{\theta}_{i\mathrm{SR}}^* = N_i^{-1}\sum_{j\in s_i} y_{ij}^* + N_i^{-1}\sum_{j\in U_i/s_i}\mathbf{x}_{ij}^\top\hat{\boldsymbol{\beta}}_R^* + \left(1-n_iN_i^{-1}\right)\hat{v}_{iR}^*$.

4. Based on B bootstrap samples, the estimator of the mean square prediction error of $\hat{\theta}_{i\mathrm{C}}$ is

$$\mathrm{MSE}(\hat{\theta}_{i\mathrm{C}}) = B^{-1}\sum_{b=1}^{B}\left(\hat{\theta}_{i\mathrm{C}}^{*(b)}-\theta_i^{*(b)}\right)^2,$$

where $\theta_i^{*(b)}$ and $\hat{\theta}_{i\mathrm{C}}^{*(b)}$ correspond to (2.4.5) and (2.4.6), respectively, for the bth bootstrap sample.

Depending on the nature of the outliers, the proposed bootstrap method is expected to work reasonably well. Examples include outliers in the random effects and outliers in the error term. When the slope parameters are different in the mixture, the proposed bootstrap method is expected to be biased. From the results in Section 2.5, the performance of the proposed method were found encouraging under all the scenarios, at least in our experiments, see Section 2.5.

2.5. Monte Carlo experiments

2.5.1. Description of the populations

We conducted a Monte Carlo study in order to study the empirical biases and mean square errors of the new robust predictors $\hat{\theta}_{i\mathrm{C}}$ and $\hat{\theta}_{i\mathrm{CB}}$, and to compare them to the proposals of Sinha & Rao (2009) and Chambers *et al.* (2014). We also study the performance of the bootstrap method. In view of the results in Section 2.4, to compare empirically the biases of $\hat{\theta}_{i\mathrm{SR}}$ with those of the new

TABLE 2.1. Description of the three scenarios. The populations were generated according to $y_{ij} = (1 - A_{ij})y_{0ij} + A_{ij}y_{1ij}$, $A_{ij} \sim Bernoulli(0.1)$, using the unit-level models (2.5.1) and (2.5.2), assuming normality for the random effects and error terms in ζ_0 and ζ_1. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain equals 0.5.

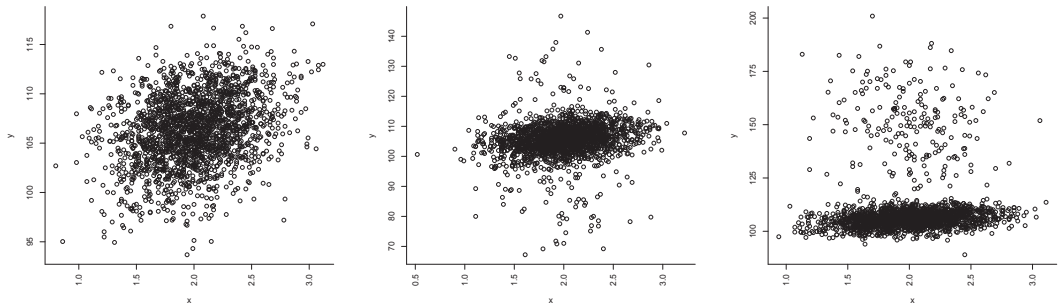
	Sources of the contamination		
Scenarios	Variances of error terms	Variances of random effects	Intercepts and slopes
$(0, 0, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 6)$	$\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1 = (100, 3)^\top$
$(e, v, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 150)$	$\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1 = (100, 3)^\top$
(e, v, b)	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 150)$	$\boldsymbol{\beta}_0 = (100, 3)^\top, \boldsymbol{\beta}_1 = (150, 1)^\top$

methods seems particularly relevant. We considered the mixture model ζ_m of two unit-level models :

$$\zeta_0 : y_{0ij} = \beta_{00} + \beta_{01}x_{ij} + v_{0i} + e_{0ij} \quad (j = 1, \ldots, N_i; i = 1, \ldots, k), \quad (2.5.1)$$

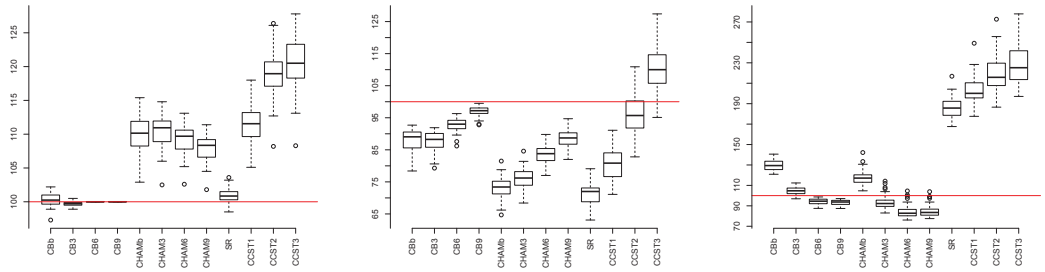
$$\zeta_1 : y_{1ij} = \beta_{10} + \beta_{11}x_{ij} + v_{1i} + e_{1ij} \quad (j = 1, \ldots, N_i; i = 1, \ldots, k). \quad (2.5.2)$$

We considered $k = 40$ and $N_1 = \ldots = N_{40} = 50$. Normal distributions for the random effects and the error terms were assumed. Thus, $v_{0i} \sim \mathcal{N}(0, \sigma^2_{v0})$, $v_{1i} \sim \mathcal{N}(0, \sigma^2_{v1})$, $e_{0ij} \sim \mathcal{N}(0, \sigma^2_{e0})$ and $e_{1ij} \sim \mathcal{N}(0, \sigma^2_{e1})$, $(k = 1, \ldots, 40; j = 1, \ldots, 50)$. The values of the auxiliary information were generated from a normal distribution such that $E(X) = 2$ and $V^{1/2}(X) = 0.35$. The mixture model ζ_m satisfied $y_{ij} = (1 - A_{ij})y_{0ij} + A_{ij}y_{1ij}$, where the A_{ij}'s were independently generated according to a Bernoulli distribution, $A_{ij} \sim Bernoulli(p)$, with $p = 0.1$. In each area of a



(a) (0,0,0) (b) (e,v,0) (c) (e,v,b)

FIGURE 2.1. Populations generated according to the mixture model (2.5.1) and (2.5.2). The model parameters are given in Table 1. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to 0.5.



(a) (0,0,0) (b) (e,v,0) (c) (e,v,b)

FIGURE 2.2. Boxplots of the relative efficiencies for the predictors defined in Table 2.2. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain equals 0.5.

given population, random samples of size $n_1 = \ldots = n_{40} = 5$ have been selected by simple random sampling without replacement.

Contaminations in the error terms and random effects have been investigated in the simulation experiments of Sinha & Rao (2009). More precisely, the random effects were generated according to the contaminated normal distribution $(1-p)\mathcal{N}(0, \sigma^2_{v0}) + p\mathcal{N}(0, \sigma^2_{v1})$. Similarly, the error terms were obtained from $(1-p)\mathcal{N}(0, \sigma^2_{e0}) + p\mathcal{N}(0, \sigma^2_{e1})$. In view of the asymptotic biases derived in Section 2.4, it is particularly relevant to investigate situations where the intercepts and slopes are different. We considered the same scenarios as in Sinha & Rao (2009), but added two more scenarios with $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_1$, where $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^\top$ and $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11})^\top$. Adopting a notation similar to that of Sinha & Rao (2009), three scenarios were studied, see Table 2.1.

From Table 2.1, scenario $(0, 0, 0)$ corresponds to the absence of contamination. For the parameters in Table 2.1 the correlation between the units of a given area is set to $\rho_0 = 0.5$ in the absence of contamination. Note that ρ_0 satisfies the relation $\sigma^2_{v0} = \rho_0\sigma^2_{e0}/(1-\rho_0)$. We also considered the case $\rho_0 = 0.05$ in the Supplementary material. Under $(0, 0, b)$, the random effects and the error terms are not contaminated, but the model parameters in (2.5.1) and (2.5.2) are different. Under $(0, v, 0)$, only the area random effect is contaminated. Another example is the case (e, v, b), where the contamination comes from the random effects, the random errors and the fixed effects. The other scenarios are interpreted in the same manner. See Figure 2.1.

TABLE 2.2. Monte Carlo absolute relative biases (ARB in percentage) and relative efficiencies (RE in percentage) for the robust predictors and the empirical bets linear unbiased predictor of the small area means (averaged over areas). The robust predictor $\hat{\theta}_{i\mathrm{C}}$ is denoted Cb, C3, C6 and C9, when the tuning constants are set to $q = b, 3, 6, 9$, respectively. Similarly, $\hat{\theta}_{i\mathrm{CB}}$ is represented by CBb, CB3, CB6 and CB9. The Sinha–Rao predictor is noted SR, and the predictor $\hat{\theta}_{i\mathrm{CCST}}$ is noted CCST1, CCST2 and CCST3, when the tuning constants are $c = 1, 2, 3$, respectively. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain equals 0.5.

	$(0,0,0)$		$(e,v,0)$		(e,v,b)	
	ARB	RE	ARB	RE	ARB	RE
EBLUP	0.025	100.0	0.047	100.0	0.071	100.0
CBb	0.025	100.3	0.042	88.1	1.803	129.7
CB3	0.024	99.7	0.042	87.8	1.076	104.7
CB6	0.025	100.0	0.045	92.8	0.545	94.2
CB9	0.025	100.0	0.047	97.0	0.281	93.3
CHAMb	0.027	110.1	0.034	73.2	1.844	117.9
CHAM3	0.027	110.5	0.036	76.0	1.112	93.6
CHAM6	0.027	109.3	0.040	83.4	0.576	84.6
CHAM9	0.027	108.0	0.042	88.4	0.307	85.2
SR	0.025	101.0	0.035	71.2	3.140	186.4
CCST1	0.027	111.5	0.038	80.9	3.055	203.6
CCST2	0.030	118.9	0.040	96.5	2.796	219.2
CCST3	0.031	120.7	0.042	110.1	2.524	227.4

2.5.2. Predictors used in the study and empirical measures

Five predictors of the small area mean θ_i were included in the study. The empirical best linear unbiased predictor defined by (2.2.2), based on the empirical best linear unbiased estimators and empirical best linear unbiased predictors, with variance components estimated by maximum likelihood, was used as a benchmark. As in Sinha & Rao (2009), the robust predictors relied on the robustified maximum likelihood estimators of $(\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top)^\top$, which are obtained by solving (2.2.8) and (2.2.9). The robust random effects were estimated by solving Fellner's equation (2.2.4). We used $b = 1.345$. Based on these estimators, the following robust predictors were considered : the Sinha–Rao predictor $\hat{\theta}_{i\mathrm{SR}}$ given by (2.2.10) ; the new robust procedures $\hat{\theta}_{i\mathrm{C}}$ and $\hat{\theta}_{i\mathrm{CB}}$ defined by (2.3.7) and (2.3.16), respectively ; and the predictor (2.2.11), denoted by $\hat{\theta}_{i\mathrm{CCST}}$. For $\hat{\theta}_{i\mathrm{C}}$, the tuning constants were set to $c_1 = q \times \mathrm{med}\{w_{ihj}(\hat{\boldsymbol{\delta}})\} \times \hat{\sigma}_{eR}$ and $c_2 = q \times \mathrm{med}\{W_{ih}(\hat{\boldsymbol{\delta}})\} \times \hat{\sigma}_{vR}$, where $(\hat{\sigma}^2_{eR}, \hat{\sigma}^2_{vR})^\top$ denote the robust estimators of the variance components based on

the robustified maximum likelihood method. We considered $d_1 = c_1$ in $\hat{\theta}_{i\mathrm{CB}}$. In Section 2.5.3, simulation results are presented for all the domains using boxplots, and we considered in these cases $q = 3, 6, 9$, for the predictors $\hat{\theta}_{i\mathrm{C}}$ and $\hat{\theta}_{i\mathrm{CB}}$. The tuning constant of the robust predictor $\hat{\theta}_{i\mathrm{CCST}}$ was set to $c = 1, 2, 3$, inspired by the simulation experiments of Chambers *et al.* (2014).

For each scenario described in Table 2.1, 1,000 populations were generated. Let $\hat{\theta}_i^{(b)}$ denote a predictor for domain i at iteration b. The empirical percent absolute relative bias for the area mean θ_i was calculated as

$$\mathrm{ARB}(\hat{\theta}_i) = \left| B^{-1} \sum_{b=1}^{B} \frac{(\hat{\theta}_i^{(b)} - \theta_i^{(b)})}{\theta_i^{(b)}} \right| \times 100 \quad (i = 1, \ldots, k).$$

Using the empirical best linear unbiased predictor $\hat{\theta}_{i\mathrm{EBLUP}}$ as the reference, we calculated the relative efficiency of $\hat{\theta}_i$ in percentage $\hat{\theta}_{i\mathrm{EBLUP}}$, we used

$$\mathrm{RE}(\hat{\theta}_i) = \frac{\mathrm{MSE}(\hat{\theta}_i)}{\mathrm{MSE}(\hat{\theta}_{i\mathrm{EBLUP}})} \times 100, \quad \mathrm{MSE}(\hat{\theta}_i) = B^{-1} \sum_{b=1}^{B} \left(\hat{\theta}_i^{(b)} - \theta_i^{(b)}\right)^2 \quad (i = 1, \ldots, k),$$

with $\mathrm{MSE}(\hat{\theta}_i)$ denoting the empirical mean squared error of $\hat{\theta}_i$. Section 2.5.3 presents simulation results based on all the domains, using boxplots and integrated measures such as averages computed over all the areas.

We also present boxplots for 95% confidence intervals using the bootstrap method. We used 500 populations and $B = 200$ bootstrap replications. In addition to the empirical best linear unbiased predictor and the proposed methods, we considered $\hat{\theta}_{i\mathrm{CCST}}$ with $c = 1, 2$, which offered the best efficiencies. Simulation results are reported under scenarios $(0, v, 0)$, $(e, v, 0)$ and (e, v, b).

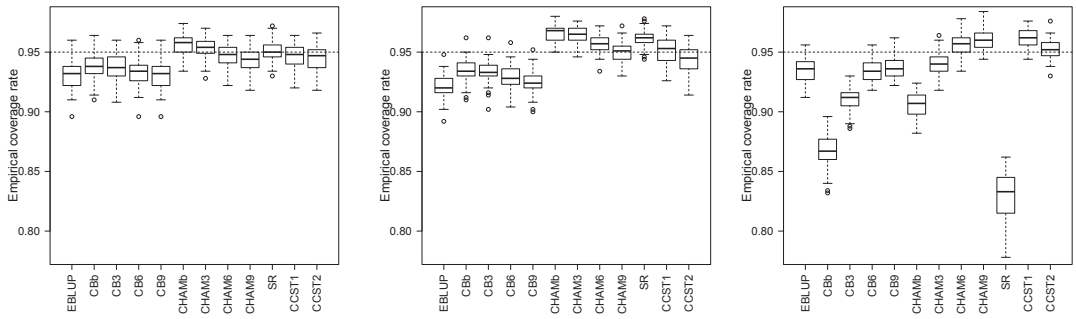
2.5.3. Results for all the domains

In Table 2.2, integrated measures of the empirical biases are given, where the averages over the areas were computed. Boxplots displaying empirical efficiency for all the domains are presented in Figure 2.2, for all the scenarios described in Table 2.1. The relative efficiencies were computed for each domain in the Monte Carlo studies; boxplots of the $k = 40$ relative empirical efficiencies with respect to the empirical best linear unbiased predictor were computed for each domain.

Table 2.2 shows that all the methods exhibited small empirical biases for the scenarios $(0, 0, 0)$, and $(e, v, 0)$. Larger biases were observed under scenario (e, v, b) for the robust methods as a function of the tuning constants. Small values of the tuning constants generated larger empirical biases. The empirical biases decreased as the tuning constants increased, as expected. From Table 2.2, all the empirical biases were small for these scenarios. From Figure 2.2, the new robust methods

were significantly more efficient than the robust predictors $\hat{\theta}_{i\mathrm{SR}}$ or $\hat{\theta}_{i\mathrm{CCST}}$, when the population was generated according to the mixture model.

Under scenario $(0, 0, 0)$, the robust predictors $\hat{\theta}_{i\mathrm{CB}}$ and $\hat{\theta}_{i\mathrm{SR}}$ appeared as efficient as the empirical best linear unbiased predictor, whereas $\hat{\theta}_{i\mathrm{C}}$ was slightly less efficient. The robust predictor $\hat{\theta}_{i\mathrm{CCST}}$ was approximately unbiased but the variance part of the mean squared error was important. From Figure 2.2, the integrated mean square errors were 10-20% larger than the empirical best linear unbiased predictor. Under scenario (e, v, b), small differences in efficiency were observed between $\hat{\theta}_{i\mathrm{C}}$ and $\hat{\theta}_{i\mathrm{CB}}$, for a given value of the tuning constant. The predictors $\hat{\theta}_{i\mathrm{SR}}$ and $\hat{\theta}_{i\mathrm{CCST}}$ showed large mean square errors; for these predictors, the bias part of the mean squared error was non negligible. Efficiency gains were possible for several areas, and the medians of the boxplots suggest that the proposed robust methods were more efficient. Large gains in efficiency were observed under the scenario $(e, v, 0)$. As expected, the efficiencies decreased as the tuning constant increased. Taking $q = 3$ or $q = 6$ seemed to give robustness and efficiency. From Table 2.2, possible reductions in integrated mean square errors of approximately 10%-20% were observed for these values of the tuning constants.



(a) (0,v,0) (b) (e,v,0) (c) (e,v,b)

FIGURE 2.3. Empirical coverage rates under the scenarios $(0, v, 0)$, $(e, v, 0)$ and (e, v, b). The nominal coverage rate is 95%. Predictors are defined in Table 2.2. Under the scenario (0,0,0), the correlation between the units of the same domain equals 0.5.

Figure 2.3 presents the empirical coverage rates of the bootstrap confidence intervals. Due to the nature of the mixture models, the empirical coverage rates were reasonably close to the nominal coverage rates under $(0, v, 0)$ and $(e, v, 0)$. When the slopes were different, the bootstrap method worked reasonably well for the new methods with moderate tuning constants. The results for $\hat{\theta}_{i\mathrm{CCST}}$ were reasonable. The biases were large for the Sinha–Rao method and the coverage rates were far from the nominal levels. Under $(0, v, 0)$, the biases of the mean

squared error estimators were reasonably small (under 10%) for the proposed methods with moderate values of the tuning constants. These biases were small under $(e, v, 0)$ for the predictors based on the concept of conditional bias but large and positive using Chambers' method (around 10-20%). Under (e, v, b), the biases of the mean squared error estimators were positive for the new predictors with moderate values of the tuning constants, but large and negative for the Sinha–Rao method.

Acknowledgement

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada. The first author received a scholarship from the National Program on Complex Data Structures by participating in an internship program at Statistics Canada. The authors would like to thank the Editor, an Associate Editor and two referees for constructive suggestions. We also thank Professor J.N.K. Rao for constructive comments on an earlier draft of this paper.

2.6. Supplementary material

Supplementary material available at *Biometrika* online includes an outline of the proof of the asymptotic expression of the bias and additional simulation experiments.

2.6.1. Main lines of the proof of the asymptotic expression of the bias

To derive expression (2.4.2), the following regularity conditions are needed.
Assumption A.1 The functions $\hat{v}_{R1} = \hat{v}_{R1}(\boldsymbol{\beta}, \boldsymbol{\delta}), \ldots, \hat{v}_{Rk} = \hat{v}_{Rk}(\boldsymbol{\beta}, \boldsymbol{\delta})$, which are the solutions of the Fellner equation (2.2.4), are assumed to be differentiable with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.
Assumption A.2 Under the mixture model ζ_m, the estimators of the fixed effects parameters satisfy

$$k^{1/2}(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_R) = O_P(1),\ k^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_m) = O_P(1).$$

Assumption A.3 Under ζ_m, the estimators of the variance components satisfy

$$k^{1/2}(\hat{\boldsymbol{\delta}}_R - \boldsymbol{\delta}_R) = O_P(1),\ k^{1/2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_m) = O_P(1)$$

for some finite vector $\boldsymbol{\delta}_m$.
Assumption A.4 Let

$$\frac{\partial \hat{v}_{hR}}{\partial \hat{\boldsymbol{\lambda}}_R}(\boldsymbol{\lambda}_R) = \left.\frac{\partial \hat{v}_{hR}}{\partial \hat{\boldsymbol{\lambda}}_R}\right|_{\hat{\boldsymbol{\lambda}}_R = \boldsymbol{\lambda}_R},$$

where $\hat{\boldsymbol{\lambda}}_R = (\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\delta}}_R^\top)^\top$ and $\boldsymbol{\lambda}_R = (\boldsymbol{\beta}_R^\top, \boldsymbol{\delta}_R^\top)^\top$. We assume that

$$\lim_{k\to\infty} \sup_h E_m \left\{ \left|\left| \frac{\partial \hat{v}_{hR}}{\partial \hat{\boldsymbol{\lambda}}_R}(\boldsymbol{\lambda}_R) \right|\right| \right\} < \infty,$$

where $||\cdot||$ is the usual Euclidian norm.
Assumption A.5 The weights and the auxiliary information satisfy the following conditions

$$\lim_{k\to\infty} E_m \left\{ N_i^{-1} \sum_{h\neq i} \sum_{j\in s_h} |w_{ihj}(\boldsymbol{\delta}_m)|||\mathbf{x}_{hj}|| \right\} < \infty,\ \lim_{k\to\infty} E_m \left\{ N_i^{-1} \sum_{h=1}^{k} |W_{ih}(\boldsymbol{\delta}_m)| \right\} < \infty,$$

$$\lim_{k\to\infty} \left\{ N^{-1} \sum_{h=1}^{k} \sum_{j\in s_h} \left| \frac{\partial w_{ihj}(\boldsymbol{\delta}_m)}{\partial \hat{\boldsymbol{\delta}}_m} \right| ||\mathbf{x}_{hj}|| \right\} < \infty.$$

Under Assumption A.1, $\hat{v}_{hR}$ is assumed to be differentiable as a function of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$. A similar assumption has been used in Fellner (1986, Section 3). Assumptions A.2 and A.3 state that the robust estimators and the non robust estimators converge to some finite limits with convergence rate of $k^{1/2}$ under the mixture

model ζ_m. When the populations are generated according to ζ_0, Sinha & Rao (2009) established the asymptotic normality of the robust estimators. Assumption A.4 states that the expected derivatives should be bounded in probability uniformly as $k \to \infty$. Note that with our assumptions the ψ function is bounded. Finally, Assumption A.5 states that the average of the weights under the model ζ_m is bounded.

Under Assumptions A.1-A.5, a first-order Taylor series expansion allows us to derive the asymptotic biases of the predictors. In formula (2.4.2), it is stated for $\hat{\theta}_{i\mathrm{C}}$. The predictor is written as

$$\begin{aligned}\hat{\theta}_{i\mathrm{C}}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R) &= \hat{\theta}_{i\mathrm{SR}}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R) + N_i^{-1} \sum_{j \in s_i} \psi_{c_1}\left\{\alpha_{iij}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)\right\} + \\ &\qquad N_i^{-1} \sum_{h \neq i} \sum_{j \in s_h} \psi_{c_2}\left\{\alpha_{ihj}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)\right\} + N_i^{-1} \sum_{h=1}^{k} \psi_{c_2}\left\{\alpha_{ih}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)\right\},\end{aligned}$$

where $\alpha_{iij}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R) = \left\{w_{iij}(\hat{\boldsymbol{\delta}}) - 1\right\} \hat{\epsilon}_{ij}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)$, $\alpha_{ihj}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R) = w_{ihj}(\hat{\boldsymbol{\delta}}) \hat{\epsilon}_{hj}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)$ and $\alpha_{ih}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R) = W_{ih}(\hat{\boldsymbol{\delta}}) \hat{v}_{hR}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)$, with $\hat{\epsilon}_{ij}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R) = y_{ij} - \mathbf{x}_{ij}^{\top} \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)$, $j \in s_i$. A first-order Taylor series expansion of the predictor $\hat{\theta}_{i\mathrm{SR}}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)$ around the point $(\boldsymbol{\beta}_R, \boldsymbol{\delta}_R)$ gives

$$\begin{aligned}\hat{\theta}_{i\mathrm{SR}}(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R) &= \hat{\theta}_{i\mathrm{SR}}(\boldsymbol{\beta}_R, \boldsymbol{\delta}_R) + O_P(k^{-1/2}), \\ &= N_i^{-1} \left\{\sum_{j \in s_i} y_{ij} + \sum_{j \in U_i / s_i} \mathbf{x}_{ij}^{\top} \boldsymbol{\beta}_R\right\} + \left(1 - n_i N_i^{-1}\right) \hat{v}_{iR}(\boldsymbol{\lambda}_R) + O_P(k^{-1/2}).\end{aligned}$$

Similarly,

$$\begin{aligned} N_i^{-1} \sum_{j \in s_i} \psi_{c_1}\left\{\alpha_{iij}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)\right\} &= N_i^{-1} \sum_{j \in s_i} \psi_{c_1}\left\{\alpha_{iij}(\boldsymbol{\delta}_m, \boldsymbol{\beta}_R, \boldsymbol{\delta}_R)\right\} + O_P(k^{-1/2}), \\ N_i^{-1} \sum_{h \neq i} \sum_{j \in s_h} \psi_{c_2}\left\{\alpha_{ihj}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)\right\} &= N_i^{-1} \sum_{h \neq i} \sum_{j \in s_h} \psi_{c_2}\left\{\alpha_{ihj}(\boldsymbol{\delta}_m, \boldsymbol{\beta}_R, \boldsymbol{\delta}_R)\right\} + O_P(k^{-1/2}), \\ N_i^{-1} \sum_{h=1}^{k} \psi_{c_2}\left\{\alpha_{ih}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\delta}}_R)\right\} &= N_i^{-1} \sum_{h=1}^{k} \psi_{c_2}\left\{\alpha_{ih}(\boldsymbol{\delta}_m, \boldsymbol{\beta}_R, \boldsymbol{\delta}_R)\right\} + O_P(k^{-1/2}).\end{aligned}$$

Thus

$$\begin{aligned}\hat{\theta}_{i\mathrm{C}} - \theta_i &= -N_i^{-1} \sum_{j \in U_i / s_i} \hat{\epsilon}_{ij}(\boldsymbol{\beta}_R, \boldsymbol{\delta}_R) + N_i^{-1} \sum_{j \in s_i} \psi_{c_1}\left\{\alpha_{iij}(\boldsymbol{\delta}_m, \boldsymbol{\beta}_R, \boldsymbol{\delta}_R)\right\} + \\ &\qquad N_i^{-1} \sum_{h \neq i} \sum_{j \in s_h} \psi_{c_2}\left\{\alpha_{ihj}(\boldsymbol{\delta}_m, \boldsymbol{\beta}_R, \boldsymbol{\delta}_R)\right\} \\ &\qquad + N_i^{-1} \sum_{h=1}^{k} \psi_{c_2}\left\{\alpha_{3j}(\boldsymbol{\delta}_m, \boldsymbol{\beta}_R, \boldsymbol{\delta}_R)\right\} + O_P(k^{-1/2}).\end{aligned}$$

Under model ζ_m the expectation of y_{hj} is $E_m(y_{hj}) = \mathbf{x}_{hj}^\top \boldsymbol{\beta}_m$. This shows that the asymptotic bias is given by 2.4.2 by taking expectations as $k \to \infty$.

2.6.2. Simulation results

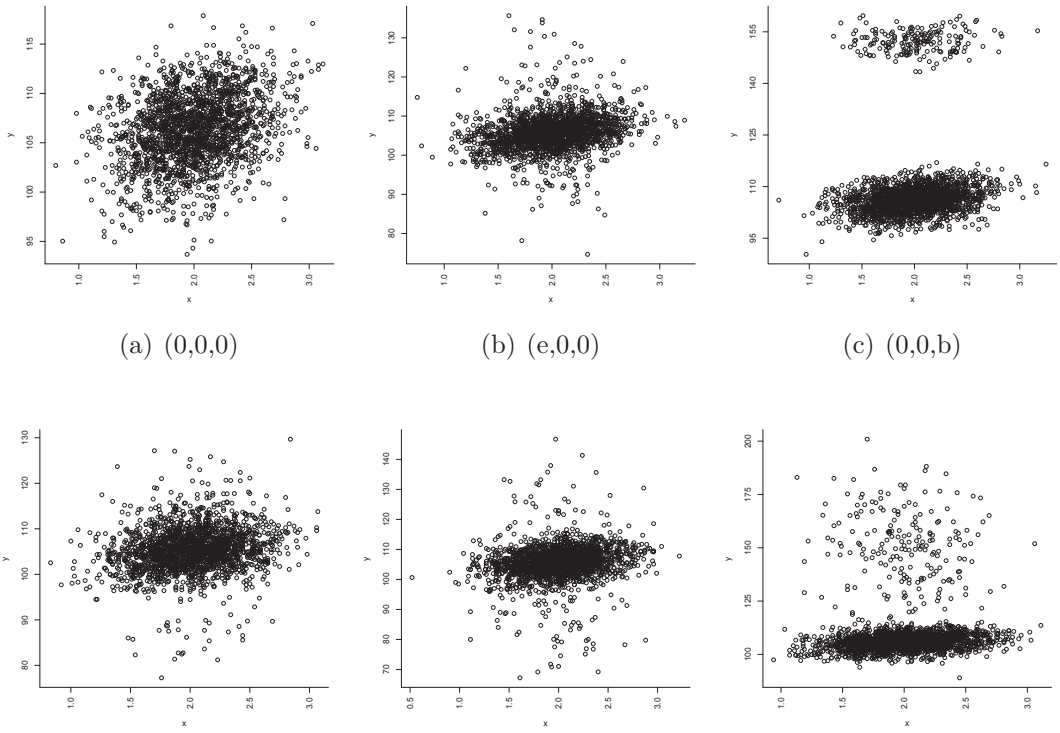
More complete simulation results are reported in this section. As in the article, we considered the mixture model ζ_m of two unit-level models

$$\zeta_0 : y_{0ij} = \beta_{00} + \beta_{01} x_{ij} + v_{0i} + e_{0ij} \quad (j = 1, \ldots, N_i; i = 1, \ldots, k), \quad (2.6.1)$$

$$\zeta_1 : y_{1ij} = \beta_{10} + \beta_{11} x_{ij} + v_{1i} + e_{1ij} \quad (j = 1, \ldots, N_i; i = 1, \ldots, k), \quad (2.6.2)$$

with $k = 40$ and $N_1 = \ldots = N_{40} = 50$. Normal distributions for the random effects and the error terms were also assumed. Thus, $v_{0i} \sim \mathcal{N}(0, \sigma^2_{v0})$, $v_{1i} \sim \mathcal{N}(0, \sigma^2_{v1})$, $e_{0i} \sim \mathcal{N}(0, \sigma^2_{e0})$ and $e_{1i} \sim \mathcal{N}(0, \sigma^2_{e1})$, $i = 1, \ldots, 50$, $k = 1, \ldots, 40$.

FIGURE 2.4. Populations generated according to the mixture model (2.6.1) and (2.6.2). The model parameters are given in Table 2.3. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to 0.5.



(a) (0,0,0) (b) (e,0,0) (c) (0,0,b)

(d) (0,v,0) (e) (e,v,0) (f) (e,v,b)

TABLE 2.3. Description of the six scenarios. The populations were generated according to $y_{ij} = (1 - A_{ij})y_{0ij} + A_{ij}y_{1ij}$, $A_{ij} \sim Bernoulli(0.1)$, using the unit-level models (2.6.1) and (2.6.2), assuming normality for the random effects and error terms in ζ_0 and ζ_1. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to $\rho_0 = 0.5$.

	Sources of the contamination		
Scenarios	Variances (error terms)	Variances (random effects)	Intercepts and slopes
$(0, 0, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 6)$	$\beta_0 = \beta_1 = (100, 3)^\top$
$(0, 0, b)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 6)$	$\beta_0 = (100, 3)^\top, \beta_1 = (150, 1)^\top$
$(0, v, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 150)$	$\beta_0 = \beta_1 = (100, 3)^\top$
$(e, 0, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 6)$	$\beta_0 = \beta_1 = (100, 3)^\top$
$(e, v, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 150)$	$\beta_0 = \beta_1 = (100, 3)^\top$
(e, v, b)	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 150)$	$\beta_0 = (100, 3)^\top, \beta_1 = (150, 1)^\top$

TABLE 2.4. Description of the six scenarios. The populations were generated according to $y_{ij} = (1 - A_{ij})y_{0ij} + A_{ij}y_{1ij}$, $A_{ij} \sim Bernoulli(0.1)$, using the unit-level models (2.6.1) and (2.6.2), assuming normality for the random effects and error terms in ζ_0 and ζ_1. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to $\rho_0 = 0.05$.

	Sources of the contamination		
Scenarios	Variances (error terms)	Variances (random effects)	Intercepts and slopes
$(0, 0, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6/19, 6/19)$	$\beta_0 = \beta_1 = (100, 3)^\top$
$(0, 0, b)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6/19, 6/19)$	$\beta_0 = (100, 3)^\top, \beta_1 = (150, 1)^\top$
$(0, v, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6/19, 69)$	$\beta_0 = \beta_1 = (100, 3)^\top$
$(e, 0, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6/19, 6/19)$	$\beta_0 = \beta_1 = (100, 3)^\top$
$(e, v, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6/19, 150/19)$	$\beta_0 = \beta_1 = (100, 3)^\top$
(e, v, b)	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6/19, 150/19)$	$\beta_0 = (100, 3)^\top, \beta_1 = (150, 1)^\top$

The auxiliary information was generated as in the article. In each area of a given population, random samples of size $n_i = 5$, $i = 1, \ldots, 40$, have been selected by simple random sampling without replacement. Adopting a notation similar to the one of Sinha & Rao (2009), six scenarios were studied, which are summarized in Tables 2.3 and 2.4.

From Tables 2.3 and 2.4, scenario $(0, 0, 0)$ corresponds to the absence of contamination. For the parameters in Table 2.3 the correlation between the units of a given area is set to $\rho_0 = 0.5$ in the absence of contamination. We studied scenarios where the correlation was smaller, and $\rho_0 = 0.05$ for the scenario $(0, 0, 0)$ in Table 2.4. The variance parameters of the random effects are also smaller in Table 2.4. Note that ρ_0 satisfies the relation : $\sigma^2_{v0} = \rho_0\sigma^2_{e0}/(1-\rho_0)$. Under $(0, 0, b)$,

TABLE 2.5. Monte Carlo relative biases (in percentage) and relative efficiencies (in percentage) for the robust predictors and the empirical best linear unbiased predictor of the small area means (averaged over areas). The parameters are given in Table 2.3 and $\rho_0 = 0.5$ under ζ_0.

	$(0,0,0)$	$(0,v,0)$	$(e,0,0)$	$(e,v,0)$	$(0,0,b)$	(e,v,b)
			Absolute relative bias			
EBLUP	0.025	0.037	0.039	0.047	0.065	0.071
CBb	0.025	0.036	0.031	0.042	1.741	1.803
CB3	0.024	0.036	0.034	0.042	0.977	1.076
CB6	0.025	0.037	0.037	0.045	0.463	0.545
CB9	0.025	0.037	0.038	0.047	0.199	0.281
CHAMb	0.027	0.030	0.031	0.034	1.782	1.844
CHAM3	0.027	0.033	0.035	0.036	1.012	1.112
CHAM6	0.027	0.035	0.040	0.040	0.494	0.576
CHAM9	0.027	0.036	0.041	0.042	0.224	0.307
SR	0.025	0.032	0.029	0.035	3.142	3.140
CCST1	0.027	0.034	0.034	0.038	3.122	3.055
CCST2	0.030	0.038	0.038	0.040	2.876	2.796
CCST3	0.031	0.039	0.043	0.042	2.574	2.524
			Relative efficiency			
EBLUP	100.0	100.0	100.0	100.0	100.0	100.0
CBb	100.3	94.2	76.2	88.1	136.0	129.7
CB3	99.7	92.4	82.6	87.8	105.9	104.7
CB6	100.0	96.8	94.5	92.8	94.0	94.2
CB9	100.0	99.4	98.9	97.0	94.2	93.3
CHAMb	110.1	88.4	66.2	73.2	123.3	117.9
CHAM3	110.5	90.1	75.1	76.0	93.7	93.6
CHAM6	109.3	95.8	87.6	83.4	83.6	84.6
CHAM9	108.0	98.1	91.7	88.4	85.8	85.2
SR	101.0	85.3	62.2	71.2	208.2	186.4
CCST1	111.5	95.9	73.0	80.9	214.3	203.6
CCST2	118.9	110.6	91.7	96.5	217.9	219.2
CCST3	120.7	120.0	107.0	110.1	230.7	227.4

the random effects and the error terms are not contaminated, but the model parameters in (2.6.1) and (2.6.2) are different. Under $(0, v, 0)$, only the area random effect is contaminated. Another example is the case (e, v, b), where the contamination comes from the random effects, the random errors and the fixed effects. The other scenarios are interpreted in the same manner. The different populations when $\rho_0 = 0.5$ are displayed in Figure 2.4.

Results for all the domains

Boxplots displaying empirical bias and efficiency for all the domains are presented in Figures 2.5, 2.6, 2.7 and 2.8, for all the scenarios described in Tables 2.3

TABLE 2.6. Monte Carlo relative biases (in percentage) and relative efficiencies (in percentage) for the robust predictors and the empirical best linear unbiased predictor of the small area means (averaged over areas). The parameters are given in Table 2.4 and $\rho_0 = 0.05$ under ζ_0.

	$(0,0,0)$	$(0,v,0)$	$(e,0,0)$	$(e,v,0)$	$(0,0,b)$	(e,v,b)
			Absolute relative bias			
EBLUP	0.013	0.018	0.019	0.020	0.040	0.040
CBb	0.013	0.017	0.017	0.018	2.215	2.230
CB3	0.013	0.017	0.018	0.019	1.300	1.352
CB6	0.013	0.018	0.019	0.020	0.628	0.698
CB9	0.013	0.018	0.019	0.020	0.323	0.383
CHAMb	0.014	0.018	0.018	0.018	2.216	2.230
CHAM3	0.015	0.018	0.019	0.019	1.299	1.351
CHAM6	0.015	0.019	0.021	0.020	0.626	0.696
CHAM9	0.014	0.019	0.021	0.021	0.321	0.381
SR	0.013	0.017	0.017	0.017	3.293	3.294
CCST1	0.022	0.024	0.028	0.027	3.196	3.118
CCST2	0.028	0.029	0.037	0.035	2.951	2.849
CCST3	0.029	0.032	0.042	0.040	2.651	2.567
			Relative efficiency			
EBLUP	100.0	100.0	100.0	100.0	100.0	100.0
CBb	99.1	103.9	85.6	88.1	203.3	193.1
CB3	99.9	100.6	90.8	91.7	135.3	132.5
CB6	100.0	99.7	96.5	96.7	101.6	101.0
CB9	100.0	100.0	99.1	99.0	94.4	93.5
CHAMb	113.1	107.6	91.6	92.4	204.7	193.9
CHAM3	114.8	106.3	96.9	97.0	136.9	133.4
CHAM6	112.2	105.6	102.5	102.1	103.4	102.2
CHAM9	110.0	105.2	104.3	103.8	96.5	94.9
SR	99.7	103.3	85.2	86.4	323.2	303.4
CCST1	194.1	139.3	159.0	152.0	325.5	325.8
CCST2	281.8	189.5	250.8	236.3	330.3	356.2
CCST3	310.3	211.5	312.7	293.4	348.2	372.4

and 2.4. More precisely, the relatives biases were computed for each domain in the Monte Carlo studies and boxplots of the $k = 40$ empirical biases were then represented. Similarly, relative empirical efficiencies with respect to the empirical best linear unbiased predictor were computed for each domain. The boxplots are displayed in Figures 2.6 and 2.8. In Tables 2.5 and 2.6, integrated measures are given, and the averages over the areas were computed.

From Figures 2.5 and 2.7, all the methods showed small empirical biases for the scenarios $(0,0,0)$, $(0,v,0)$, $(e,0,0)$ and $(e,v,0)$. Larger biases were observed under scenarios $(0,0,b)$ and (e,v,b) for the robust methods as a function of the

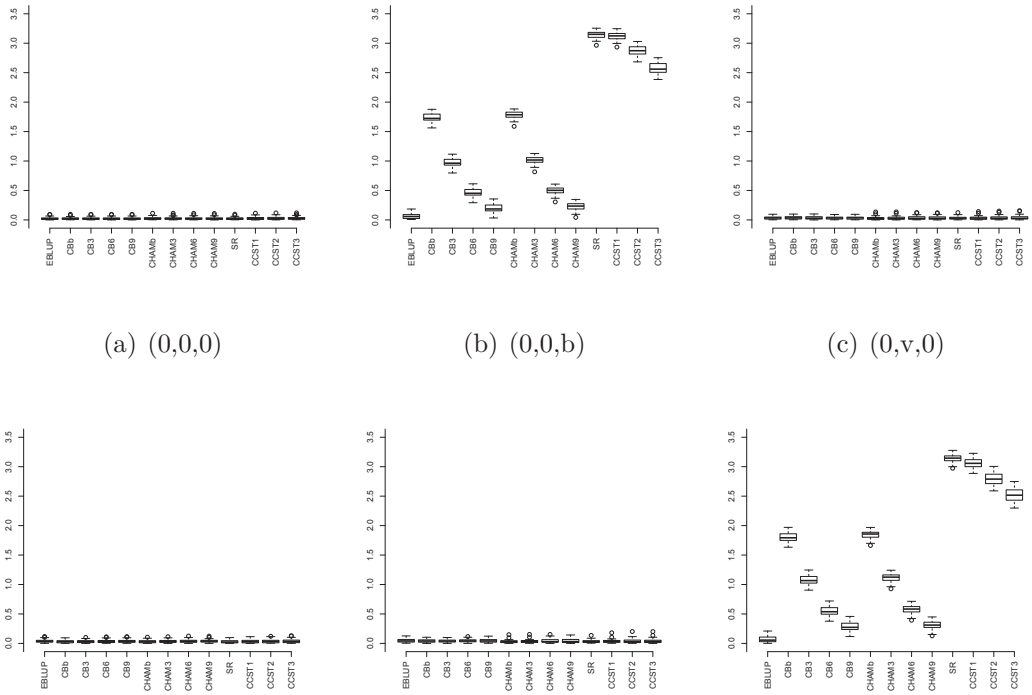
tuning constants. Small values of the tuning constants generated larger empirical biases. The empirical biases decreased as the tuning constants increased, as expected. From Tables 2.5 and 2.6, all the empirical biases were small for these scenarios. From Figures 2.6 and 2.8, the new robust methods were significantly more efficient than the robust predictors $\hat{\theta}_{i\text{SR}}$ or $\hat{\theta}_{i\text{CCST}}$, when the population was generated according to the mixture model. We observed more variability in the empirical efficiencies between the areas when $\rho_0 = 0.5$ than when $\rho_0 = 0.05$.

We now discuss the results in more details. Under scenario $(0, 0, 0)$, the robust predictors $\hat{\theta}_{i\text{CB}}$ and $\hat{\theta}_{i\text{SR}}$ appeared as efficient as the empirical best linear unbiased predictor, whereas $\hat{\theta}_{i\text{C}}$ was slightly less efficient. The robust predictor $\hat{\theta}_{i\text{CCST}}$ was approximately unbiased but the variance part of the mean squared error was important. From Tables 2.5 and 2.6, the integrated mean squared errors were 10-20% larger than the empirical best linear unbiased predictor when $\rho_0 = 0.5$, and larger by at least a factor two when $\rho_0 = 0.05$.

Under scenarios $(0, 0, b)$ and (e, v, b), small differences in efficiency were observed between $\hat{\theta}_{i\text{C}}$ and $\hat{\theta}_{i\text{CB}}$, for a given value of the tuning constant. The predictors $\hat{\theta}_{i\text{SR}}$ and $\hat{\theta}_{i\text{CCST}}$ showed large mean squared errors; for these predictors, the bias part of the mean squared error was non negligible. Efficiency gains were possible for several areas, and the median of the boxplots suggest that the proposed robust methods were more efficient. However, based on the integrated measures, it was more difficult to have gains in efficiency when $\rho_0 = 0.05$. The robust predictors $\hat{\theta}_{i\text{C}}$ and $\hat{\theta}_{i\text{CB}}$ with $q = b$ were less efficient than the empirical best linear unbiased predictor. In fact, for these scenarios, larger values of the tuning constants were necessary. When $\rho_0 = 0.5$, large gains in efficiency were observed under the scenarios $(0, v, 0)$, $(e, 0, 0)$ and $(e, v, 0)$. As expected, the efficiencies decreased as the tuning constant increased. Taking $q = 3$ or $q = 6$ seemed to give robustness and efficiency. From Tables 2.5 and 2.6, possible reductions in integrated mean squared errors of approximately 10%-20% were observed for these values of the tuning constants. The empirical efficiencies were smaller when the correlation coefficient was smaller.

Figure 2.9 presents the empirical coverage rates of the bootstrap confidence intervals for the nominal coverage rates 90% and 95%. Due to the nature of the mixture models, the empirical coverage rates were reasonably close to the nominal coverage rates under $(0, v, 0)$ and $(e, v, 0)$. When the slopes were different, the proposed bootstrap method worked reasonably well for the new methods for moderate to large values of the tuning constants. The results for $\hat{\theta}_{i\text{CCST}}$ were reasonable. However, the biases were very large for the Sinha–Rao method and the coverage rates were far from the nominal levels.

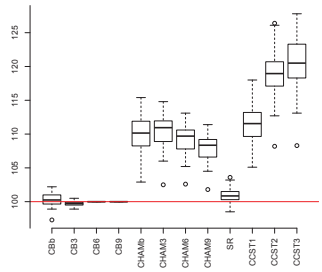
FIGURE 2.5. Boxplots of the absolute relative biases. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to 0.5.



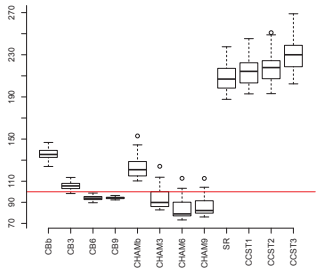
(a) (0,0,0) (b) (0,0,b) (c) (0,v,0)

(d) (e,0,0) (e) (e,v,0) (f) (e,v,b)

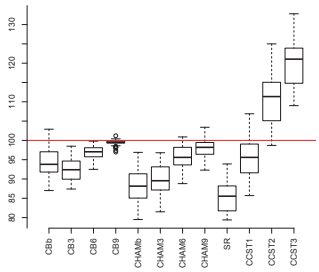
FIGURE 2.6. Boxplots of the relative efficiencies. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to 0.5.



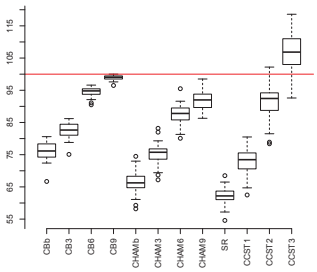
(a) (0,0,0)



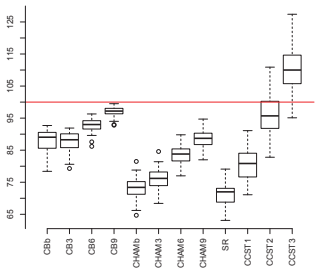
(b) (0,0,b)



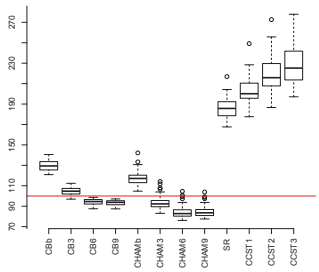
(c) (0,v,0)



(d) (e,0,0)

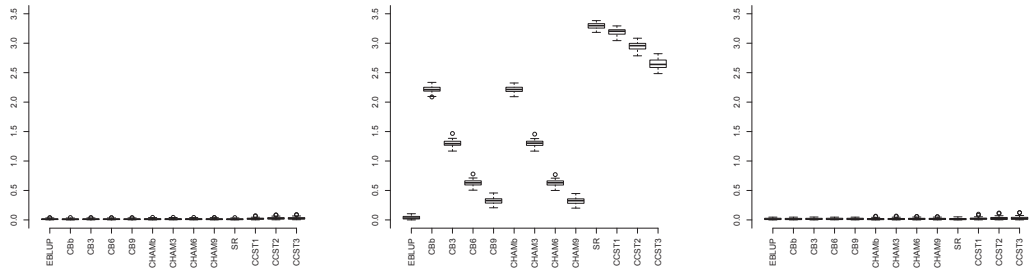


(e) (e,v,0)

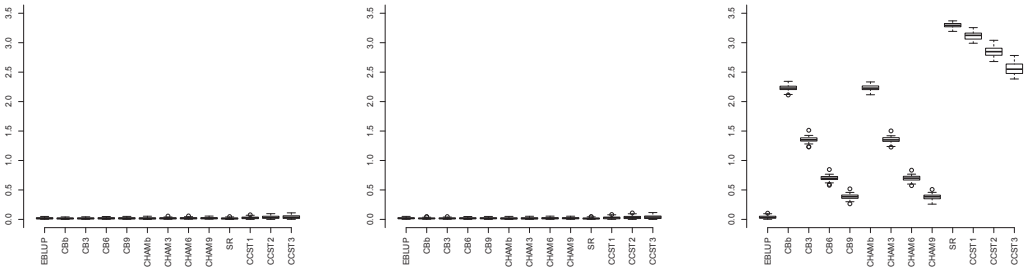


(f) (e,v,b)

FIGURE 2.7. Boxplots of the absolute relative biases. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to 0.05.

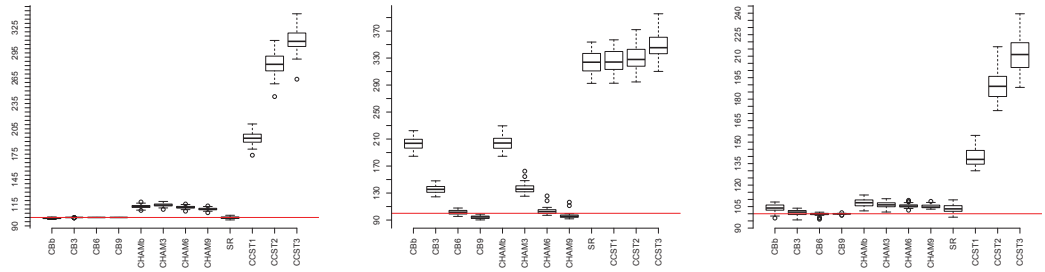


(a) (0,0,0) (b) (0,0,b) (c) (0,v,0)



(d) (e,0,0) (e) (e,v,0) (f) (e,v,b)

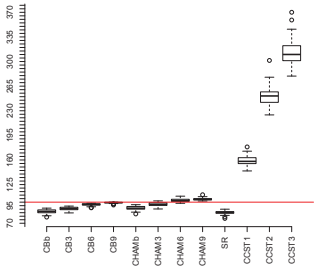
FIGURE 2.8. Boxplots of the relative efficiencies. Under the scenario $(0, 0, 0)$, the correlation between the units of the same domain is set to 0.05.



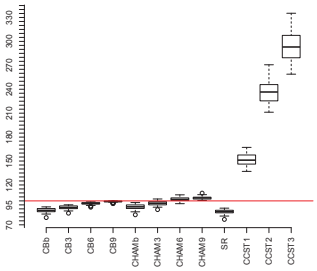
(a) (0,0,0)

(b) (0,0,b)

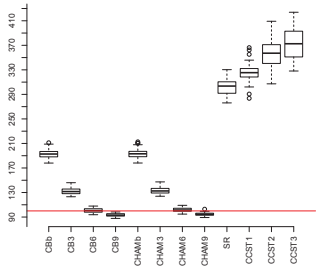
(c) (0,v,0)



(d) (e,0,0)

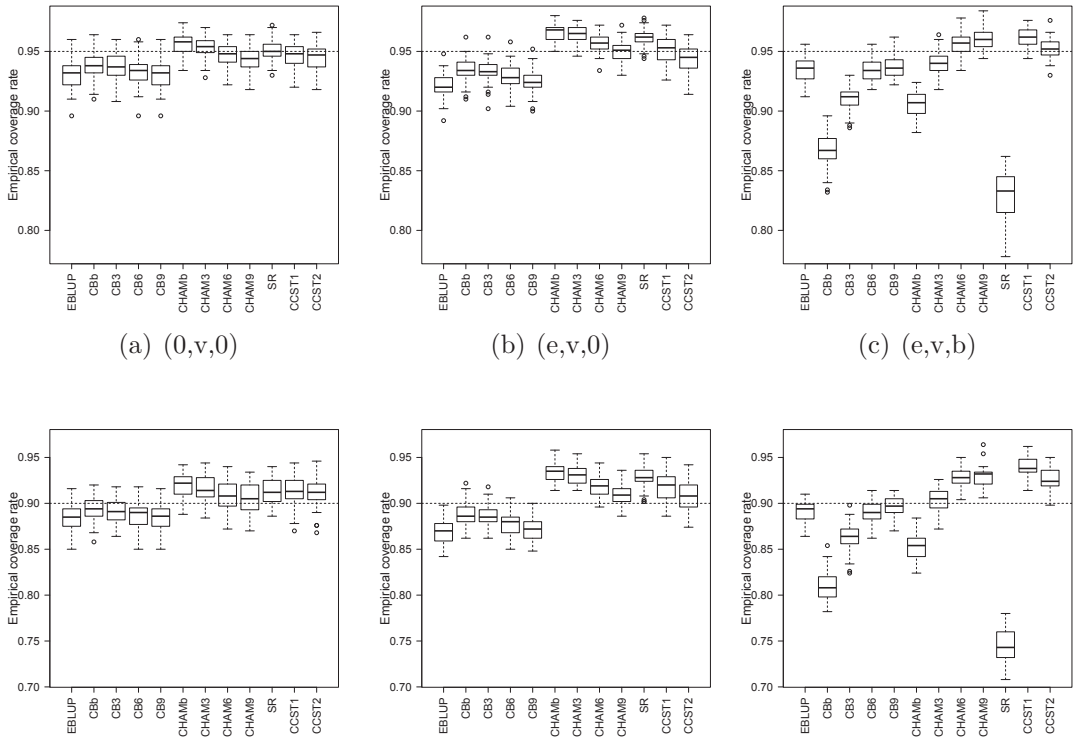


(e) (e,v,0)



(f) (e,v,b)

FIGURE 2.9. Empirical coverage rates under the scenarios $(0, v, 0)$, $(e, v, 0)$ and (e, v, b). In a)-c) the nominal coverage rates are 95%; in d)-f) the nominal coverage rates are 90%. Under the scenario (0,0,0), the correlation between the units of the same domain is set to $\rho_0 = 0.5$.



(a) (0,v,0) (b) (e,v,0) (c) (e,v,b)

(d) (0,v,0) (e) (e,v,0) (f) (e,v,b)

Bibliographie

[1] Beaumont, J.-F., Haziza, D. & Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, **100**, 555–569.

[2] Beaumont, J.-F. & Rivest, L.-P. (2009). Dealing with outliers in survey data. In *Handbook of Statistics, Sample Surveys, Design Methods and Applications*, **29A**, Ed. D. Pfeffermann and C. R. Rao, pp. 247–280. Amsterdam : North Holland.

[3] Chambers, R. L. (1986). Outliers robust finite population estimation. *Journal of the American Statistical Association*, **81**, 1063–1069.

[4] Chambers, R. L., Chandra, H., Salvati, N. & Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society. Series B*, **76**, 47–69.

[5] Chambers, R. L., Chandra, H., & Tzavidis, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methodology*, **37**, 153–170.

[6] Chambers, R. L. & Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, **93**, 255–268.

[7] Duchesne, P. (1999). Robust calibration estimators. *Survey Methodology*, **25**, 43–56.

[8] Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics*, **28**, 51–60.

[9] Ghosh, M., Maiti, T. & Roy, A. (2008). Influence functions and robust Bayes and empirical Bayes small area estimation. *Biometrika*, **95**, 573–585.

[10] Huggins, R. M. (1993). A robust approach to the analysis of repeated measures. *Biometrics*, **49**, 715–720.

[11] Lee, H. (1995). Outliers in business surveys. In *Business Survey Methods, Wiley Series in Probability and Mathematical Statistics*, Ed. B. G Cox, D. A. Binder, B. N. Chinnappa, A. Christianson, M. J. Colledge and P. S. Kott, pp. 503–526, New York : John Wiley.

[12] Moreno-Rebollo, J. L., Muñoz-Reyes, A. & Muñoz-Pichardo, J. (1999). Influence diagnostic in survey sampling : conditional bias. *Biometrika*, **86**, 923–928.

[13] Muñoz-Pichardo, J., Muñoz-Garcia, J., Moreno-Rebollo, J. L. & Pino-Mejias, R. (1995). A new approach to influence analysis in linear models. *Sankhya A*, **57**, 393–409.

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

[15] Rao, J. N. K. (2005). Interplay between sample survey theory and practice : an appraisal. Survey Methodology, **31**, 117–138.

[16] Richardson, A. M. & Welsh, A. H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics*, **51**, 1429–1439.

[17] Sinha, S. K. & Rao, J. N. K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, **37**, 381–399.

[18] Stahel, W. A. & Welsh, A. (1997). Approaches to robust estimation in the simplest variance components model. *Journal of Statistical Planning and Inference*, **57**, 295–319.

[19] Tzavidis, N., Marchetti, S. & Chambers, R. L. (2010). Robust estimation of small-area means and quantiles. *Australian & New Zealand Journal of Statistics*, **52**, 167–186.

[20] Welsh, A. H. & Ronchetti, E. (1998). Bias-calibrated estimation from sample surveys containing outliers. *Journal of the Royal Statistical Society. Series B*, **60**, 413–428.

Chapitre 3

BOOTSTRAPPING MEAN SQUARED ERRORS OF ROBUST SMALL AREA ESTIMATORS

ABSTRACT

This chapter proposes a new bootstrap estimation procedure for mean squared errors of robust small area predictors. Unlike existing approaches, we formally prove the asymptotic validity of the proposed bootstrap method. Moreover, since the proposed method is semiparametric, i.e., it does not rely on specific distributional assumptions about the errors and random effects of the unit-level model underlying the small area estimation, it is particularly attractive and more widely applicable. We assess the finite sample performance of our bootstrap estimator through Monte Carlo simulations. The results show that our procedure performs satisfactorily well and outperforms existing ones. Application of the proposed method is illustrated by analyzing a well-known outlier-contaminated small county crops area data from North-Central Iowa farms and Landsat satellite images.

Key words : Bootstrap, Small-area estimation, Empirical best linear unbiased predictor, Linear mixed model, Outliers, Random effects.

3.1. Introduction

The growing demand for reliable statistics of small geographical areas and subgroups of populations has generated a lot of interest for small area estimation (SAE) during the last decades (see Rao 2003, 2005, Pfeffermann 2013, for a recent review of methods used for SAE). Many theoretically sound and yet practical estimation procedures have been proposed of which the empirical bayes (EB) and the empirical best linear unbiased predictor (EBLUP) are used. In particular, the empirical best linear unbiased predictor is asymptotically unbiased and efficient

under correct model specification and distributional assumptions. However, it becomes highly sensitive in the presence of outliers or departures from the assumed normality of the random effects in the underlying model (see, e.g., Fellner 1986, Stahel & Welsh 1997, Sinha & Rao 2009).

Robustified versions of the classical estimators have therefore been recently investigated by several authors to downweight influential observations in the data when estimating the model (Sinha & Rao 2009, Chambers *et al.* 2014, Jiongo *et al.* 2013). The robust estimation of these small area quantities also requires the estimation of the Mean Square Error (MSE) which provides a measure of the precision of the point estimators. Sinha & Rao (2009) proposed a parametric bootstrap procedure based on the robust EBLUP estimators of the underlying linear mixed model to estimate the MSE. But Jiongo *et al.* (2013) point out that the use of robust variance estimates to generate bootstrap replicas leads to bootstrap samples whose variability are significantly smaller than the variability in the original data. Other analytical and bootstrap procedures for the MSE of the robust empirical best linear unbiased predictors (REBLUPs) have been proposed in Chambers *et al.* (2014) and Jiongo *et al.* (2013), respectively . However, their theoretical validity has not been formally established and their empirical performance are not fully satisfactory (as evidenced by the simulations results in Section 3.4 below).

In this paper, we propose a new semiparametric bootstrap procedure for estimating the mean-squared error of robust small area estimators. We focus on small area methods based on unit-level models that linearly relate the small area quantities of inferential interest to some unit level auxiliary covariates and includes random effects associated with the small areas. Since robust estimates of the variance components are typically smaller than their nonrobust counterparts and could yield bootstrap data on a smaller scale than the original data (Field *et al.* 2010), our bootstrap procedure uses (non-robust) maximum likelihood estimators to generate bootstrap samples, and robust bootstrap predictors to estimate the MSE. This produces bootstrap samples whose variability is similar to the one associated with the original data, and the resulting MSE estimator therefore has improved coverage rates. We formally prove the theoretical validity of our bootstrap procedure, examine its empirical performance through simulations and illustrate its use via a real data application. Our new MSE estimator is attractive for several reasons : first, unlike existing bootstrap estimators (e.g., Sinha & Rao 2009, Jiongo *et al.* 2013) it does not require specific distributional assumptions

about the error and random effects of the underlying unit-level model. It is therefore unaffected by misspecification biases that could arise in case of the not so uncommon non-normality. Although, for simplicity, we use the usual normality assumption to derive the results. Second, since our bootstrap samples mimic the original data, the use of the same estimation procedure for robust bootstrap predictors leads to a consistent MSE estimator. Third, the proposed procedure is easy to implement and our simulation results show that it performs satisfactorily well in finite samples and outperforms existing ones.

The validity of our bootstrap procedure uses an approach similar to the one used by Bickel & Friedman (1981) and Freedman (1981). Our procedure is an extension of the Freedman (1981) methodology to the MSE estimation of robust small area estimators in the linear mixed model framework. To our knowledge, this is the first study that provides sufficient conditions and a rigorous proof of the convergence of a MSE estimator of robust small area estimators. Although our proofs and procedure are based on the Sinha & Rao (2009) robust estimator, the derivation can be easily adapted to other existing robust predictors. A Monte Carlo simulation study computes the relative biases and relative root mean squared error rates of the proposed bootstrap MSE estimator and compares it to several analytical and bootstrap existing alternatives, including the bootstrap MSE estimator of Sinha & Rao (2009), the analytical pseudolinearization MSE estimator and linearization-based MSE estimators of Chambers *et al.* (2014), the bootstrap MSE of Jiongo *et al.* (2013) and the MSE estimator of Prasad & Rao (1990). This comparison is provided for different robust small area point estimators and various modes of data contamination.

The paper is organized as follows. Section 3.2 presents the model, notation and reviews some existing results which are summarized in Lemma 1. In Section 3.3, we present our proposed bootstrap procedure. Asymptotic properties and validity of the proposed method are also discussed. The validity proof of our bootstrap MSE proceeds in two main steps. Lemma 3 provides the requirements for the asymptotic validity of our bootstrap procedure and our main result is given in Theorem 3.3.2. Section 3.4 presents Monte Carlo simulation results showing that our procedure has satisfactory finite sample properties, and its performance is compared with the above-mentioned alternative estimators. A real data example on county crop areas is provided in Section 3.5 to show the usefulness of our method in practice. Concluding remarks are given in Section 3.6, followed by a technical appendix in Section 3.6.

3.2. Preliminairies

This section presents the basic linear mixed model that linearly relates the small area quantities of inferential interest to some unit level auxiliary covariates and includes random effects associated to the areas. We then briefly discuss the Sinha & Rao (2009) robust estimator that is used to construct our bootstrap MSE estimation.

3.2.1. Underlying Model

Consider a population $\mathcal{U}$ of size N, partitioned into k domains (areas) $\mathcal{U}_1, \ldots, \mathcal{U}_k$, of known sizes $N_1, \ldots, N_k$, respectively ; that is, $\mathcal{U} = \bigcup_{i=1}^{k} \mathcal{U}_i$ such that $\mathcal{U}_i \cap \mathcal{U}_l = \emptyset$, $i \neq l$, and $N = \sum_{i=1}^{k} N_i$. Let y define the variable of interest, and denote by y_{ij} the response value for unit j belonging to area i, $i = 1, \ldots, k$, $j = 1, \ldots, N_i$. The area mean associated with $\mathcal{U}_i$ is given by

$$\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}, \qquad (i = 1, \ldots, k),$$

Let s be the sample of size n, selected from the population $\mathcal{U}$ according to a given sampling plan $p(s)$ which we assume to be ignorable so that the population model also holds in the sample. The overall sample s can be partitioned as $s = \bigcup_{i=1}^{k} s_i$, where $s_i = s \cap \mathcal{U}_i$, of size n_i is the sample observed for sampled area i, $n = \sum_{i=1}^{k} n_i$. Note that n_i is random unless a planned sample of fixed size is taken in that area.
Traditional area-specific direct estimation methods (design-based or model-based) are not suitable in the small area context because of small (or even zero) area-specific sample sizes n_i. As a result, indirect estimation methods that borrow information across related areas through explicit models and auxiliary information, such as census and administrative data, are used for small area estimation. Denote by $\mathbf{x}_{ij} = (x_{1ij}, \ldots, x_{pij})^\top$ a p-dimensional deterministic vector of covariate values available for unit (i, j) and by $\bar{\mathbf{x}}_i = n_i^{-1} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$ the column vector of sample means of these covariates for area i. The corresponding vector of true area means is given by $\bar{\mathbf{X}}_i = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{x}_{ij}$, $i = 1, \ldots, k$, and is assumed to be available as well.

The nested error unit-level model considered can be expressed as

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i, \tag{3.2.1}$$

where $\boldsymbol{\beta}$ is an unknown p-dimensional fixed-effect regression parameter vector. The regressor $\mathbf{x}_{ij}$ is a p-dimensional vector of observed responses. The random

effects v_i are independently and identically distributed $\mathcal{N}(0, \sigma_v^2)$. The error terms e_{ij} are assumed to be independent and identically distributed $\mathcal{N}(0, \sigma_e^2)$, and independent of the v_i. We denote for notation simplicity $\mathcal{N}(0, \sigma_v^2)$ and $\mathcal{N}(0, \sigma_e^2)$ by F_v and F_e respectively.

Model (3.2.1) can be rewritten as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + v_i\mathbf{1}_{n_i} + e_i, \quad i = 1, \ldots, k, \tag{3.2.2}$$

where i is the area index, $\mathbf{y}_i$ is an n_i-dimensional vector of observed responses, $\mathbf{X}_i$ is a known $n_i \times p$ full rank design matrix, and $\mathbf{1}_{n_i}$ is a n_i-vector of ones. Denote by $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top)^\top$, the vector of model parameters, where $\boldsymbol{\delta}$ is the vector of variance parameters $\boldsymbol{\delta} = (\sigma_e^2, \sigma_v^2)^\top$. The variance-covariance matrix of $\mathbf{y}_i$ is given by $\mathbf{V}_i = \sigma_e^2\mathbf{I}_{n_i} + \sigma_v^2\mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top$, where $\mathbf{I}_{n_i}$ is the identity matrix of order n_i. The random effect component, v_i, accounts for the between-area variations that are not explained by the available auxiliary information $\mathbf{X}_i$.

3.2.2. Robust Estimation

For our bootstrap MSE procedure, we consider the class of robust estimators $\hat{\boldsymbol{\theta}}_R$ that are solutions to the following estimating equation :

$$\mathbf{S}(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}) \equiv \sum_{i=1}^{k} \boldsymbol{\Psi}(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) = \mathbf{0}, \tag{3.2.3}$$

where $\boldsymbol{\Psi}(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) = \left(\boldsymbol{\Psi}_1(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})^\top, \boldsymbol{\Psi}_2(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})^\top\right)^\top$; $\boldsymbol{\Psi}_1$ is a p-dimensional vector of estimating functions associated to the regression parameter $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}_2 = \left(\Psi_{21}^\top, \Psi_{22}^\top\right)^\top$ is a 2-dimensional vector of estimating functions associated to the variance parameters $\boldsymbol{\delta} = (\sigma_e^2, \sigma_v^2)^\top$. This class of estimators includes robust maximum likelihood estimators developed by Sinha & Rao (2009) for which

$$\begin{aligned}
\boldsymbol{\Psi}_1(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) &= \mathbf{X}_i^\top\mathbf{V}_i^{-1}\mathbf{U}_i^{1/2}\boldsymbol{\Psi}_b(\mathbf{r}_i) \\
\boldsymbol{\Psi}_{2l}(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) &= \boldsymbol{\Psi}_b^\top(\mathbf{r}_i)\mathbf{U}_i^{1/2}\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\delta}_l}\mathbf{V}_i^{-1}\mathbf{U}_i^{1/2}\boldsymbol{\Psi}_b(\mathbf{r}_i) - \mathrm{tr}\left(\mathbf{K}_\mathrm{i}\mathbf{V}_\mathrm{i}^{-1}\frac{\partial \mathbf{V}_\mathrm{i}}{\partial \boldsymbol{\delta}_\mathrm{l}}\right),
\end{aligned}$$

where $l = 1, 2$, $\mathbf{r}_i = \mathbf{U}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$; $\boldsymbol{\Psi}_b(\mathbf{r}_i) = (\psi_b(r_{i1}), \ldots, \psi_b(r_{in_i}))^\top$ is an n_i-vector of bounded functions, $\mathbf{U}_i = \mathrm{diag}(\mathbf{V}_\mathrm{i})$ is a diagonal matrix whose elements are the diagonal elements of the matrix $\mathbf{V}_i$, and $\mathbf{K}_i = E\left\{\psi_b^2(r)\right\}\mathbf{I}_{n_i}$ where r has the standard normal distribution $r \sim \mathcal{N}(0, 1)$. An example of function ψ_b is the Huber-type function defined by

$$\psi_b(u) = \min\{|b|, \max(-|b|, u)\}, \tag{3.2.4}$$

where b is a user-chosen positive constant. In a classical robust estimation framework under normality, a popular choice of the tuning constant b dictated by efficiency considerations is $b = 1.345$. Smoother versions of these functions can also be used as desired. Note that the case where $b \to \infty$, or, equivalently, $\psi_b(u) = u$ corresponds to the classical (nonrobust) maximum likelihood estimation.

Newton-Raphson algorithms to solve for these robust estimators numerically can be found in Sinha & Rao (2009). From the robustly estimated parameters, $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\delta}}_R^\top)^\top$, obtained from (3.2.3), the Sinha-Rao robust empirical best linear unbiased predictor (REBLUP) for the area mean $\bar{Y}_i$, denoted $\hat{\bar{Y}}_{i\mathrm{SR}}$, is of a plug-in type given by

$$\hat{\bar{Y}}_{i\mathrm{SR}} = N_i^{-1} \sum_{j \in s_i} y_{ij} + \left(1 - n_i N_i^{-1}\right) \bar{\mathbf{x}}_{ic}^\top \hat{\boldsymbol{\beta}}_R + \left(1 - n_i N_i^{-1}\right) \hat{v}_{iR}. \quad (3.2.5)$$

where $\bar{\mathbf{x}}_{ic} = \dfrac{1}{N_i - n_i} \sum_{j \in \mathcal{U}_i/s_i} x_{ij}$, and the robust predictors of the random effects, $\hat{v}_{iR} \equiv \hat{v}_{iR}(\hat{\boldsymbol{\delta}}_R)$, are obtained by solving the following Fellner (1986) system of estimating equations, conditionally on $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\delta}}_R^\top)^\top$:

$$\sigma_e^{-1} \sum_{i=1}^{k} \mathbf{X}_i^\top \boldsymbol{\Psi} \{\sigma_e^{-1}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - v_i \mathbf{1}_{n_i})\} = \mathbf{0}, \quad (3.2.6)$$

$$\sigma_e^{-1} \mathbf{1}_{n_i}^\top \boldsymbol{\Psi} \{\sigma_e^{-1}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - v_i \mathbf{1}_{n_i})\} - \sigma_v^{-1} \psi_b(\sigma_v^{-1} v_i) = \mathbf{0}, \quad (3.2.7)$$

where $i = 1, \ldots, k$. An alternative expression for the Sinha-Rao estimator is given by

$$\hat{\bar{Y}}_{i\mathrm{SR}} = N_i^{-1} \sum_{j \in s_i} y_{ij} + \left(1 - n_i N_i^{-1}\right) \bar{\mathbf{x}}_{ic}^\top \hat{\boldsymbol{\beta}}_R + \left(1 - n_i N_i^{-1}\right) \hat{\rho}_{iR} \sum_{j \in s_i} \hat{f}_{ijR} \left(y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R\right)$$

where

$$\hat{\rho}_{iR} = \frac{\sigma_{vR}^2 \sum_{j=1}^{n_i} \hat{a}_{ijR}}{\sigma_{vR}^2 \sum_{j=1}^{n_i} \hat{a}_{ijR} + \sigma_{eR}^2 \hat{b}_{iR}} \quad \text{and} \quad \hat{f}_{ijR} = \frac{\hat{a}_{ijR}}{\sum_{j=1}^{n_i} \hat{a}_{ijR}},$$

with

$$\hat{a}_{ijR} = \frac{\psi_b\left\{\hat{\sigma}_{eR}^{-1}\left(y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR}\right)\right\}}{\hat{\sigma}_{eR}^{-1}\left(y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_R - \hat{v}_{iR}\right)} \quad \text{and} \quad \hat{b}_{iR} = \frac{\psi_b\left(\hat{\sigma}_{vR}^{-1} \hat{v}_{iR}\right)}{\hat{\sigma}_{vR}^{-1} \hat{v}_{iR}}.$$

Denote by $\boldsymbol{\theta}_R = (\boldsymbol{\beta}_R^\top, \boldsymbol{\delta}_R^\top)^\top$ the probability limit of $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\delta}}_R^\top)^\top$ (also usually referred to as the robust target parameter). An expression for the prediction error, i.e. the difference between the predictor and the true area mean, can be obtained as follows :

$$\left(1 - n_i N_i^{-1}\right)^{-1} \left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i\right) = (\bar{\mathbf{x}}_{ic} - \hat{\rho}_{iR} \bar{\mathbf{x}}_{iR})^\top \left(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_R\right)$$

$$-(1-\hat{\rho}_{iR})v_i + \hat{\rho}_{iR}\bar{e}_{iR} - \bar{e}_{ic}$$
$$+\left(\bar{\mathbf{x}}_{ic} - \hat{\rho}_{iR}\bar{\mathbf{x}}_{iR}\right)^{\top}\left(\boldsymbol{\beta}_R - \boldsymbol{\beta}\right), \quad (3.2.8)$$

where

$$\bar{\mathbf{x}}_{iR} = \sum_{j=1}^{n_i} \hat{f}_{ijR}\mathbf{x}_{ij}, \quad \bar{e}_{iR} = \sum_{j=1}^{n_i} \hat{f}_{ijR}e_{ij} \quad \text{and} \quad \bar{e}_{ic} = \frac{1}{N_i - n_i}\sum_{j\in\mathcal{U}_i/s_i} e_{ij}.$$

As it will become clearer later, the expression for the prediction error provides a useful means to establish the convergence requirements for the validity of the bootstrapped MSE developed in this paper. Specifically, we will show that sufficient conditions to establish the convergence of our bootstrap using the Bickel & Freedman (1981) approach is to establish the convergence of the random effects v_i, the average error of the units of the area of interest, $\bar{e}_{iR}$, the average error of nonsampled units of the area of interest, $\bar{e}_{ic}$, and the robust ML estimator of the fixed effects $\hat{\boldsymbol{\beta}}_R$. For the latter, asymptotic properties of the whole parameter vector $\hat{\boldsymbol{\theta}}_R$ are needed. We state these properties in what follows.

3.2.3. Asymptotic Properties of the Robust Parameter Estimator

Denote by $E_m[\cdot]$ the expectation using Model (3.2.2). The asymptotic properties of the robust estimator $\hat{\boldsymbol{\theta}}_R$ are based on the following assumptions, which can be found in other related studies. These assumptions allow to rule out cases for which the limiting distributions of the estimated parameters either degenerate or blow-up.

Assumption A1. $\lim_{k\to\infty}\dfrac{k}{n} = c \in [0,1]$

Assumption A2. The covariates $\mathbf{X}_i$ are distributed over a bounded support.

Assumption A3. The $p \times p$ matrix $\mathbf{J}_1$ defined by

$$\mathbf{J}_1(\boldsymbol{\theta}) = \lim_{k\to\infty}\sum_{i=1}^{k}\mathbf{I}_{1k}^{-1/2}\mathbf{X}_i^{\top}\mathbf{V}_i^{-1}\mathbf{U}_i^{1/2}E_m\left\{\boldsymbol{\Psi}_b(\mathbf{r}_i)\boldsymbol{\Psi}_b(\mathbf{r}_i)^{\top}\right\}\mathbf{U}_i^{1/2}\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{I}_{1k}^{-1/2}$$

exists, is positive definite and continuous in $\boldsymbol{\theta}$; where $\mathbf{I}_{1k}$ is the $p \times p$ diagonal matrix defined by

$$\mathbf{I}_{1k} = \operatorname{diag}(k, n, \ldots, n) = \begin{pmatrix} k & 0 & \ldots & 0 \\ 0 & n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & n \end{pmatrix}$$

Assumption A4. The 2×2 matrix $\mathbf{J}_2$ defined by

$$\mathbf{J}_2(\boldsymbol{\theta}) = \lim_{k\to\infty} \sum_{i=1}^{k} \mathbf{I}_{2k}^{-1/2} E_m \left\{ \boldsymbol{\Psi}_2(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) \boldsymbol{\Psi}_2^{\top}(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) \right\} \mathbf{I}_{2k}^{-1/2}$$

exists, is positive definite and continuous in $\boldsymbol{\theta}$; where $\mathbf{I}_{2k} = \operatorname{diag}(k, n) = \begin{pmatrix} k & 0 \\ 0 & n \end{pmatrix}$.

Assumption A5. The $(p+2) \times (p+2)$ matrix $\mathbf{G}$ defined by $\mathbf{G}_k(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{G}(\boldsymbol{\theta})$ exists, is finite, positive definite and continuous in $\boldsymbol{\theta}$; where

$$\mathbf{G}_k(\boldsymbol{\theta}) = -\begin{bmatrix} \sum_{i=1}^{k} \mathbf{I}_{1k}^{-1/2} \dfrac{\partial \boldsymbol{\Psi}_1(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \mathbf{I}_{1k}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^{k} \mathbf{I}_{2k}^{-1/2} \dfrac{\partial \boldsymbol{\Psi}_2(\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\delta}} \mathbf{I}_{2k}^{-1/2} \end{bmatrix};$$

and the above convergence in probability is uniform on compacts of $\boldsymbol{\theta}$.

Assumption A6. $\mathbf{I}_k^{-1} \mathbf{S}(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}_R) \xrightarrow{p} \mathbf{0}$, where $\mathbf{I}_k = \begin{pmatrix} \mathbf{I}_{1k} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{2k} \end{pmatrix}$, $\mathbf{I}_{1k}$, $\mathbf{I}_{2k}$ are defined in A3 and A4.

Assumption A7. $\mathbf{I}_k^{-1/2} \mathbf{S}(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}_R) \xrightarrow{d} \mathcal{N}_{p+2}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{R}})$,
where $\boldsymbol{\Sigma}_R = \begin{pmatrix} \mathbf{J}_1(\boldsymbol{\theta}_R) & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2(\boldsymbol{\theta}_R) \end{pmatrix}$.

Assumption A1 states that the ratio of the number of areas over the total number of observations is asymptotically a constant fraction. This condition is weaker than the one required by Field *et al.* (2008) to establish the validity of the random effect bootstrap (for linear mixed models). Field *et al.* (2008) require that each of the area's sample size converges to infinity as the number of areas increases. In contrast, in our framework, all the areas could remain small as the number of areas increases. This condition is therefore similar to Assumption 3.2 of Miller (1977) which is a direct application of those that Weiss (1971, 1973, 1975) used to establish the asymptotic properties of maximum likelihood estimators in some nonstandard cases. As pointed out by Miller (1977), such an assumption is

reasonable and easily holds in most practical situations.

Assumptions A3 and A4 are similar to Assumptions 3.4 and 3.5 of Miller (1977). The matrices $\mathbf{J}_1$ and $\mathbf{J}_2$ defined within these assumptions determine the asymptotic covariance matrices of the fixed and random effects estimates respectively. Assumptions A3 and A4 ensure the existence and positive definiteness of these matrices. It should be noted that if either $\mathbf{J}_1$ or $\mathbf{J}_2$ does not exist or is not positive definite, then the associated estimates do not converge to a nondegenerate distribution. As explained by Miller (1977), any design or set of designs that might be used in practice would naturally satisfy these two assumptions.

Assumptions A5-A7 are equivalent to conditions A.1-A.4 of Huggins (1993). Assumption A5 is usually checked in an ad-hoc manner. For example, the existence of bounded derivatives or the Hölder or Lipschitz continuity of $\mathbf{G}_k(\cdot)$ on compacts of $\boldsymbol{\theta}$ would suffice for these conditions to hold. Assumptions A6 and A7 readily follow from the law of large numbers, the central limit theorem and the appropriate standard regularity conditions. Note that if the errors e_{ij} and the random effects v_i belong to asymmetrical distributions, then the covariance between the location part and the variance part of the estimation function may not be equal to zero. Thereby, the assumptions A5-A7 may not be verified.

The above assumptions guarantee that conditions A.1-A.4 of Huggins (1993) are satisfied. The following result is therefore a corollary of Theorem A of Huggins (1993), and is thus given without proof.

Lemma 1. *Under Assumptions A1-A7,*

$$\mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_R) \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \mathbf{G_R^{-1}\Sigma_R G_R^{-1}}), \tag{3.2.9}$$

where $\hat{\boldsymbol{\theta}}_R = (\hat{\boldsymbol{\beta}}_R^\top, \hat{\boldsymbol{\delta}}_R^\top)^\top$ *is the unique solution to (3.2.3), and* $\boldsymbol{\theta}_R$ *is its probability limit.*

Likewise, if we take $\psi_b(t) = t$ *in all of the functions given above, we obtain :*

$$\mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \mathbf{G_0^{-1}\Sigma_0 G_0^{-1}}), \tag{3.2.10}$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\delta}}^\top)^\top$ *is solution to (3.2.3), and* $\hat{\boldsymbol{\theta}}$ *is the maximum likelihood estimator of the true parameter vector* $\boldsymbol{\theta}_0$.

Proof : See Theorem A of Huggins (1993)

3.3. The Proposed MSE Bootstrap Estimator

This section proposes a bootstrap procedure designed for estimating the MSE of the robust empirical best linear unbiased predictors described in the previous section. We show that if the bootstrap samples are generated similarly to the process that generated the original data the bootstrap procedure for the MSE will be valid. Our bootstrap is semi-parametric and therefore avoids the possible bias due to incorrect specification of the random effects or the error distribution (Opsomer *et al.* 2008).

3.3.1. Description of the Bootstrap Method

We present the method of generating the bootstrap samples and estimating the MSE of the robust estimators. The method is described for the Sinha-Rao robust predictor and can be easily adapted for other predictors.The bootstrap procedure works as follows :

Step 1 : Here we proceed in two sub steps. Firstly, we compute

$$\hat{u}_i = \frac{1}{\sqrt{\hat{\rho}_i}}\hat{v}_i, \quad i = 1, \ldots, k, \quad \text{and} \quad \hat{e}_{lg} = y_{lg} - \mathbf{x}_{lg}^\top \hat{\boldsymbol{\beta}} - \frac{\hat{\tau}_l}{\hat{\rho}_l}\hat{v}_l,$$

where

$$l = 1, \ldots, k; g = 1, \ldots, n_l \quad \hat{\tau}_i = 1 - \sqrt{1 - \hat{\rho}_i}, \quad \hat{\rho}_i = \frac{n_i \hat{\sigma}_v^2}{\hat{\sigma}_e^2 + n_i \hat{\sigma}_v^2},$$

and $\hat{v}_i$ is the empirical best linear unbiased predictor of the random effect, given by

$$\hat{v}_i = \hat{\rho}_i(\bar{y}_i - \bar{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}). \tag{3.3.1}$$

Then we generate k random variables u_i^*, $i = 1, \ldots, k$, by drawing independently with replacement among $\hat{u}_i - \frac{1}{k}\sum_{l=1}^k \hat{u}_l$; and generate N random variables e_{ij}^*, $i = 1, \ldots, k$, $j = 1, \ldots, N_i$, by drawing independently with replacement among $\hat{e}_{lg} - \frac{1}{n}\sum_{l=1}^k \sum_{g=1}^{n_l} \hat{e}_{lg}$, $l = 1, \ldots, k$, $g = 1, \ldots, n_l$, respectively. The use of different subscripts notations here is made to emphasize the fact that the randomly selected component of e^* for which the coordinate (i, j) is assigned, e_{ij}^*, is independent of the corresponding area and unit from the original residual $\hat{e}_{ij}$.

Note that the estimates $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\delta}}^\top)^\top$, where $\hat{\boldsymbol{\delta}} = (\hat{\sigma}_e^2, \hat{\sigma}_v^2)^\top$, used at this step (and in steps 2 and 3) are the nonrobust estimators of $\boldsymbol{\theta}_0$. The use of nonrobust estimators leads to generated bootstrap samples whose behavior is similar to the original sample (e.g. same scale and variability), regardless of the original sample's distribution.

Step 2 : Compute the mean of the bootstrap population :

$$\bar{Y}_i^* = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + u_i^* + N_i^{-1} \sum_{j=1}^{N_i} e_{ij}^*. \tag{3.3.2}$$

Step 3 : Generate a bootstrap sample $(\mathbf{X}_i, \mathbf{y}_i^*)$, $i = 1, \ldots, k$, from the model

$$y_{ij}^* = \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + u_i^* + e_{ij}^*, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \tag{3.3.3}$$

where $\{e_{ij}^*; \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i\}$ is a sample of size n drawn from the population of bootstrapped errors using the same sampling plan $p(s)$ that was used to draw the original sample. Equation (3.3.3) can be rewritten in the form

$$\mathbf{y}_i^* = \mathbf{X}_i \hat{\boldsymbol{\beta}} + u_i^* \mathbf{1}_{n_i} + \mathbf{e}_i^*, \quad i = 1, \ldots, k. \tag{3.3.4}$$

Step 4 : Robust bootstrap estimators $\hat{\boldsymbol{\beta}}_R^*$, $\hat{\boldsymbol{\delta}}_R^*$, and $\hat{v}_{iR}^*$ are computed from the bootstrap samples. The Sinha-Rao robust bootstrap estimator for small area means, $\hat{\bar{Y}}_{iSR}^*$, is obtained as

$$\hat{\bar{Y}}_{i\mathrm{SR}}^* = N_i^{-1} \sum_{j \in s_i} y_{ij}^* + \left(1 - n_i N_i^{-1}\right) \bar{\mathbf{x}}_{ic}^\top \hat{\boldsymbol{\beta}}_R^* + \left(1 - n_i N_i^{-1}\right) \hat{v}_{iR}^*. \tag{3.3.5}$$

Step 5 : Repeat the above process a large number of times, say B times, to obtain B bootstrap samples and compute the estimator of the mean squared error of $\hat{\bar{Y}}_{iSR}$ by

$$\widehat{MSE}\left(\hat{\bar{Y}}_{i\mathrm{SR}}\right) = B^{-1} \sum_{b=1}^{B} \left(\hat{\bar{Y}}_{iSR}^{*(b)} - \bar{Y}_i^{*(b)}\right)^2,$$

where $\hat{\bar{Y}}_{iSR}^{*(b)}$ and $\bar{Y}_i^{*(b)}$ correspond to Expressions (3.3.5) and (3.3.2), respectively, for the b^{th} bootstrap sample.

Although the procedure does not specify the number of bootstrap samples to be generated, it is recommended to choose a number sufficiently large such that further increases do not substantially affect the estimated values. The bootstrap procedure described above has some particular advantages. It is semiparametric, i.e. it is flexible to any distribution of the error or random effects of the underlying unit-level model so that the generated bootstrap samples would pick up variants of the original distribution stemming from possibly data contamination. Moreover, as noted above, the use of nonrobust estimators in steps 1-3 allow the bootstrap samples to preserve the scale and variability of the original data. Hence, the proposed MSE estimator is expected to work reasonably well regardless of the nature of the outliers i.e., whether they are in the fixed-effects, the random effects or the error term, and should be robust to the misspecification of the random

components of the model. As shown by the simulation results given in Section 3.4, this presents a significant advantage over alternative bootstrap MSE procedures (e.g. Sinha & Rao 2009, Jiongo *et al.* 2013).

3.3.2. Validity of the Boostrap Estimator

Denote by $d_t, t = 1, 2, \ldots$, the Mallows (1972) metric for probabilities in $\mathbb{R}^{p+2}$, relative to the Euclidean norm $||.||$. If μ and ν are two probabilities in $\mathbb{R}^{p+2}$, then $d_t(\mu, \nu)$ is the infimum of $[E\left(||\mathbf{U} - \mathbf{V}||^t\right)]^{1/t}$ over all pairs of random vectors $\mathbf{U}$ and $\mathbf{V}$, where $\mathbf{U}$ has law μ and $\mathbf{V}$ has law ν. Also, for two random variables $\mathbf{U}$ and $\mathbf{V}$, write $d_t(\mathbf{U}, \mathbf{V})$ for the d_t-distance between the distributions of $\mathbf{U}$ and $\mathbf{V}$. Only the cases $t = 1, 2, 3$ or 4 are of interest in this paper.

Let $\hat{F}_{uk}$ be the empirical distribution of $\hat{u}_i$, $i = 1, \ldots, k$, centered at their mean, and let F_{uk} be the empirical distribution of u_i, $i = 1, \ldots, k$. Likewise, let $\hat{F}_{ek}$ be the empirical distribution of $\hat{e}_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, centered at their mean, and let F_{ek} be the empirical distribution of the e_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$. Define by $\Phi_k(F_{v,e})$ the distribution of $\mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_R)$, and by $\Phi_k(\hat{F}_{u,e})$ the distribution of $\mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}}_R^* - \hat{\boldsymbol{\theta}}_R)$, where $\hat{\boldsymbol{\theta}}_R^*$ is the robust estimate of $\hat{\boldsymbol{\theta}}$ obtained from the bootstrap sample $(\mathbf{X}_i, \mathbf{y}_i^*)$, $i = 1, \ldots, k$.

Denote by $E_*[\cdot]$ the bootstrap expectation. To derive the asymptotic properties of the bootstrap estimators we use the following bootstrap analogues of Assumptions A3 to A7 stated in Section 3.2.3, which are given conditionally on the original sample $(\mathbf{X}_i, \mathbf{y}_i)$, $i = 1, \ldots, k$.

Assumption B3. The $p \times p$ matrix $\mathbf{J}_1^*$ defined by

$$\mathbf{J}_1^*(\boldsymbol{\theta}) = \lim_{k\to\infty} \sum_{i=1}^{k} \mathbf{I}_{1k}^{-1/2} \mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{U}_i^{1/2} E_* \left\{ \boldsymbol{\Psi}_b(\mathbf{r}_i^*) \boldsymbol{\Psi}_b(\mathbf{r}_i^*)^\top \right\} \mathbf{U}_i^{1/2} \mathbf{V}_i^{-1} \mathbf{X}_i \mathbf{I}_{1k}^{-1/2}$$

exists, is positive definite and is a continuous function of $\boldsymbol{\theta}$.

Assumption B4. The 2×2 matrix $\mathbf{J}_2^*$ defined by

$$\mathbf{J}_2^*(\boldsymbol{\theta}) = \lim_{k\to\infty} \quad \sum_{i=1}^{k} \mathbf{I}_{2k}^{-1/2} E_* \left\{ \boldsymbol{\Psi}_2(\mathbf{y}_i^*, \mathbf{X}_i, \boldsymbol{\theta}) \boldsymbol{\Psi}_2^\top(\mathbf{y}_i^*, \mathbf{X}_i, \boldsymbol{\theta}) \right\} \mathbf{I}_{2k}^{-1/2}$$

exists, is positive definite and is a continuous function of $\boldsymbol{\theta}$.

Assumption B5. The $(p+2) \times (p+2)$ matrix $\mathbf{G}^*$ defined by $\mathbf{G}^*(\boldsymbol{\theta}) = \lim_{k\to\infty} \mathbf{G}_k^*(\boldsymbol{\theta})$ exists, is positive definite and continuous in $\boldsymbol{\theta}$; where

$$\mathbf{G}_k^*(\boldsymbol{\theta}) = -\begin{bmatrix} \sum_{i=1}^k \mathbf{I}_{1k}^{-1/2} \dfrac{\partial \boldsymbol{\Psi}_1(\mathbf{y}_i^*, \mathbf{X}_i, \boldsymbol{\theta})}{\partial \hat{\boldsymbol{\beta}}} \mathbf{I}_{1k}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^k \mathbf{I}_{2k}^{-1/2} \dfrac{\partial \boldsymbol{\Psi}_2(\mathbf{y}_i^*, \mathbf{X}_i, \boldsymbol{\theta})}{\partial \hat{\boldsymbol{\delta}}} \mathbf{I}_{2k}^{-1/2} \end{bmatrix}.$$

The above convergence of $\mathbf{G}_k^*(\boldsymbol{\theta})$ is uniform over compacts of $\boldsymbol{\theta}$.

Assumption B6. $\mathbf{I}_k^{-1}\mathbf{S}(\mathbf{y}^*, \mathbf{X}, \hat{\boldsymbol{\theta}}_R) \xrightarrow{p} 0$
Assumption B7. $\mathbf{I}_k^{-1/2}\mathbf{S}(\mathbf{y}^*, \mathbf{X}, \hat{\boldsymbol{\theta}}_R) \xrightarrow{d} \mathcal{N}_{p+2}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{R}}^*)$,
where $\boldsymbol{\Sigma}_R^* = \begin{pmatrix} \mathbf{J}_1^*(\boldsymbol{\theta}_R) & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2^*(\boldsymbol{\theta}_R) \end{pmatrix}$.

With these conditions, a bootstrap version of Lemma 1 is given by the following result.

Lemma 2. *Under Assumptions A1-A2, B3-B7 , and conditionally on the sample,*

$$\mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}}_R^* - \hat{\boldsymbol{\theta}}_R) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{G}^{*-1}_{\mathbf{R}} \boldsymbol{\Sigma}^*_{\mathbf{R}} \mathbf{G}^{*-1}_{\mathbf{R}}); \tag{3.3.6}$$

Likewise, when we take $\psi_b(t) = t$, we get :

$$\mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{G}^{*-1}_{\mathbf{0}} \boldsymbol{\Sigma}^*_{\mathbf{0}} \mathbf{G}^{*-1}_{\mathbf{0}}), \tag{3.3.7}$$

To show the validity of our bootstrap, the first step is to show that the bootstrap matrices given in the above conditions converge in probability to the original matrices, and the distribution of the bootstrap estimators asymptotically mimics the behavior of the original estimators in probability. These results are gathered in Lemma 3 and Theorem 3.3.1.

Lemma 3. *Let Assumptions A1-A2, B3-B7 hold. Then, for $k \to \infty$, and uniformly over $\boldsymbol{\theta}$,*

$$d_4\left(F_v, \hat{F}_{uk}\right) \xrightarrow{p} 0 \quad \textit{and} \quad d_4\left(F_e, \hat{F}_{ek}\right) \xrightarrow{p} 0 \tag{3.3.8}$$

$$\mathbf{J}_1^*(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{J}_1(\boldsymbol{\theta}) \tag{3.3.9}$$

$$\mathbf{J}_2^*(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{J}_2(\boldsymbol{\theta}) \tag{3.3.10}$$

$$\mathbf{G}^*(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{G}(\boldsymbol{\theta}) \tag{3.3.11}$$

Proof. It uses the Mallows (1972) metric for $t = 4$, and results from Bickel & Freedman (1981). See the Appendix at Section 3.6.

We next show that the asymptotic distribution of the robust bootstrap estimator is asymptotically equivalent to the asymptotic distribution of the robust initial estimator, conditionally on the sample.

Théorème 3.3.1. *Under Assumptions A1-A7 and B3-B7, and conditionally on the sample,*

$$d_2^{p+2}\left\{\Phi_k(F_{v,e}), \Phi_k(\hat{F}_{u,e})\right\} \xrightarrow{p} 0 \quad as\ k \to \infty.$$

Proof The proof follows immediately from the above results and Lemma 8.3 of Bickel & Freedman (1981). Denote $\boldsymbol{\xi}_R = \mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_R)$ and $\boldsymbol{\xi}_R^* = \mathbf{I}_k^{1/2}(\hat{\boldsymbol{\theta}}_R^* - \hat{\boldsymbol{\theta}}_R)$. Recall that their finite sample distribution are defined by $\Phi_k(F_{v,e})$ and $\Phi_k(\hat{F}_{u,e})$, respectively. By Lemmas 1 and 2, their asymptotic distributions are given by $\mathcal{N}(\mathbf{0}, \mathbf{G_R^{-1}\Sigma_R G_R^{-1}})$ and $\mathcal{N}(\mathbf{0}, \mathbf{G^*{}_R^{-1}\Sigma_R^* G^*{}_R^{-1}})$, respectively. It then follows by Lemma 3 and the Levy's Continuity Theorem that conditionally on the sample,

$$\boldsymbol{\xi}_R^* \xrightarrow{d} \boldsymbol{\xi}_R.$$

It also easily follows that

$$E_*\left[\|\boldsymbol{\xi}_R^*\|^2\right] \longrightarrow \operatorname{tr}\left(\mathbf{G}^{*-1}_R \boldsymbol{\Sigma}^*_R \mathbf{G}^{*-1}_R\right), \quad \text{and} \quad E_m\left[\|\boldsymbol{\xi}_R\|^2\right] \longrightarrow \operatorname{tr}\left(\mathbf{G}_R^{-1}\boldsymbol{\Sigma}_R\mathbf{G}_R^{-1}\right)$$

which, by Lemma 3 and the continuous mapping theorem, implies that

$$\left|E_*\left[\|\boldsymbol{\xi}_R^*\|^2\right] - E_m\left[\|\boldsymbol{\xi}_R\|^2\right]\right| \xrightarrow{p} \mathbf{0}.$$

It then follows by Lemma 8.3 a) of Bickel and Freedman (1981) that

$$d_2^{p+2}\left\{\Phi_k(F_{v,e}), \Phi_k(\hat{F}_{u,e})\right\} \xrightarrow{p} 0 \quad \text{as } k \longrightarrow \infty.$$

The following theorem is the main result of this paper. It states that under conditions given above, the proposed bootstrap MSE estimator of the Sinha & Rao (2009) robust empirical best linear unbiased predictor is a consistent estimator of the MSE.

Théorème 3.3.2. *Under Assumptions A1-A7 and B3-B7,*

$$\left|E_*\left(\hat{\bar{Y}}^*_{iSR} - \bar{Y}_i^*\right)^2 - E_m\left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i\right)^2\right| \xrightarrow{p} 0 \quad as \quad k \to \infty.$$

Démonstration. By Lemma 8.3 a) of Bickel and Freedman (1981), it is sufficient to show that $d_2\left(\hat{\bar{Y}}^*_{iSR} - \bar{Y}_i^*, \hat{\bar{Y}}_{iSR} - \bar{Y}_i\right) \xrightarrow{p} 0$. Denote $\hat{\boldsymbol{\gamma}}_R = \mathbf{I}_{1k}^{1/2}\left(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_R\right)$ and $\hat{\boldsymbol{\gamma}}_R^* = \mathbf{I}_{1k}^{1/2}\left(\hat{\boldsymbol{\beta}}_R^* - \hat{\boldsymbol{\beta}}_R\right)$. Then, from Equation (3.2.8) above, we can write $\left(1 - n_i N_i^{-1}\right)^{-1}\left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i\right)$ as an affine function of $\hat{\boldsymbol{\gamma}}_R$, v_i, $\bar{e}_{iR}$, $\bar{e}_{ic}$. That is,

$\left(1 - n_i N_i^{-1}\right)^{-1} \left(\hat{\bar{Y}}_{iSR} - \bar{Y}_i\right) = \mathbf{\Lambda}_i \left(\hat{\boldsymbol{\gamma}}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic}\right)$. It then follows from Assumptions A1 and A2 that there exists a positive constant, $M > 0$ such that

$$\| \mathbf{\Lambda}_i \left(\hat{\boldsymbol{\gamma}}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic}\right) \|^2 \leq M \left[1 + \| \left(\hat{\boldsymbol{\gamma}}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic}\right)^\top \|^2\right].$$

Given that, by Theorem 3.3.1 and Condition (3.3.8) of Lemma 3 above, we must have

$$d_2 \left(\left(\hat{\boldsymbol{\gamma}}_R^*, v_i^*, \bar{e}_{iR}^*, \bar{e}_{ic}^*\right)^\top, \left(\hat{\boldsymbol{\gamma}}_R, v_i, \bar{e}_{iR}, \bar{e}_{ic}\right)^\top \right) \xrightarrow{p} 0.$$

It then follows by Lemma 8.5 of Bickel and Freedman (1981) that

$$d_2 \left(\hat{\bar{Y}}_{iSR}^* - \bar{Y}_i^*, \hat{\bar{Y}}_{iSR} - \bar{Y}_i \right) \xrightarrow{p} 0$$

□

3.4. Monte Carlo Simulations

In this section we carry out Monte Carlo simulations to explore the finite sample performance of the proposed bootstrap procedure for estimating the MSE. For this purpose we consider four small area estimators. The empirical best linear unbiased estimator, EBLUP, the robust estimator of Sinha & Rao (2009), SR, the robust estimator of Chambers *et al.* (2014), CCST3, and the robust estimator of Jiongo *et al.* (2013) based on the conditional bias concept of Beaumont *et al.* (2013), CB. For each of these small area estimators, the performance of the proposed bootstrap MSE procedure, denoted JNBOOT, is assessed and compared with several other alternative MSE estimators. For the small area estimators CB, we compare our results with the bootstrap MSE estimators of Sinha & Rao (2009), denoted SRBOOT, and Jiongo *et al.* (2013), denoted JHDBOOT. For the robust estimators SR and CCST3, we also compare our results with the analytical linearization MSE and linearization-based MSE estimators developed by Chambers *et al.* (2014), denoted CCT and CCST, respectively. Finally, for the EBLUP, we compare our results with all of the above including the estimator of Prasad and Rao (1990), denoted PR. [1]

3.4.1. Simulation Design

We consider the same type of contamination design as in Jiongo *et al.* (2013). Outliers are generated from a mixture model ζ_m satisfying $y_{ij} = (1 - A_{ij}) y_{0ij} + A_{ij} y_{1ij}$, where the A_{ij} are independently generated according to a Bernoulli distribution with parameter $p = 0.1$, and y_{0ij} and y_{1ij} are given by two mixed linear

1. Note that for the EBLUP, the bootstrap procedures SRBOOT and JHDBOOT are equivalent. Hence, only JHDBOOT is reported for this case.

models defined by

$$\zeta_0 : y_{0ij} = \beta_{00} + \beta_{01}x_{ij} + v_{0i} + e_{0ij}, \quad (j = 1, \ldots, N_i; i = 1, \ldots, k), \quad (3.4.1)$$

$$\zeta_1 : y_{1ij} = \beta_{10} + \beta_{11}x_{ij} + v_{1i} + e_{1ij}, \quad (j = 1, \ldots, N_i; i = 1, \ldots, k). \quad (3.4.2)$$

We take $k = 40$ and $N_1 = \ldots = N_{40} = 50$. The error terms and random effects are assumed to be normally distributed and are given by $v_{0i} \sim \mathcal{N}(0, \sigma^2_{v0})$, $v_{1i} \sim \mathcal{N}(0, \sigma^2_{v1})$, $e_{0ij} \sim \mathcal{N}(0, \sigma^2_{e0})$ and $e_{1ij} \sim \mathcal{N}(0, \sigma^2_{e1})$, $(k = 1, \ldots, 40; j = 1, \ldots, 50)$.

The values of the auxiliary variable are generated from a normal distribution with mean $E(X) = 2$ and standard deviation $V^{1/2}(X) = 0.35$. In each area of the population, random samples of size $n_1 = \ldots = n_{40} = 5$ have been selected by simple random sampling without replacement. Three contamination scenarios are considered and Table 3.1 provides a description of each scenario. Scenario $(0, 0, 0)$ corresponds to the absence of contamination, while scenario $(e, v, 0)$ corresponds to having both the random errors and the area random effects contaminated. Finally, scenario (e, v, b) corresponds to the situation where the contamination comes from the random effects, the random errors and the fixed effects. These scenarios are similar to those given in Sinha & Rao (2009), but an additional scenario has been added here to allow for $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})$ to be different from $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11})$. For the parameters given in Table 3.1, the correlation between the units in a given area is equal to $\rho_0 = 0.5$ in the absence of contamination, where ρ_0 satisfies the relation $\sigma^2_{v0} = \rho_0\sigma^2_{e0}/(1 - \rho_0)$. Figure 3.1 provides a picture of the simulated data for various modes of contamination.

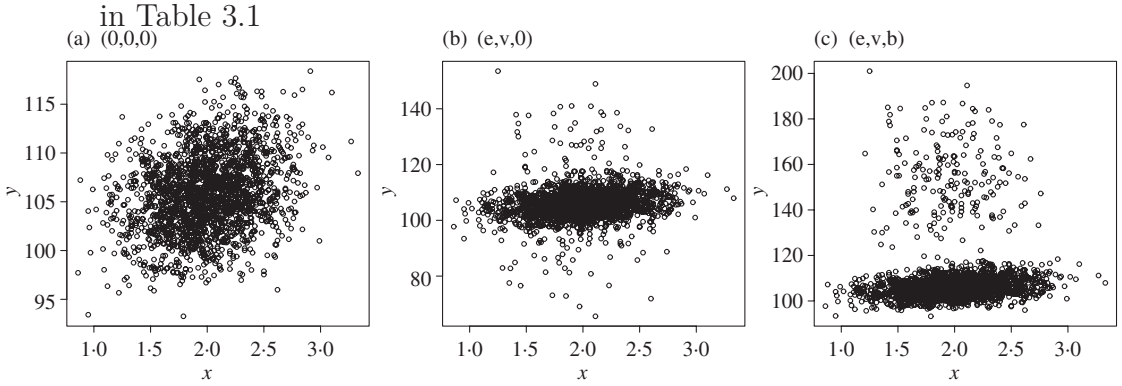
TABLE 3.1. Description of the contamination scenarios.

The populations are generated according to $y_{ij} = (1 - A_{ij})y_{0ij} + A_{ij}y_{1ij}$, $A_{ij} \sim$ *Bernoulli*(0.1), *using the unit-level models (3.4.1) and (3.4.2), assuming normality for the random effects and error terms in* ζ_0 *and* ζ_1. *Under the scenario* $(0, 0, 0)$, *the correlation between the units of the same domain equals 0.5.*

	Sources of the contamination		
Scenarios	Variances of error terms	Variances of random effects	Intercepts and slopes
$(0, 0, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 6)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 6)$	$\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1 = (100, 3)^\top$
$(e, v, 0)$	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 150)$	$\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1 = (100, 3)^\top$
(e, v, b)	$(\sigma^2_{e0}, \sigma^2_{e1}) = (6, 150)$	$(\sigma^2_{v0}, \sigma^2_{v1}) = (6, 150)$	$\boldsymbol{\beta}_0 = (100, 3)^\top, \boldsymbol{\beta}_1 = (150, 1)^\top$

For each scenario, we generate $T = 500$ populations and $B = 200$ bootstrap replications. The tuning constant for the small area estimators CB is set as defined in Beaumont *el al.* (2013) which leads to a robust predictor given by $\hat{Y}_{iCB} = \hat{Y}_{iEBLUP} - \frac{1}{2}\left(\hat{B}^{min}_{iR} + \hat{B}^{max}_{iR}\right)$, where $\hat{B}^{min}_{iR}$ and $\hat{B}^{max}_{iR}$ are the minimum

FIGURE 3.1. Scatter plots of the populations generated from the mixture model (3.4.1) and (3.4.2). The model parameters are given in Table 3.1



and the maximum of the robust estimator of the conditional bias respectively. See Jiongo *el al.* (2013) for the details on the estimation of the conditional bias. The tuning constant of the robust predictor CCST3 is set at $b = 3$ as in the simulation experiments of Chambers *et al.* (2014). Although the robust estimation of the small area means is not the subject of this paper, we present the results of the relative absolute bias and root relative mean squared errors (RRMSE) of each of the small area estimators considered, for completeness.

Let $\hat{\bar{Y}}_i$ denotes an arbitrary estimator of the small area mean $\bar{Y}_i$. Then the absolute relative bias for the area mean $\bar{Y}_i$ associated to $\hat{\bar{Y}}_i$ is given by

$$\mathrm{ARB}(\hat{\bar{Y}}_i) = 100 \times \left| T^{-1} \sum_{t=1}^{T} \frac{\hat{\bar{Y}}_i^{(t)} - \bar{Y}_i^{(t)}}{\bar{Y}_i^{(t)}} \right|, \quad (i = 1, \ldots, k), \tag{3.4.3}$$

and the root relative mean squared error is given by

$$\mathrm{RRMSE}(\hat{\bar{Y}}_i) = 100 \times \sqrt{T^{-1} \sum_{t=1}^{T} \left(\frac{\hat{\bar{Y}}_i^{(t)} - \bar{Y}_i^{(t)}}{\bar{Y}_i^{(t)}} \right)^2}, \quad (i = 1, \ldots, k).$$

For the estimation of the MSE, we also compute the empirical values of the relative bias (RB) and the root relative mean squared error. Denote by $\widehat{\mathrm{MSE}}(\hat{\bar{Y}}_i)$ the estimator of the mean squared error of $\hat{\bar{Y}}_i$. The relative bias associate to $\widehat{\mathrm{MSE}}(\hat{\bar{Y}}_i)$ is given by

$$\mathrm{RB}\left(\widehat{\mathrm{MSE}}(\hat{\bar{Y}}_i)\right) = 100 \times T^{-1} \sum_{t=1}^{T} \frac{\widehat{\mathrm{MSE}}(\hat{\bar{Y}}_i)^{(t)} - \mathrm{MSE}(\hat{\bar{Y}}_i)}{\mathrm{MSE}\hat{\bar{Y}}_i)}, \quad (i = 1, \ldots, k),$$

and the root relative mean squared error of $\widehat{\text{MSE}}(\hat{\bar{Y}}_i)$ is calculated as

$$\text{RRMSE}\left(\widehat{\text{MSE}}(\hat{\bar{Y}}_i)\right) = 100\times\sqrt{T^{-1}\sum_{t=1}^{T}\left(\frac{\widehat{\text{MSE}}(\hat{\bar{Y}}_i)^{(t)} - \text{MSE}(\hat{\bar{Y}}_i)}{\text{MSE}(\hat{\bar{Y}}_i)}\right)^2}, \quad (i = 1, \ldots, k).$$

Section 3.4.2 presents simulation results based on all the domains. We use boxplots and measures of central tendency such as the median of all the areas. Simulation results are obtained under scenarios $(0, 0, 0)$, $(e, v, 0)$ and (e, v, b) described in Table 3.1.

3.4.2. Simulation Results

TABLE 3.2. Monte Carlo absolute relative biases (%) and relative root mean squared error (%) for the predictors of the small area means (at the median over areas).

Scenario	Absolute relative bias				Root relative mean squared error			
	EBLUP	CB	SR	CCST3	EBLUP	CB	SR	CCST3
$(0, 0, 0)$	0.03	0.03	0.04	0.04	0.96	0.98	0.97	1.06
$(e, v, 0)$	0.06	0.06	0.05	0.06	1.75	1.57	1.47	1.84
(e, v, b)	0.10	0.18	3.13	2.37	2.93	2.73	3.86	4.41

The results reported in Table 3.2 present the percent Monte Carlo absolute relative biases (ARB %) and the percent relative root mean squared error (RRMSE %) for the EBLUP and the robust predictors of the small area means, where the computation is given for the median of all the areas. The results show that the estimator CB proposed by Jiongo *et al.* (2013) perform well with the given value of the tuning constant regardless of the mode of contamination. That is, whether the contamination occurs at the errors level, the random effect level or the fixed effects level. On the other hand, the biases of the Sinha & Rao (2009) and the Chambers *et al.* (2014) predictors are quite similar for the case (e,v,0), while the former tend to yield smaller mean squared error than the latter. For the case (e,v,b), the Chambers *et al.* (2014) predictor has less bias than the Sinha & Rao (2009), but is more variable.

The results reported in Table 3.3 present the percent Monte Carlo relative biases (RB %) and the percent root relative mean squared error (RRMSE %) of the mean squared error estimator of the predictors of the small area means, obtained at the median of the areas. In the absence of outliers (scenario $(0, 0, 0)$ in Table 3.3), only the analytical pseudolinearization MSE estimators (CCT) and linearization-based MSE estimators (CCST) of the MSE are biased when the

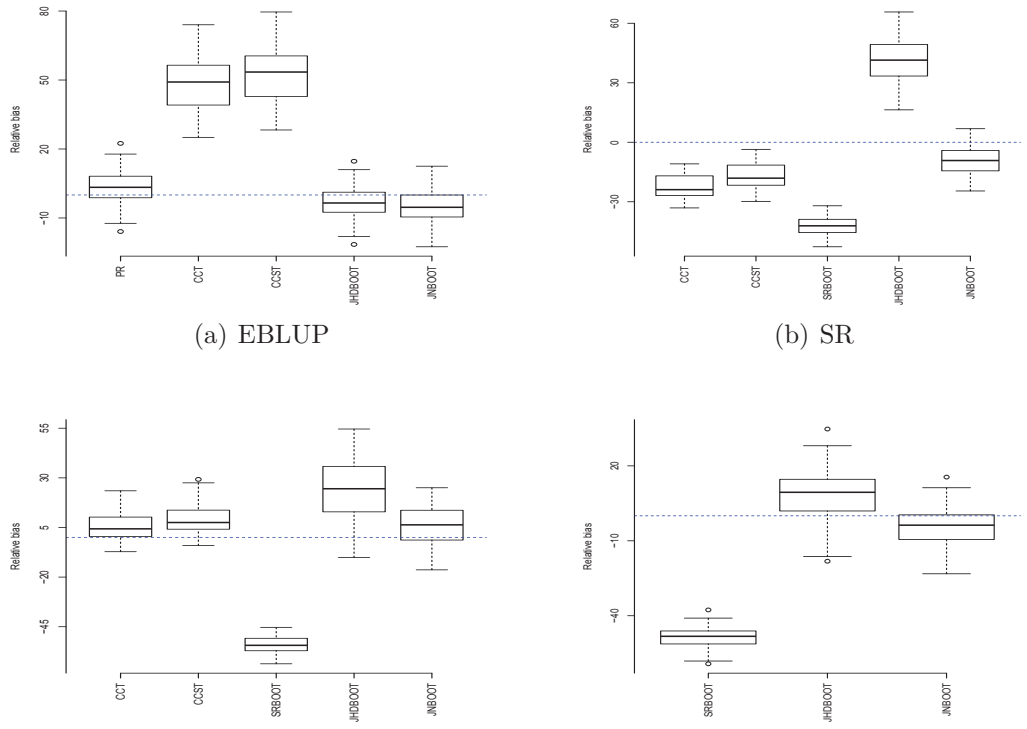
TABLE 3.3. Monte Carlo relative biases (RB %) and root relative mean squared error (RRMSE %) for the mean squared error estimator of the predictors of small area means (at the median of the areas).

		$(0,0,0)$		$(e,v,0)$		(e,v,b)	
SAE	MSE	RB	RRMSE	RB	RRMSE	RB	RRMSE
EBLUP	PR	0.39	11.74	3.33	31.28	34.87	57.09
	CCT	-2.31	43.41	49.13	195.70	238.50	549.90
	CCST	0.21	44.96	53.48	221.90	243.50	587.80
	JHDBOOT	-1.28	15.19	-3.48	32.63	12.22	49.27
	JNBOOT	-1.02	15.60	-5.36	33.02	10.93	48.61
SR	CCT	-3.34	56.76	-23.90	48.96	-70.03	71.71
	CCST	-0.88	57.43	-18.10	89.58	-65.56	81.95
	SRBOOT	2.77	17.67	-42.14	43.46	-91.44	91.46
	JHDBOOT	1.66	15.59	41.42	62.58	-37.65	46.82
	JNBOOT	0.97	15.90	-9.21	25.36	-9.33	38.36
CCST3	CCT	40.75	97.16	4.09	150.40	-46.72	122.00
	CCST	42.81	104.80	7.50	175.80	-43.84	138.80
	SRBOOT	1.11	18.04	-54.39	55.08	-91.55	91.57
	JHDBOOT	-0.25	16.61	24.51	85.56	65.38	81.61
	JNBOOT	-0.56	16.65	6.32	33.08	1.00	44.48
CB	SRBOOT	0.74	17.19	-48.25	49.20	-82.22	82.32
	JHDBOOT	0.05	15.30	9.41	50.85	40.03	70.02
	JNBOOT	-0.51	15.68	-3.75	28.32	3.95	41.12

Chambers *et al.* (2014) robust small area predictor CCST3 is used. All the other MSE estimators are equivalent in terms of bias and display negligible biases, regardless of the small area estimator considered. Likewise, for the MSE estimators it can be noted that only analytical pseudolinearization MSE estimators (CCT) and linearization-based MSE estimator (CCST) are unstable throughout. In contrast, all the bootstrap estimators are stable and equivalent to each other regardless of the small area estimator considered. For the particular case of the empirical best linear unbiased predictor (EBLUP), the Prasad & Rao (1990) MSE estimator (PR) is more stable than the bootstrap estimators.

Consider the case where outliers are present in the errors and the random effects (scenario $(e,v,0)$ in Table 3.3). We also note that, except for the Chambers *et al.* (2014) robust estimator (CCST3), the analytical pseudolinearization MSE estimator (CCT) and linearization-based MSE estimator (CCST) of the MSE are biased for all other small area estimators considered. In general, it can be seen that for all the robust predictors of the small area mean under consideration, the bootstrap MSE of Sinha & Rao (2009), SRBOOT, is negatively biased while the bootstrap MSE of Jiongo *et al.* (2013), JHDBOOT, is positively biased. Moreover,

FIGURE 3.2. Boxplots of the relative biases of the MSE estimators. Scenario $(e, v, 0)$



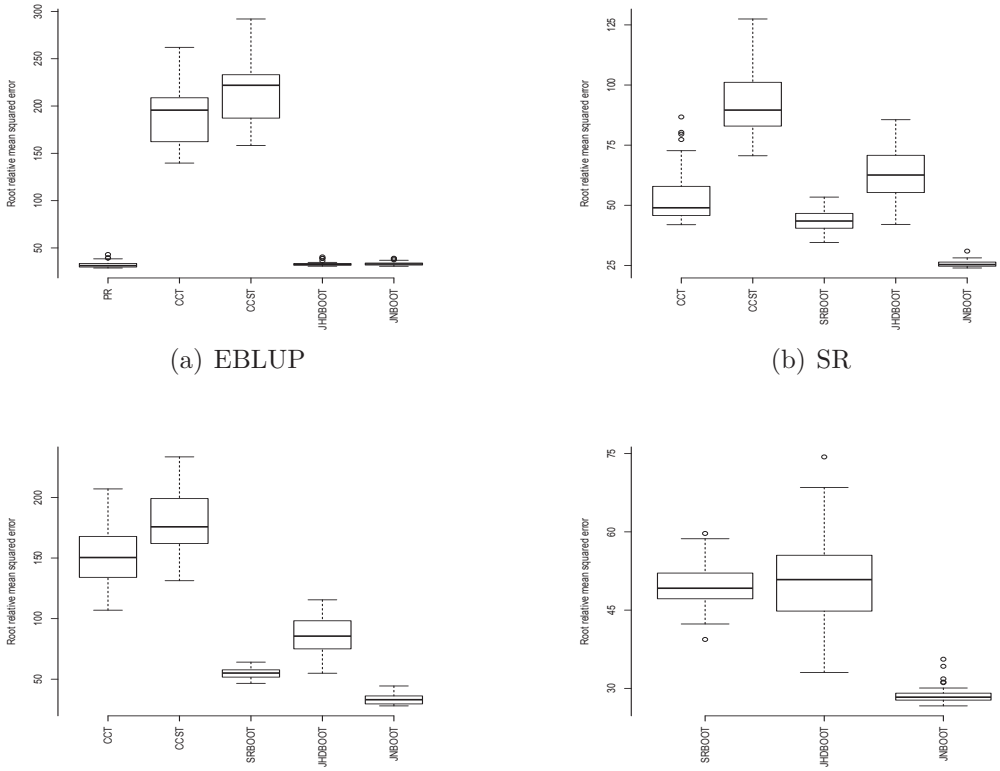
(a) EBLUP (b) SR

(c) CCST3 (d) CB

these biases are generally higher than 20% in absolute value. In sharp contrast, the proposed bootstrap MSE, JNBOOT, performs the best and exhibits smaller biases, always less than 12% in absolute value. This clearly shows the striking difference in terms of performance between our estimator and the others. As for the variability of these MSE estimators, the proposed bootstrap is again the most efficient. Indeed, regardless of the robust predictor considered the relative root mean square error of the competitors CCT, CCST, SRBOOT or JHDBOOT is generally higher than the RRMSE of the proposed bootstrap, JNBOOT. Figure 3.2 and 3.3 depict the boxplots of the relative biases and the root relative mean squared error of the MSE estimators, respectively, which further shows graphically the superiority of the proposed bootstrap method over existing alternatives for all the robust small area predictors.

Finally, if we look at the case where outliers are simultaneously present in the error terms, the random effects and the fixed effects (that is, scenario (e, v, b) in Table 3.3), the estimators, CCT, CCST, SRBOOT or JHDBOOT, are highly biased, regardless of the robust small area predictor considered. In contrast, the

FIGURE 3.3. Boxplots of the root relative mean squared error of the MSE estimators. Scenario $(e, v, 0)$



(a) EBLUP (b) SR

(c) CCST3 (d) CB

proposed bootstrap, JNBOOT, is unbiased with relative bias as low as 1.00% for the robust predictors CCST3 of Chambers *et al.* (2014), and 3.95% for the robust predictors CB of Jiongo *et al.* (2013). The relative bias of the proposed estimator is moderate (−9.21%) for SR of Sinha & Rao (2009), yet by far still lower than that of the other predictors. This shows once again that the JNBOOT outperforms its competitors in terms of bias. In terms of efficiency, the proposed bootstrap also has the best performance. The relative root mean squared error of the alternative MSE estimators are mostly at least three to four times as high as that of the proposed MSE estimator. Even for the particular case of EBLUP where the CCT and the CCST display extremely large biases and error rates, the proposed JNBOOT still performs the best and is within reasonable ranges. Unreported boxplots (available from the authors) further confirm the superiority of the proposed bootstrap method, JNBOOT, over all the existing alternatives considered, as in the previous scenario.

3.5. Application : County crop areas

The data used in this application are taken from Battese *et al.* (1988). They estimate the acreage of corn and soybeans of $k = 12$ counties (small areas) of North-Central Iowa from Lansat satellite images and observations from $n = 37$ segments obtained from a farms survey. The data include (a) the sample size for each area, (b) the number of acres of corn and soybeans for each unit of the sample (as collected in the survey), (c) the number of image pixels classified as corn or soybeans for each unit in the sample, and (d) the population mean of each area of pixels classified as corn or soybeans. For a detailed description of these data, see Battese *et al.* (1988).

Battese *et al.* (1988) identified an outlier and deleted it from their study. Sinha & Rao (2009) incorporated this outlier in their estimation procedure to investigate the influence of this observation on the EBLUP, and also to assess the ability of their robust method to identify and reduce the influence of this unit on the estimation. These data are of interest to us because they provide a good example in which the proposed bootstrap MSE estimator for outlier-robust predictors of small area means can be applied.

The model is given by :

$$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + v_i + e_{ij}, \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i, \tag{3.5.1}$$

where the random effects v_i are independently and identically distributed as $\mathcal{N}(0, \sigma_v^2)$; the error terms e_{ij} are independently and identically distributed as $\mathcal{N}(0, \sigma_e^2)$; the random effects v_i and the error terms e_{ij} are independent for all $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$; x_{1ij} and x_{2ij} correspond to the number of corn pixels and soybeans pixels, respectively ; and finally y_{ij} is the number of acres of corn. The parameters estimation results of Model (3.5.1) can be found in Sinha & Rao (2009).

The small area population mean is given by $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$. Tables 3.4 and 3.5 present the bootstrap MSE estimates for the small area predictors EBLUP and SR, respectively, based on $B = 1000$ bootstrap replications. The results for the proposed bootstrap MSE, JNBOOT, and the Jiongo *et al.* (2013) bootstrap estimator, JHDBOOT, are similar for all the crop areas, regardless of the predictor used. For the (non-robust) EBLUP, the results obtained for the analytical MSE estimators, PR, CCT and CCST, are quite different and usually higher than

TABLE 3.4. EBLUP Predicted hectares of corn with estimated standard errors.

County	Sample Segments	EBLUP Predictor	Standard errors PR	CCT	CCST	JHDBOOT	JNBOOT
Cerro Gordo	1	122.2	8.4	11.5	n.a.	7.7	7.8
Hamilton	1	123.2	8.4	8.8	n.a.	7.8	7.4
Worth	1	113.8	8.4	23.6	n.a.	7.7	8.0
Humboldt	2	115.4	8.5	8.6	8.7	7.6	7.4
Franklin	3	136.1	8.0	15.4	15.4	6.4	6.8
Pocanhontas	3	108.4	8.1	9.4	9.4	6.9	6.7
Winnibago	3	116.8	8.0	7.2	7.2	7.1	6.8
Wright	3	122.6	8.1	5.9	5.9	6.9	7.0
Webster	4	111.0	7.8	8.7	8.7	6.5	6.5
Hancock	5	124.4	7.4	6.0	6.0	6.0	6.1
Kossuth	5	113.4	7.4	10.7	10.7	6.1	5.9
Hardin	6	131.3	7.2	5.3	5.4	6.3	5.9

those of the bootstrap MSE estimators, JHDBOOT and JNBOOT. Note that it is not possible to calculate the analytical linearization-based MSE estimator CCST of Chambers *et al.* (2014) for the first three domains (Cerro Gordo, Hamilton, and Worth) because their sample sizes are equal to $n_i = 1$. As for the robust predictor SR, it can be noted that the standard errors of the bootstrap of Sinha & Rao (2009), SRBOOT are significantly higher than the proposed bootstrap, JNBOOT, for the first three areas for which the sample sizes are equal to $n_i = 1$. These differences decrease when the sample size of the area increases. The standard errors of the two bootstrap procedures tend to be similar when the sample size of the area is 4 and above. On the other hand, the analytical estimators CCT and CCST are quite different from the bootstrap MSE estimators, but are similar to one another except for the first three areas where the CCST estimates are not available.

The estimation results from the methods discussed clearly display considerable differences in values and standard errors and could lead to different inferences about the small area means. The analytical MSE estimators give higher standard errors compared to the standard errors produced by bootstrap MSE estimators including the proposed bootstrap JNBOOT. Hence, it is important to compare the accuracy of these methods in a rigorous way. Meanwhile, our results suggest that the latter should be preferred since its theoretical validity has just been established by our main theoretical result and it is empirically more accurate as per the simulations outcomes obtained in Section 3.4.

TABLE 3.5. SR Predicted hectares of corn with estimated standard errors.

County	Sample segments	SR Predictor	Standard errors CCT	CCST	SRBOOT	JHDBOOT	JNBOOT
Cerro Gordo	1	123.7	6.1	n.a.	9.8	7.6	7.7
Hamilton	1	125.3	7.8	n.a.	9.6	7.7	7.3
Worth	1	110.2	19.0	n.a.	9.6	7.7	7.8
Humboldt	2	114.1	7.8	7.8	8.7	7.6	7.2
Franklin	3	140.8	10.6	10.6	7.4	6.5	6.8
Pocanhontas	3	110.8	8.3	8.3	7.5	6.9	6.7
Winnibago	3	115.2	7.1	7.1	7.4	7.2	6.8
Wright	3	122.7	6.2	6.2	7.6	6.9	6.9
Webster	4	113.5	7.5	7.4	6.9	6.5	6.4
Hancock	5	124.1	5.9	5.9	6.4	6.1	6.3
Kossuth	5	109.4	8.1	8.1	6.5	6.1	6.0
Hardin	6	136.9	5.8	5.9	6.3	6.4	6.0

3.6. Concluding Remarks

In this paper, we considered the problem of bootstrapping the mean squared error of robust small area estimators. The underlying model is the unit-level model where error variance, random effects and fixed effects can be estimated using existing approaches. Given that robust estimates of the variance components are typically smaller than their nonrobust counterparts it is difficult to construct bootstrap data on the same scale as the original data (Field *et al.* 2010). We overcome this difficulty by using the nonrobust maximum likelihood estimators for generating the bootstrap samples and apply the robust estimation technique on this sample to obtain outlier-robust bootstrap predictors. It is from this starting point that our proposed MSE estimator is built.

Existing bootstrap MSE procedures that have been proposed in this literature are not justified theoretically, whereas we formally prove the theoretical validity of our proposed bootstrap. This is the first time, to our knowledge, that the asymptotic validity of a bootstrap method for MSE has been formally established for robust small area estimation. Moreover, the semi-parametric nature of the proposed method makes it particularly attractive, as it help to protect again the misspecification of the errors en the random effects. Our theoretical results are derived using an approach similar to Bickel & Freedman (1981) and Freedman (1981) and convergence results established by Huggins (1993). The proofs of the proposed bootstrap MSE estimator are provided for the robust estimator of Sinha & Rao (2009), but the argument can be easily extended to accommodate other

robust predictors.

We examined the behaviour of the proposed method through Monte Carlo simulations and compared its performance with five other methods : the bootstrap MSE estimator of Sinha & Rao (2009), the analytical pseudolinearization MSE estimator and linearization-based MSE estimator of Chambers *et al.* (2014), the bootstrap MSE of Jiongo *et al.* (2013) and the MSE estimator of Prasad & Rao (1990). The results showed that for all the different robust small area estimators and all the various modes of contamination considered, the proposed bootstrap MSE performs the best, both in terms of bias and efficiency. An empirical application using county crops area data from North-Central Iowa farms and Landsat satellite images illustrates the usefulness of the proposed method is practice.

Finally, we note that, although our bootstrap MSE estimator was developed under the linear mixed model, it should be possible to develop a version of this MSE estimator under a semiparametric mixed model that allow for possibly nonparametric effects in the dependency. This presents an avenue for further research.

Appendix : Proofs

This section provides the proofs of Conditions (3.3.8) - (3.3.11) stated in Lemma 3.

Proof of Lemma 3

Proof of (3.3.8) : Using the triangular inequality and a binomial expansion, it can be shown that

$$\frac{1}{8}d_4\left(F_v,\hat{F}_{uk}\right)^4 \leq d_4\left(F_v,F_u\right)^4 + d_4\left(F_u,\hat{F}_{uk}\right)^4$$

and

$$\frac{1}{8}d_4\left(F_u,\hat{F}_{uk}\right)^4 \leq d_4\left(F_u,F_{uk}\right)^4 + d_4\left(F_{uk},\hat{F}_{uk}\right)^4.$$

This implies that

$$d_4\left(F_v,\hat{F}_{uk}\right)^4 \leq 8d_4\left(F_v,F_u\right)^4 + 64d_4\left(F_u,F_{uk}\right)^4 + 64d_4\left(F_{uk},\hat{F}_{uk}\right)^4.$$

Notice that $\hat{u}_i = u_i + O_p(k^{-1/2})$, where $u_i = \sqrt{\rho_i}\left(v_i + \bar{e}_i\right)$ and $\bar{e}_i = \sum_{j=1}^{n_i} e_{ij}/n_i$. By the stability of the family distribution of v_i and e_{ij}, it follows that u_i is distributed as F_v.

We then have $d_4\left(F_v, F_u\right)^4 = 0$, and by Lemma 8.4 of Bickel & Freedman (1981) we also have $d_4\left(F_u, F_{uk}\right)^4 \xrightarrow{p} 0$ as $k \longrightarrow \infty$.

On the other hand, since F_{uk} and $\hat{F}_{uk}$ are two empirical distributions, this implies that $d_4\left(F_{uk}, \hat{F}_{uk}\right)^4 \leq \frac{1}{k}\sum_{i=1}^{k}\left(\hat{u}_i - u_i\right)^4 = O_p(k^{-2})$, so that $d_4\left(F_v, \hat{F}_{uk}\right) \xrightarrow{p} 0$ as $k \longrightarrow \infty$.

Likewise, we have

$$d_4\left(F_e, \hat{F}_{ek}\right)^4 \leq 8d_4\left(F_e, F_\epsilon\right)^4 + 64d_4\left(F_\epsilon, F_{\epsilon k}\right)^4 + 64d_4\left(F_{\epsilon k}, \hat{F}_{ek}\right)^4,$$

where $\epsilon_{ij} = (1-\tau_i)v_i + e_{ij} - \tau_i\bar{e}_i, i = 1, \ldots, k \quad j = 1, \ldots, n_i$.

Note that the sampling residuals are $\tilde{e}_{ij} = \hat{e}_{ij} - \frac{1}{n}\sum_{g=1}^{k}\sum_{l\in s_g}\hat{e}_{gl}$ and that the ϵ_{ij} are independent and identically distributed with the same distribution F_e. Hence $d_4\left(F_e, F_\epsilon\right) = 0$, and by lemma 8.4 of Bickel & Freedman (1981) $d_4\left(F_\epsilon, F_{\epsilon k}\right)^4 \xrightarrow{p} 0$.

Finally, notice that we can write $\hat{e}_{ij} = \epsilon_{ij} + O_p(k^{-1/2})$, which implies that $\tilde{e}_{ij} - \epsilon_{ij} = O_p(k^{-1/2})$. It follows that

$$d_4\left(F_{\epsilon k}, \hat{F}_{ek}\right)^4 \;\leq\; \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\tilde{e}_{ij} - \epsilon_{ij}\right)^4 = O_p(k^{-2}) \xrightarrow{p} 0 \quad \text{as} \quad k \longrightarrow \infty.$$

Hence, $d_4\left(F_e, \hat{F}_{ek}\right)^4 \xrightarrow{p} 0$ as $k \longrightarrow \infty$. □

Proof of (3.3.9) : Denote : $\mathbf{X}_i = \left(\mathbf{1}_{n_i}, \tilde{\mathbf{X}}_i\right)$, where $\mathbf{X}_i$ is a $n_i \times (p-1)$ matrix of auxiliary variables.

$$\mathbf{J}_{1k} = \sum_{i=1}^{k}\mathbf{I}_{1k}^{-1/2}\mathbf{X}_i^\top\mathbf{V}_i^{-1}\mathbf{U}_i^{1/2}E_m\left\{\mathbf{\Psi}_b(\mathbf{r}_i)\mathbf{\Psi}_b(\mathbf{r}_i)^\top\right\}\mathbf{U}_i^{1/2}\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{I}_{1k}^{-1/2},$$

$$a_2 = E_m\left\{\psi_b^2(r_{ij})\right\}, \quad a_{11} = E_m\left\{\psi_b(r_{ij_1})\psi_b(r_{ij_2})\right\}, \quad \text{for} \quad j_1 \neq j_2.$$

Then, a straightforward calculation shows that

$$\mathbf{J}_{1k} = \begin{pmatrix} J_{111k} & \mathbf{J}_{112k} \\ \mathbf{J}_{112k}^\top & \mathbf{J}_{122k} \end{pmatrix},$$

where

$$\begin{aligned} J_{111k} &= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\rho_i - \rho_i^2)(a_2 - a_{11})\right\}, \\ \mathbf{J}_{112k} &= \sqrt{\frac{k}{n}}\left[\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11} + \frac{\sigma_e^2}{\sigma_v^2}(\rho_i - \rho_i^2)(a_2 - a_{11})\right\}\bar{\tilde{\mathbf{X}}}_i\right], \end{aligned}$$

$$\begin{aligned}\mathbf{J}_{122k} &= \frac{k}{n}\left[\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11}+\frac{\sigma_e^2}{\sigma_v^2}(\rho_i-\rho_i^2)(a_2-a_{11})\right\}\bar{\tilde{\mathbf{X}}}_i^\top\bar{\tilde{\mathbf{X}}}_i\right]\\ &\quad+\sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{\mathbf{X}}_i^\top\tilde{\mathbf{X}}_i-\bar{\tilde{\mathbf{X}}}_i^\top\bar{\tilde{\mathbf{X}}}_i\right\},\end{aligned}$$

and $\rho_i=\dfrac{n_i\sigma_v^2}{\sigma_e^2+n_i\sigma_v^2}$.

Taking the expression to the limit as $k\longrightarrow\infty$, yields $\mathbf{J}_1=\begin{pmatrix} J_{111} & \mathbf{J}_{112}\\ \mathbf{J}_{112}^\top & \mathbf{J}_{122}\end{pmatrix}$, where

$$\begin{aligned}J_{111} &= \frac{\sigma_e^4}{\sigma_v^4}\nu_1 a_{11}+\frac{\sigma_e^2}{\sigma_v^2}(\nu_1-\nu_2)(a_2-a_{11}),\\ \mathbf{J}_{112} &= \sqrt{c}\lim_{k\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11}+\frac{\sigma_e^2}{\sigma_v^2}(\rho_i-\rho_i^2)(a_2-a_{11})\right\}\bar{\tilde{\mathbf{X}}}_i\right],\\ \mathbf{J}_{122} &= \lim_{k\to\infty}\left[c\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\sigma_e^4}{\sigma_v^4}\rho_i^2 a_{11}+\frac{\sigma_e^2}{\sigma_v^2}(\rho_i-\rho_i^2)(a_2-a_{11})\right\}\bar{\tilde{\mathbf{X}}}_i^\top\bar{\tilde{\mathbf{X}}}_i\right]\\ &\quad+\lim_{k\to\infty}\left[\sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{\mathbf{X}}_i^\top\tilde{\mathbf{X}}_i-\bar{\tilde{\mathbf{X}}}_i^\top\bar{\tilde{\mathbf{X}}}_i\right\}\right];\end{aligned}$$

with $\nu_1=\lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\rho_i$, and $\nu_2=\lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\rho_i^2$.

Using the same reasoning we can obtain the bootstrap version $\mathbf{J}_1^*$ of $\mathbf{J}_1$ defined by

$$\mathbf{J}_1^*=\begin{pmatrix} J_{111}^* & \mathbf{J}_{112}^*\\ \mathbf{J}_{112}^{*\top} & \mathbf{J}_{122}^*\end{pmatrix},$$

where $J_{111}^*=\dfrac{\hat\sigma_e^4}{\hat\sigma_v^4}\hat\nu_1 a_{11}^*+\dfrac{\hat\sigma_e^2}{\hat\sigma_v^2}(\hat\nu_1-\hat\nu_2)(a_2^*-a_{11}^*)$,

$$\begin{aligned}\mathbf{J}_{112}^* &= \sqrt{c}\lim_{k\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\hat\sigma_e^4}{\hat\sigma_v^4}\hat\rho_i^2 a_{11}^*+\frac{\hat\sigma_e^2}{\hat\sigma_v^2}(\hat\rho_i-\hat\rho_i^2)(a_2^*-a_{11}^*)\right\}\bar{\tilde{\mathbf{X}}}_i\right],\\ \mathbf{J}_{122}^* &= \lim_{k\to\infty}\left[c\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\hat\sigma_e^4}{\hat\sigma_v^4}\hat\rho_i^2 a_{11}^*+\frac{\hat\sigma_e^2}{\hat\sigma_v^2}(\hat\rho_i-\hat\rho_i^2)(a_2^*-a_{11}^*)\right\}\bar{\tilde{\mathbf{X}}}_i^\top\bar{\tilde{\mathbf{X}}}_i\right]\\ &\quad+\lim_{k\to\infty}\left[\sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{\mathbf{X}}_i^\top\tilde{\mathbf{X}}_i-\bar{\tilde{\mathbf{X}}}_i^\top\bar{\tilde{\mathbf{X}}}_i\right\}\right];\end{aligned}$$

and $\hat\nu_1=\frac{1}{k}\sum_{i=1}^{k}\hat\rho_i$, $\hat\nu_2=\frac{1}{k}\sum_{i=1}^{k}\hat\rho_i^2$, $a_2^*=E_*\left\{\psi_b^2(r_{ij}^*)\right\}$, $a_{11}^*=E_*\left\{\psi_b(r_{ij_1}^*)\psi_b(r_{ij_2}^*)\right\}$, $j_1\neq j_2$.

By Lemmas 2.1 and 8.5 of Bickel and Freedman (1981), a_{11}^* and a_2^* converge in probability to a_{11} and a_2 respectively. Lemma 1 obtained above implies that

$(\hat{\sigma}_v^2, \hat{\sigma}_e^2, \hat{\nu}_1, \hat{\nu}_2)$ converge in probability to $(\sigma_v^2, \sigma_e^2, \nu_1, \nu_2)$. Hence, by continuity, $\mathbf{J}_1^* \xrightarrow{p} \mathbf{J}_1$. □

Proof of (3.3.10) : The proof of (3.3.10) proceeds exactly as for (3.3.9). The derivation is however more cumbersome because it requires to calculate fourth-order moments. Thus after a lengthy algebraic expansion, one could write

$$\mathbf{J}_{2k} = \begin{pmatrix} J_{211k} & J_{212k} \\ J_{212k} & J_{222k} \end{pmatrix},$$

with

$$J_{211k} = \frac{1}{\sigma_e^4}\left(\frac{1-\eta}{\eta}\right)^2\left(\frac{a_{11}-\eta a_2}{1-\eta}\right)^2\left(\bar{\rho}-\frac{1}{\eta}\bar{\rho^2}\right)^2 k + A_{11k},$$

where $\eta = \dfrac{\sigma_v^2}{\sigma_e^2+\sigma_v^2}$, and A_{11k} is a bounded sequence of real numbers given by

$$\begin{aligned} A_{11k} &= \frac{1}{\sigma_e^4\eta^2}\left(a_4 - 4a_{31} - 4a_{22} - 12a_{211} - 6a_{1111}\right)\left(\bar{\rho} - 3\bar{\rho^2} + 3\bar{\rho^3} - \bar{\rho^4}\right) \\ &+ \frac{1}{\sigma_e^4\eta^2}\left\{4a_{31} + 4a_{22} - 18a_{211} + 11a_{1111} - (a_2 - a_{11})^2\right\}\left(\bar{\rho^2} - 2\bar{\rho^3} + \bar{\rho^4}\right) \\ &+ \frac{1-\eta}{\sigma_e^4\eta^3}\left\{6a_{211} - 6a_{1111} - 2a_{11}(a_2 - a_{11})\right\}\left(\bar{\rho^3} - \bar{\rho^4}\right) \\ &+ \frac{(1-\eta)^2}{\sigma_e^4\eta^4}\left(a_{1111} - a_{11}^2\right)\bar{\rho^4} \\ &= A_{11}\left(\sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \bar{\rho}, \bar{\rho^2}, \bar{\rho^3}, \bar{\rho^4}\right). \end{aligned}$$

The numbers $a_4, a_{31}, a_{22}, a_{211}, a_{1111}$ are fourth-order moments defined by

$$a_4 = E_m\left\{\psi_b^4(r_{ij})\right\}, \quad a_{31} = E_m\left\{\psi_b^3(r_{ij_1})\psi_b(r_{ij_2})\right\}, \quad a_{22} = E_m\left\{\psi_b^2(r_{ij_1})\psi_b^2(r_{ij_2})\right\},$$

$$a_{211} = E_m\left\{\psi_b^2(r_{ij_1})\psi_b(r_{ij_2})\psi_b(r_{ij_3})\right\}, \quad a_{1111} = E_m\left\{\psi_b(r_{ij_1})\psi_b(r_{ij_2})\psi_b(r_{ij_3})\psi_b(r_{ij_4})\right\},$$

where $j_1 \neq j_2$, $j_1 \neq j_3$, $j_1 \neq j_4$, , $j_2 \neq j_3$, $j_2 \neq j_4$, and $j_3 \neq j_4$,
and the numbers $\bar{\rho^l}$, $l = 1, 2, 3, 4$ are defined by $\bar{\rho^l} = \dfrac{1}{k}\sum_{i=1}^k \rho_i^l$, $l = 1, 2, 3, 4$.

Note that $A_{11}(\cdot)$, as defined above, is a continuous function of its arguments.

Likewise,

$$J_{212k} = \frac{1}{\sigma_e^4}\left(\frac{1-\eta}{\eta}\right)\left(\frac{a_{11}-\eta a_2}{1-\eta}\right)^2\left(\bar{\rho}-\frac{1}{\eta}\bar{\rho^2}\right)\left(1-\bar{\rho}\sqrt{\frac{k}{n}}\right)\sqrt{nk}$$

$$+ \quad \frac{1}{\sigma_e^4\eta^2}\left\{a_{211}-a_{1111}+a_{11}(a_2-a_{11})\right\}\sqrt{\frac{n}{k}}+A_{12k}\sqrt{\frac{k}{n}},$$

where, as for the above derivation, A_{12k} is a bounded sequence of real numbers which depends on the fourth moments of $\psi_b(r_{ij})$ and the sample moments of ρ_i^l, $l=1,2,3,4$. That is, $A_{12k}=A_{12}\left(\sigma_e^2,\eta,a_4,a_{31},a_{22},a_{211},a_{1111},\bar{\rho},\bar{\rho^2},\bar{\rho^3},\bar{\rho^4}\right)$, and $A_{12}(\cdot)$ is a continuous function of its arguments. The last component J_{222k} of matrix $\mathbf{J}_{2k}$ is given by

$$\begin{aligned}
J_{222k} &= \frac{1}{\sigma_e^4}\left(\frac{a_{11}-\eta a_2}{1-\eta}\right)^2\left\{1-\left(\bar{\rho}+\frac{1}{\eta}\bar{\rho}-\frac{1}{\eta}\bar{\rho^2}\right)\frac{k}{n}\right\}^2 n \\
&+ \frac{1}{\sigma_e^4(1-\eta)^2}\left\{a_{22}-2a_{211}+a_{1111}-(a_2-a_{11})^2\right\}\left(\frac{1}{n}\sum_{i=1}^{k}n_i^2\right) \\
&+ \frac{1}{\sigma_e^4(1-\eta)^2}\left\{a_4-3a_{22}-4a_{31}+12a_{211}-a_{1111}+2(a_2-a_{11})^2\right\}+A_{22k}\times\frac{k}{n},
\end{aligned}$$

where, as above, A_{22k} is a bounded sequence of real numbers and can be written as $A_{22k}=A_{22}\left(\sigma_e^2,\eta,a_4,a_{31},a_{22},a_{211},a_{1111},\bar{\rho},\bar{\rho^2},\bar{\rho^3},\bar{\rho^4}\right)$, where $A_{12}(\cdot)$ is also a continuous function of its arguments. Since Assumption A1 implies that $\frac{k}{n}$ converges to a possibly zero constant c, and the limit of $\mathbf{J}_{2k}$ is assumed to always exist and be finite by Assumption A4, then we must have

$$\begin{aligned}
&a_{11}-\eta a_2=0, \quad a_{211}-a_{1111}+a_{11}(a_2-a_{11})=0, \\
&\text{and} \quad a_{22}-2a_{211}+a_{1111}-(a_2-a_{11})^2=0.
\end{aligned}$$

It follows that $\mathbf{J}_2=\lim_{k\to\infty}\mathbf{J}_{2k}=\begin{pmatrix} J_{211} & J_{212} \\ J_{212} & J_{222}\end{pmatrix}$, where

$$\begin{aligned}
J_{211} &= \frac{1}{\sigma_e^4\eta^2}\left(a_4-4a_{31}-4a_{22}-12a_{211}-6a_{1111}\right)\left(\nu_1-3\nu_2+3\nu_3-\nu_4\right) \\
&+ \frac{1}{\sigma_e^4\eta^2}\left\{4a_{31}+4a_{22}-18a_{211}+11a_{1111}-(a_2-a_{11})^2\right\}\left(\nu_2-2\nu_3+\nu_4\right) \\
&+ \frac{1-\eta}{\sigma_e^4\eta^3}\left\{6a_{211}-6a_{1111}-2a_{11}(a_2-a_{11})\right\}\left(\nu_3-\nu_4\right) \\
&+ \frac{(1-\eta)^2}{\sigma_e^4\eta^4}\left(a_{1111}-a_{11}^2\right)\nu_4, \\
&= A_{11}\left(\sigma_e^2,\eta,a_4,a_{31},a_{22},a_{211},a_{1111},\nu_1,\nu_2,\nu_3,\nu_4\right),
\end{aligned}$$

$$J_{212} = \sqrt{c}A_{12}\left(\sigma_e^2,\eta,a_4,a_{31},a_{22},a_{211},a_{1111},\nu_1,\nu_2,\nu_3,\nu_4\right),$$

$$
\begin{aligned}
J_{222} &= \frac{1}{\sigma_e^4(1-\eta)^2}\left\{a_4 - 3a_{22} - 4a_{31} + 12a_{211} - a_{1111} + 2(a_2 - a_{11})^2\right\} \\
&+ cA_{22}\left(\sigma_e^2, \eta, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, \nu_1, \nu_2, \nu_3, \nu_4\right).
\end{aligned}
$$

Likewise, the bootstrap version $\mathbf{J}_2^*$ of $\mathbf{J}_2$ is given by

$$
\mathbf{J}_2^* = \lim_{k\to\infty} \mathbf{J}_{2k}^* = \begin{pmatrix} J_{211}^* & J_{212}^* \\ J_{212}^* & J_{222}^* \end{pmatrix},
$$

where

$$
\begin{aligned}
J_{211}^* &= A_{11}\left(\hat{\sigma}_e^2, \hat{\eta}, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4\right), \\
J_{212}^* &= \sqrt{c}A_{12}\left(\hat{\sigma}_e^2, \hat{\eta}, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4\right), \\
J_{222}^* &= \frac{1}{\hat{\sigma}_e^4(1-\hat{\eta})^2}\left\{a_4^* - 3a_{22}^* - 4a_{31}^* + 12a_{211}^* - a_{1111}^* + 2(a_2^* - a_{11}^*)^2\right\} \\
&+ cA_{22}\left(\hat{\sigma}_e^2, \hat{\eta}, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4\right)
\end{aligned}
$$

By Lemmas 2.1 and 8.5 of Bickel and Freedman (1981), a_{11}^*, a_2^*, a_4^*, a_{31}^*, a_{22}^*, a_{211}^*, a_{1111}^*, converge in probability to a_{11}, a_2, a_4, a_{31}, a_{22}, a_{211}, a_{1111}, respectively. Since, by Lemma 1 above $(\hat{\sigma}_v^2, \hat{\sigma}_e^2, \hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\nu}_4)$ converges in probability to $(\sigma_v^2, \sigma_e^2, \nu_1, \nu_2, \nu_3, \nu_4)$ it then follows by the continuous mapping theorem that $\mathbf{J}_2^* \xrightarrow{p} \mathbf{J}_2$. □

Proof of (3.3.11) : We use the same reasoning as for the proofs of (3.3.9) and (3.3.10). A straightforward expansion of the components of $\mathbf{G}_k(\boldsymbol{\theta})$ allows to see that we can express it as a sum of two components, one which is a bounded sequence and another which depends on a_2, a_{11}, d_2, d_{11}, where $d_2 = E_m\left\{r_{ij}\psi_b'(r_{ij})\psi_b(r_{ij})\right\}$ and $d_{11} = E_m\left\{r_{ij_1}\psi_b'(r_{ij_1})\psi_b(r_{ij_2})\right\}, j_1 \neq j_2$.

By Assumptions A1 and A5 which respectively assume that $\frac{k}{n}$ converges to a possibly zero constant $c \in [0,1]$ and that the limit $\mathbf{G}(\boldsymbol{\theta})$ of $\mathbf{G}_k(\boldsymbol{\theta})$ always exists and is finite, we must have $a_2 - a_{11} - d_2 + d_{11} = 0$. It then follows that

$$
\mathbf{G}(\boldsymbol{\theta}) = \begin{bmatrix} G_{111} & \mathbf{G}_{112} & 0 & 0 \\ \mathbf{G}_{112}^\top & \mathbf{G}_{122} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & G_{211} & G_{212} \\ 0 & 0 & G_{212} & G_{222} \end{bmatrix}, \text{ where}
$$

$$
\begin{aligned}
G_{111} &= \frac{d_1}{\sigma_e^2}\frac{1-\eta}{\eta}\nu_1, \\
\mathbf{G}_{112} &= \frac{d_1}{\sigma_e^2}\frac{1-\eta}{\eta}\sqrt{c}\lim_{k\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k}\rho_i\bar{\bar{\mathbf{X}}}_i\right],
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{G}_{122} &= \frac{d_1}{\sigma_e^2}\lim_{k\to\infty}\left[\frac{1-\eta}{\eta}c\frac{1}{k}\sum_{i=1}^{k}\rho_i\bar{\tilde{\mathbf{X}}}_i^{\top}\bar{\tilde{\mathbf{X}}}_i+\sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{\mathbf{X}}_i^{\top}\tilde{\mathbf{X}}_i-\bar{\tilde{\mathbf{X}}}_i^{\top}\bar{\tilde{\mathbf{X}}}_i\right\}\right],\\
G_{211} &= \frac{1}{\sigma_e^4}\left(\frac{1-\eta}{\eta}\right)^2\left\{(1-\eta)a_2+d_{11}\right\}\nu_2,\\
G_{212} &= \frac{1}{\sigma_e^4}\left(\frac{1-\eta}{\eta}\right)\sqrt{c}\left\{(1-\eta)a_2+d_{11}\right\}(\nu_1-\nu_2),\\
G_{222} &= \frac{a_2}{\sigma_e^2}+\frac{c}{\sigma_e^2}\left[a_2\left\{2\nu_2-3\nu_1+(\nu_1-\nu_2)\eta\right\}+\frac{1-\eta}{\eta}(\nu_1-\nu_2)d_{11}\right];
\end{aligned}
$$

and $d_1 = E_m\left\{\psi_b^{'}(r_{ij}).\right\}$

Likewise, we have $\mathbf{G}^*(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} G_{111}^* & \mathbf{G}_{112}^* & 0 & 0 \\ \mathbf{G}_{112}^{*\top} & \mathbf{G}_{122}^* & \mathbf{0} & \mathbf{0} \\ 0 & 0 & G_{211}^* & G_{212}^* \\ 0 & 0 & G_{212}^* & G_{222}^* \end{bmatrix}$, where

$$
\begin{aligned}
G_{111}^* &= \frac{d_1^*}{\hat{\sigma}_e^2}\frac{1-\hat{\eta}}{\hat{\eta}}\hat{\nu}_1,\\
\mathbf{G}_{112}^* &= \frac{d_1^*}{\hat{\sigma}_e^2}\frac{1-\hat{\eta}}{\hat{\eta}}\sqrt{c}\lim_{k\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k}\hat{\rho}_i\bar{\tilde{X}}_i\right],\\
\mathbf{G}_{122}^* &= \frac{d_1^*}{\hat{\sigma}_e^2}\lim_{k\to\infty}\left[\frac{1-\hat{\eta}}{\hat{\eta}}c\frac{1}{k}\sum_{i=1}^{k}\hat{\rho}_i\bar{\tilde{X}}_i^{\top}\bar{\tilde{X}}_i+\sum_{i=1}^{k}\frac{n_i}{n}\left\{\frac{1}{n_i}\tilde{X}_i^{\top}\tilde{X}_i-\bar{\tilde{X}}_i^{\top}\bar{\tilde{X}}_i\right\}\right],\\
G_{211}^* &= \frac{1}{\hat{\sigma}_e^4}\left(\frac{1-\hat{\eta}}{\hat{\eta}}\right)^2\left\{(1-\hat{\eta})a_2^*+d_{11}^*\right\}\hat{\nu}_2,\\
G_{212}^* &= \frac{1}{\hat{\sigma}_e^4}\left(\frac{1-\hat{\eta}}{\hat{\eta}}\right)\sqrt{c}\left\{(1-\hat{\eta})a_2^*+d_{11}^*\right\}(\hat{\nu}_1-\hat{\nu}_2),\\
G_{222}^* &= \frac{a_2^*}{\hat{\sigma}_e^2}+\frac{c}{\hat{\sigma}_e^2}\left[a_2^*\left\{2\hat{\nu}_2-3\hat{\nu}_1+(\hat{\nu}_1-\hat{\nu}_2)\hat{\eta}\right\}+\frac{1-\hat{\eta}}{\hat{\eta}}(\hat{\nu}_1-\hat{\nu}_2)d_{11}^*\right];
\end{aligned}
$$

with

$$
d_1^* = E_*\left\{\psi_b^{'}(r_{ij}^*).\right\},\ d_2^* = E_*\left\{r_{ij}^*\psi_b^{'}(r_{ij}^*)\psi_b(r_{ij}^*)\right\},\ d_{11}^* = E_*\left\{r_{ij_1}^*\psi_b^{'}(r_{ij_1}^*)\psi_b(r_{ij_2}^*)\right\}, j_1 \neq j_2,
$$

It follows by the continuous mapping theorem that $\mathbf{G}^*(\hat{\boldsymbol{\theta}})$ converges in probability to $\mathbf{G}(\boldsymbol{\theta})$ as $k\longrightarrow\infty$. □

Bibliographie

[1] Battese, G. E., Harter R. M. and W. A. Fuller (1988). An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28-36.

[2] Beaumont, J.-F., Haziza, D. and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, **100**, 555-569.

[3] Bickel, P.-J., Freedman, D.-A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196–1217.

[4] Chambers, R. L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, **81**, 1063-1069.

[5] Chambers, R., Chandra, H., Salvati, N. and Tzavidis, N. (2014), Outlier robust small area estimation. *Journal of the Royal Statistical Society, Series B*, **76**, 47-69.

[6] Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics*, **28**, 51-60.

[7] Freedman, D.-A. (1981). Bootstrapping regression models. *The Annals of Statistics*, **9**, 1218-1228.

[8] Field, C. A., Pang, Z. and Welsh, A. H. (2010). Bootstrapping robust estimates for clustered data. *Journal of American Statistical Association*, **105**, 1606–1616.

[9] Field, C. A., Pang, Z. and Welsh, A. H. (2008). Bootstrapping data with multiple levels of variation. *The Canadian Journal of Statistics*, **36**, 521–539.

[10] Huggins, R.M. (1993). On the robust analysis of variance components models for pedigree data. *Australian Journal of Statistics*, **35**, 43–57.

[11] Jiongo, V., D., Haziza, D. and Duchesne, P. (2013), Controlling the bias of robust small-area estimators. *Biometrika*, **100**, 843-858.

[12] Mallows, C.L. (1972). A note on asymptotic joint normality. *Annal of Mathematical Statistics*, **43**, 508–515.

[13] Miller, J.J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Annals of Statistics*, **5**, 746-762.

[14] Opsomer, J. P., Claeskens, G., Ranalli, M. G., Kauemann, G. and Breidt, F. J. (2008), Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society, Series B*, **70**, 265-286.

[15] Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, **28**, 1, 40-68.

[16] Prasad, N. G. N. and 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.

[17] Rao, J. N. K. (2003). *Small Area Estimation*. Wiley, Hoboken, NJ.

[18] Rao, J. N. K. (2005). Inferential issues in small area estimation : some new developments. *Statistics in Transition* **7**, 513-526.

[19] Sinha, S.K., and Rao, J. N.K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, **37**, 381–399.

[20] Stahel, W. A. and A. Welsh (1997). Approaches to robust estimation in the simplest variance components Model. *Journal of Statistical Planning and Inference*, **57**, 295-319.

[21] Weiss, L. (1971). Asymptotic properties of maximum likelihood estimators in some nonstandard cases. *Journal of the American Statistical Association*, **66**, 345-350.

[22] Weiss, L. (1973). Asymptotic properties of maximum likelihood estimators in some nonstandard cases II. *Journal of the American Statistical Association*, **68**, 428-430.

[23] Weiss, L. (1975). The asymptotic distribution of the likelihood ratio in some nonstandard cases. *Journal of the American Statistical Association*, **70**, 204-208.

Chapitre 4

TRIPLE ROBUSTESSE EN PRÉSENCE DE DONNÉES IMPUTÉES

Resumé

L'imputation est la principale méthode de traitement de la non-réponse partielle dans les enquêtes. Elle consiste à remplacer les valeurs manquantes par des valeurs artificielles. L'imputation par la régression déterministe, qui inclut l'imputation par le ratio et l'imputation par la moyenne est souvent utilisée dans les enquêtes. Ces méthodes d'imputation peuvent être biaisées si le modèle d'imputation ou le modèle de non-réponse n'est pas correctement spécifié. Des estimateurs imputés qui possèdent les propriétés de double robustesse ont alors été développés. Cependant, en présence des valeurs aberrantes, ces estimateurs doublement robustes peuvent être très instables. En utilisant le concept de biais conditionnel comme mesure de l'influence (Beaumont, Haziza & Ruiz-Gazen, 2013), nous proposons une version robuste aux valeurs aberrantes. Cet estimateur qui est fondé sur un estimateur imputé possédant les propriétés de double robustesse est donc triplement robuste. Les résultats des études de simulation sont présentés.

Mots clés : Robustesse, Imputation, Influence, Biais conditionnel

4.1. Introduction

En présence de non-réponse, les estimateurs non ajustés pour la non-réponse peuvent être fortement biaisés si les répondants diffèrent des non-répondants au regard des variables étudiées et si les taux de non-réponse sont grands. L'objectif premier des méthodes de traitement de la non-réponse est donc de réduire le biais de non-réponse. Dans cet article, nous considérons le problème de la non-réponse partielle, qui est, dans la plupart des cas, traité par imputation.

En l'absence d'erreurs non dues à l'échantillonnage (par exemple, erreurs de couverture, erreurs de non-réponse, etc), les statisticiens d'enquêtes utilisent des

procédures d'estimation qui sont (asymptotiquement) sans biais sous le plan. Autrement dit, la validité de ces estimateurs ne dépend pas de la validité d'un modèle sous-jacent. En présence de non-réponse, l'usage de modèle est incontournable. On distingue deux types de modèles : le modèle de non-réponse qui est un ensemble d'hypothèses à propos du mécanisme (inconnu) de non-réponse et le modèle d'imputation qui est un ensemble d'hypothèses à propos de la distribution de la variable que l'on cherche à imputer.

Dans cet article, nous considérons les procédures d'imputation doublement robustes. Une procédure d'imputation est dite doublement robuste si l'estimateur imputé résultant est asymptotiquement sans biais et consistant lorsque le modèle de non-réponse ou le modèle d'imputation est correctement spécifié. Une procédure doublement robuste conduit donc à un estimateur asymptotiquement sans biais si l'un des deux modèles est correctement spécifié. Dans le contexte des enquêtes, les procédures doublement robustes ont été étudiées, entre autre, par Kott (1994), Haziza & Rao (2006), Haziza & Picard (2012) et Kim & Haziza (2014).

Bien que les procédures d'imputation offrent une certaine protection contre la mauvaise spécification de l'un des deux modèles, les estimateurs imputés sont sensibles à la présence d'unités influentes. Une unité influente fait partie de la population d'intérêt. Ces dernières sont fréquentes lorsque les variables d'intérêt sont fortement asymétriques ; par exemple, les variables de revenu. Les unités influentes ont généralement un impact important sur la volatilité (variance) des estimateurs. Le but sera donc de réduire l'influence de ces unités, ce qui mènera à des estimateurs biaisés mais stables. Dans cet article, nous proposons des estimateurs robustes à la présence d'unités influentes. Ces derniers ont une forme similaire à l'estimateur robuste proposé par Beaumont, Haziza & Ruiz-Gazen (2013) dans le cas de données complètes. Les estimateurs proposés dans cet article sont dits triplement robustes car ils présentent la propriété de double robustesse mais ils sont également robustes à la présence d'unités influentes.

4.2. Notation et cadres de travail

Soit U une population finie de taille N. Nous cherchons à estimer le total dans la population $Y = \sum_{i \in U} y_i$, où y_i désigne la i-ème valeur de la variable d'intérêt y, $i = 1, \cdots, N$. Un échantillon s, de taille n, est sélectionné selon un plan de sondage $p(s)$. Soit $d_i = 1/\pi_i$, le poids de sondage de l'unité i, où π_i désigne sa probabilité d'inclusion d'ordre 1 dans l'échantillon. En l'absence de non-réponse, un estimateur basé sur les données complètes est donné par l'estimateur par

dilatation

$$\hat{Y}_\pi \quad = \quad \sum_{i \in U} d_i I_i y_i, \tag{4.2.1}$$

où I_i est une variable indicatrice de sélection de l'unité i telle que $I_i = 1$ si $i \in s$ et $I_i = 0$, sinon. L'estimateur (4.2.1) est sans biais sous le plan pour Y. Autrement dit, $E_p(\hat{Y}_\pi) = Y$, où $E_p(.)$ désigne l'espérance par rapport au plan de sondage $p(s)$.

Quand certaines des valeurs de la variable d'intérêt y sont manquantes, un estimateur de Y est l'estimateur imputé :

$$\hat{Y}_I = \sum_{i \in U} d_i r_i I_i y_i + \sum_{i \in U} d_i (1 - r_i) I_i y_i^*, \tag{4.2.2}$$

où r_i est une variable indicatrice de réponse de l'unité i telle que $r_i = 1$ si l'unité i a répondu à la variable y et $r_i = 0$, sinon, et y_i^* désigne la valeur imputée pour remplacer la valeur manquante y_i. On note également s_r et s_m les ensembles de répondants et de non-répondants respectivement.

4.2.1. Les modèles en présence

Nous présentons maintenant le modèle de non-réponse ainsi que le modèle d'imputation à l'aide desquels nous construirons les valeurs imputées et étudierons les propriétés de l'estimateur (4.2.2).

D'une part, nous supposons que les unités répondent indépendamment les unes des autres et que la probabilité de réponse p_i, de l'unité i peut être modélisée au moyen d'un modèle paramétrique

$$p_i = Prob(r_i = 1) = p(\mathbf{z}_i, \boldsymbol{\alpha}_0), \tag{4.2.3}$$

où $\boldsymbol{\alpha}_0$ est un vecteur de paramètres inconnus et $\mathbf{z}$ est un vecteur de variables auxiliaires disponibles pour toutes les unités échantillonnées (répondants et non-répondants). Le modèle (4.2.3) est appelé *modèle de non-réponse*. Un cas particulier de (4.2.3) est le modèle logistique

$$p_i = \frac{\exp\left(\mathbf{z}_i^\top \boldsymbol{\alpha}_0\right)}{1 + \exp\left(\mathbf{z}_i^\top \boldsymbol{\alpha}_0\right)}.$$

D'autre part, nous supposons que la variable d'intérêt obéit au modèle suivant :

$$y_i = m\left(\mathbf{z}_i, \boldsymbol{\beta}_0\right) + \epsilon_i, \quad i = 1, \ldots, N, \tag{4.2.4}$$

où $\boldsymbol{\beta}_0$ est un vecteur de paramètres inconnus et ϵ_i, $i = 1, \ldots, N$, sont des variables aléatoires indépendantes satisfaisant

$$E_m(\epsilon_i) = 0, \quad E_m(\epsilon_i^2) = \sigma^2 c_i, \quad E_m(\epsilon_i \epsilon_j) = 0 \quad i \neq j,$$

et $E_m(.)$ désigne l'espérance par rapport au modèle (4.2.4) et $c_i = \gamma(\mathbf{z}_i)$. La fonction $\gamma(.)$ est supposée connue. Le modèle (4.2.4) est appelé *modèle d'imputation*.

Dans cet article, nous considérons le cas de l'imputation par la régression déterministe doublement robuste pour laquelle la valeur imputée y_i^* est donnée par

$$y_i^* = m(\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r) \quad i \in s_m, \tag{4.2.5}$$

où $\hat{\boldsymbol{\beta}}_r$ est obtenu comme solution de l'équation estimante

$$\hat{U}_r\left(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}\right) = N^{-1} \sum_{i \in U} d_i r_i I_i \left(\hat{p}_i^{-1} - 1\right) c_i^{-1} \left\{y_i - m\left(\mathbf{z}_i, \boldsymbol{\beta}\right)\right\} h(\mathbf{z}_i, \boldsymbol{\beta}) = 0. \tag{4.2.6}$$

Dans l'équation précédente, $\hat{p}_i = p(\mathbf{z}_i, \hat{\boldsymbol{\alpha}})$ désigne la probabilité de réponse estimée pour l'unité i et $\hat{\boldsymbol{\alpha}}$ est un estimateur de $\boldsymbol{\alpha}_0$ solution de l'équation estimante

$$\hat{\mathbf{S}}(\boldsymbol{\alpha}) = N^{-1} \sum_{i \in s} d_i \left\{r_i - p(\mathbf{z}_i, \boldsymbol{\alpha})\right\} \mathbf{l}_i(\boldsymbol{\alpha}) = \mathbf{0}, \tag{4.2.7}$$

où $\mathbf{l}_i(\boldsymbol{\alpha}) = \frac{\partial ln\left(\frac{p_i}{1-p_i}\right)}{\partial \boldsymbol{\alpha}}$ voir Beaumont (2005), Haziza et Rao (2006) et Kim et Kim (2007).

4.2.2. Approches pour l'inférence

Afin d'étudier les propriétés des estimateurs imputés (par exemple, biais et variance), nous considérons deux approches distinctes : l'approche par modèle de non-réponse (NM) et l'approche par modèle d'imputation (IM). Avant de présenter ces deux approches, il convient de décrire les trois sources d'aléa sous-jacentes : (i) le modèle d'imputation qui génère le vecteur $\mathbf{y} = (y_1, ..., y_N)'$; (ii) le plan de sondage qui génère le vecteur $\mathbf{I} = (I_1, ..., I_N)'$ et (iii) le mécanisme de non-réponse qui génère le vecteur $\mathbf{r} = (r_1, ..., r_N)'$.

Dans le cas de l'approche NM, l'inférence est menée par rapport à la distribution conjointe du plan de sondage et du modèle de non-réponse (4.2.3). Notons, que le vecteur $\mathbf{y}$ est traité comme fixe. Cette approche a été étudiée, entre autre, par Rao (1996), Shao & Steel (1999), Beaumont (2005), Kim & Park (2006), Haziza & Rao (2006) et Haziza (2009).

Dans le cas de l'approche IM, l'inférence est menée par rapport à la distribution conjointe du plan de sondage et du modèle d'imputation (4.2.4). Notons qu'il n'est pas nécessaire de postuler un modèle de non-réponse explicite (comme,

par exemple, le modèle (4.2.3)). Cependant, on suppose que le mécanisme de non-réponse est ignorable. Autrement dit, on supposera que

$$E_m\left(y_i \mid \mathbf{z}_i, r_i = 1\right) = E_m\left(y_i \mid \mathbf{z}_i, r_i = 0\right).$$

L'approche IM a été étudiée, entre autre, par Särndal (1992), Shao & Steel (1999), Brick, Kalton & Kim (2004) et Haziza (2009).

4.2.3. Décomposition de l'erreur totale et biais de non-réponse

L'erreur totale de $\hat{Y}_I$, $\hat{Y}_I - Y$, peut-être décomposée comme suit :

$$\hat{Y}_I - Y = \left(\hat{Y}_\pi - Y\right) + \left(\hat{Y}_I - \hat{Y}_\pi\right). \tag{4.2.8}$$

Le terme $\hat{Y}_\pi - Y$ en (4.2.8) désigne l'erreur due à l'échantillonnage alors que le terme $\hat{Y}_I - \hat{Y}_\pi$ désigne l'erreur due à la non-réponse.

Dans le cas de l'approche NM, le biais de l'estimateur imputé $\hat{Y}_I$ est défini selon

$$Biais\left(\hat{Y}_I\right) = E_p E_q\left(\hat{Y}_I - Y|\mathbf{I}\right) = E_p B_q\left(\hat{Y}_I|\mathbf{I}\right),$$

où $B_q\left(\hat{Y}_I|\mathbf{I}\right) = E_q\left(\hat{Y}_I - \hat{Y}_\pi|\mathbf{I}\right)$ désigne le biais conditionnel de non-réponse sous l'approche NM et $E_q(.)$ désigne l'espérance par rapport au modèle de non-réponse (4.2.3). Lorsque $E_p B_q\left(\hat{Y}_I|\mathbf{I}\right) = 0$, on dira que l'estimateur imputé $\hat{Y}_I$ est sans biais par rapport à pq.

Dans le cas de l'approche IM, le biais de l'estimateur imputé $\hat{Y}_I$ est défini selon

$$Biais\left(\hat{Y}_I\right) = E_{mpq}\left(\hat{Y}_I - Y\right) = E_{pqm}\left(\hat{Y}_I - Y|\mathbf{I}, \mathbf{r}\right) = E_{pq} B_m\left(\hat{Y}_I|\mathbf{I}, \mathbf{r}\right),$$

où $B_m\left(\hat{Y}_I|\mathbf{I}, \mathbf{r}\right) = E_m\left(\hat{Y}_I - \hat{Y}_\pi|\mathbf{I}, \mathbf{r}\right)$ désigne le biais conditionnel de non-réponse sous l'approche IM. Lorsque $E_{pq} B_m\left(\hat{Y}_I|\mathbf{I}, \mathbf{r}\right) = 0$, on dira que l'estimateur imputé $\hat{Y}_I$ est sans biais par rapport à mpq.

Sous certaines conditions de régularité, on peut montrer que l'estimateur imputé $\hat{Y}_I$ donné par (4.2.2) obtenu au moyen des valeurs imputées (4.2.5), est asymptotiquement sans biais et consistant pour Y si le modèle de non-réponse (4.2.3) est correctement spécifié et/ou le modèle d'imputation (4.2.4) est correctement spécifié ; voir Haziza et Rao (2006) et Kim et Rao (2011). Autrement dit, l'estimateur $\hat{Y}_I$ est doublement robuste. Bien qu'il possède la propriété de double robustesse, ce dernier est sensible à la présence de valeurs influentes. Il s'agira donc de rendre robuste $\hat{Y}_I$, ce qui nous amène à discuter du concept d'influence d'une unité.

4.3. Résultats théoriques préliminaires

Avant d'introduire le concept d'influence, nous présentons des résultats théoriques préliminaires qui nous seront utiles dans la suite.

4.3.1. Approche du modèle de non-réponse (NM)

Le théorème 4.3.1 fournit une approximation de l'erreur due à la non-réponse sous l'approche NM.

Théorème 4.3.1. *Sous les conditions de régularité données à l'annexe A.1, on a :*

$$\begin{aligned}\hat{Y}_I - \hat{Y}_\pi &= \sum_{i\in s} d_i \left\{r_i g_i(p) - 1\right\} \left\{y_i - m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)\right\} \\ &\quad + \sum_{i\in s} d_i \nu_i (r_i - p_i) + o_p(Nn^{-1/2}), \end{aligned} \tag{4.3.1}$$

où l'ordre de grandeur est par rapport à la distribution conjointe du modèle de non-réponse et du plan d'échantillonnage et où $\mathbf{B}_{\pi p}$ est solution de l'équation estimante

$$U_{pq}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_0\right) = E_{pq}\left\{U_r\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_0\right)\right\} = 0$$

et les facteurs d'ajustement $g_i(p)$ et ν_i sont respectivement donnés par (A.1) et (A.2) en annexe.

Démonstration. Voir l'appendice pour la preuve. □

4.3.2. Approche du modèle d'imputation (IM)

Le théorème 4.3.2 fournit une approximation de l'erreur due à la non-réponse sous l'approche IM.

Théorème 4.3.2. *Sous les conditions de régularité suffisantes similaires à celles du théorème 4.3.1 , on a :*

$$\hat{Y}_I - \hat{Y}_\pi = \sum_{i\in s} d_i \left\{r_i g_i(r) - 1\right\} \left\{y_i - m\left(\mathbf{z}_i, \boldsymbol{\beta}_0\right)\right\} + o_p(Nn^{-1/2}), \tag{4.3.2}$$

où l'ordre de grandeur est par rapport à la distribution conjointe du modèle d'imputation et du plan d'échantillonnage et où le facteur d'ajustement $g_i(r)$ est donné en annexe.

Démonstration. La preuve n'est pas donnée, car elle est similaire à celle du théorème 4.3.1. Voir l'appendice pour la preuve du théorème 4.3.1. □

4.4. Influence d'une unité : utilisation du biais conditionnel

En l'absence de non-réponse, Moreno-Rebollo, Munoz-Reyes & Munoz-Pichardo (1999) ont proposé le biais conditionnel d'une unité comme mesure d'influence ; voir aussi Beaumont, Haziza & Ruiz-Gazen (2013). Le biais conditionnel de l'unité échantillonnée i par rapport à l'estimateur par dilatation $\hat{Y}_\pi$ est défini par :

$$\begin{aligned} B_i^\pi(I_i=1) &= E_p(\hat{Y}_\pi - Y|I_i=1) \\ &= (d_i-1)y_i + \sum_{\substack{j\in U \\ j\neq i}} \left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right) y_j \\ &= \sum_{j\in U} \left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right) y_j. \end{aligned}$$

Le biais conditionnel $B_i^\pi(I_i=1)$ est une mesure de l'influence de l'unité i. Plus une unité a une grande influence, plus elle a un impact important sur la volatilité (variance) d'un estimateur. Notons que $B_i^\pi(I_i=1)=0$ lorsque $\pi_i=1$. Autrement dit, une unité sélectionnée avec probabilité 1 n'a aucune influence sur la volatilité de $\hat{Y}_\pi$. De plus, notons qu'une unité non-échantillonnée peut également avoir une grande influence. Le biais conditionnel de l'unité non-échantillonnée i par rapport à l'estimateur par dilatation $\hat{Y}_\pi$ est défini par :

$$B_i^\pi(I_i=0) = E_p(\hat{Y}_\pi - Y|I_i=0) = -\frac{1}{d_i-1}B_i^\pi(I_i=1).$$

Cependant, à l'étape de l'estimation, seule l'influence des unités échantillonnées peut être réduite et rien ne peut être fait pour les unités non-échantillonnées.

Dans le cas du plan de Poisson, du plan stratifié aléatoire simple sans remise et des plans à grande entropie, Beaumont, Haziza & Ruiz-Gazen (2013) ont montré que l'erreur due à l'échantillonnage peut s'écrire (exactement ou approximativement) comme suit :

$$\hat{Y}_\pi - Y = \sum_{i\in s} B_i^\pi\left(I_i=1\right) + \sum_{i\in U/s} B_i^\pi\left(I_i=0\right).$$

Autrement dit, le biais conditionnel d'une unité peut s'interpréter comme la contribution d'une unité à l'erreur due à l'échantillonnage.

Dans les prochaines sous-sections, nous définissons deux mesures d'influence par rapport à l'estimateur imputé $\hat{Y}_I$: l'une sous l'approche NM et l'autre sous l'approche IM.

4.4.1. Biais conditionnel d'une unité sous l'approche NM

Dans cette section, nous définissons l'influence d'une unité répondante sous l'approche NM par rapport à l'estimateur imputé $\hat{Y}_I$ obtenu au moyen des valeurs imputées (4.2.5). Le biais conditionnel de l'unité répondante i est défini par :

$$B_{qi}^I\left(I_i=1,r_i=1\right)=E_{pq}\left(\hat{Y}_I-Y\Big|I_i=1,r_i=1\right). \tag{4.4.1}$$

Il découle de la décomposition (4.2.8) que le biais conditionnel de l'unité répondante i, donné par (4.4.1), peut s'écrire comme suit :

$$B_{qi}^I\left(I_i=1,r_i=1\right)=E_p\left(\hat{Y}_\pi-Y\Big|I_i=1\right)+E_{pq}\left(\hat{Y}_I-\hat{Y}_\pi\Big|I_i=1,r_i=1\right). \tag{4.4.2}$$

Le premier terme à droite de l'égalité (4.4.2) représente la contribution (ou l'impact) de l'unité i à l'erreur due à l'échantillonnage $\hat{Y}_\pi-Y$, alors que le deuxième terme représente la contribution de l'unité répondante i à l'erreur due à la non-réponse $\hat{Y}_I-\hat{Y}_\pi$, sous l'approche NM. Le biais conditionnel (4.4.2) peut donc s'interpréter comme la contribution de l'unité répondante i à l'erreur totale, $\hat{Y}_I-Y$, sous l'approche NM. En ignorant les termes d'ordre inférieur, il découle de (4.3.1) que le biais conditionnel en (4.4.2) peut être approximé par

$$\begin{aligned} B_{qi}^I\left(I_i=1,r_i=1\right) \;\approx\; & \sum_{j\in U}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)y_j \\ & +d_i(1-p_i)\left[\nu_i+g_i(p)\left\{y_i-m\left(\mathbf{z}_i,\mathbf{B}_{\pi p}\right)\right\}\right]. \end{aligned} \tag{4.4.3}$$

L'unité i a une grande contribution à l'erreur due à la non-réponse lorsque (i) son poids de sondage est grand et/ou (ii) sa probabilité de réponse est petite et/ou (iii) le facteur $g_i(p)$ est grand et/ou (iv) son résidu $y_i-m\left(\mathbf{z}_i^\top,\mathbf{B}_{\pi p}\right)$ est grand. Lorsque $p_i=1$, notons que le deuxième terme à droite de l'égalité en (4.4.3) est égal à 0. Dans ce cas, l'unité i n'a aucune influence sur l'erreur due à la non-réponse.

4.4.2. Biais conditionnel d'une unité sous l'approche IM

Dans cette section, nous définissons l'influence d'une unité répondante sous l'approche IM par rapport à l'estimateur imputé $\hat{Y}_I$ obtenu au moyen des valeurs imputées (4.2.5). Le biais conditionnel de l'unité répondante i est défini par :

$$B_{mi}^I\left(y_i,I_i=1,r_i=1\right)=E_{mpq}\left(\hat{Y}_I-Y\Big|y_i,I_i=1,r_i=1\right). \tag{4.4.4}$$

Il découle de la décomposition (4.2.8) que le biais conditionnel de l'unité répondante i, donné par (4.4.4), peut s'écrire comme

$$B_{mi}^I\left(y_i,I_i=1,r_i=1\right) \;=\; E_mE_p\left(\hat{Y}_\pi-Y\Big|y_i,I_i=1\right)$$

$$+E_q E_p E_m \left(\hat{Y}_I - \hat{Y}_\pi \middle| y_i, I_i = 1, r_i = 1\right). \quad (4.4.5)$$

Le premier terme à droite de l'égalité (4.4.5) représente la contribution (ou l'impact) de l'unité i à l'erreur due à l'échantillonnage, $\hat{Y}_\pi - Y$, alors que le deuxième terme représente la contribution de l'unité répondante i à l'erreur due à la non-réponse, $\hat{Y}_I - \hat{Y}_\pi$ sous l'approche IM. Encore une fois, le biais conditionnel peut donc s'interpréter comme la contribution de l'unité répondante i à l'erreur totale, $\hat{Y}_I - Y$, sous l'approche IM. En ignorant les termes d'ordre inférieur, il découle de (4.3.2) que le biais conditionnel en (4.4.5) peut être approximé par

$$\begin{aligned} B_{mi}^I (y_i, I_i = 1, r_i = 1) &\approx E_m \left\{ \sum_{j \in U} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) y_j \middle| y_i \right\} \\ &+ E_q \left[d_i \left\{ r_i g_i(r) - 1 \right\} \left\{ y_i - m \left(\mathbf{z}_i, \boldsymbol{\beta}_0 \right) \right\} \middle| r_i = 1 \right]. \quad (4.4.6) \end{aligned}$$

L'unité i a une grande contribution à l'erreur due à la non-réponse lorsque (i) son poids de sondage est grand et/ou (ii) le facteur $g_i(r)$ est grand et/ou (iii) son résidu $y_i - m\left(\mathbf{z}_i, \boldsymbol{\beta}_0\right)$ est grand.

4.5. Estimateur imputé robuste à la présence de valeurs influentes

En l'absence de non-réponse, Beaumont, Haziza & Ruiz-Gazen (2013) ont proposé une version robuste de l'estimateur par dilatation :

$$\hat{Y}_\pi^R = \hat{Y}_\pi - \sum_{i \in s} \hat{B}_i^\pi (I_i = 1) + \sum_{i \in s} \psi_c \left\{ \hat{B}_i^\pi (I_i = 1) \right\}, \quad (4.5.1)$$

où $\hat{B}_i^\pi (I_i = 1)$ est un estimateur de $B_i^\pi (I_i = 1)$ obtenu en remplaçant les paramètres inconnus par des estimateurs robustes et $\psi_c(.)$ est une fonction dont le rôle est de réduire l'influence des unités qui ont une grande influence. Une fonction $\psi_c(.)$ populaire est la fonction de Huber donnée par :

$$\psi_c(t) = \begin{cases} c & \text{if } t > c \\ t & \text{if } |t| \leq c \\ -c & \text{if } t < -c \end{cases}$$

où c est une constante à déterminer. Nous faisons les remarques suivantes : (i) l'estimateur $\hat{Y}_\pi^R$ est consistant au sens de Cochran. Autrement dit, lorsque $s = U$, on a $\hat{Y}_\pi^R = Y$. (ii) Lorsque $c \to \infty$, l'estimateur robuste $\hat{Y}_\pi^R$ tend vers l'estimateur non-robuste $\hat{Y}_\pi$. (iii) Dans le cas stratifié aléatoire simple sans remise, $\hat{Y}_\pi^R$ coïncide (à un facteur près) avec l'estimateur de Kokic et Bell (1994).

4.5.1. Estimateur triplement robuste sous l'approche NM

Suivant Beaumont, Haziza & Ruiz-Gazen (2013), une version robuste de $\hat{Y}_I$ sous l'approche NM est donnée par

$$\hat{Y}_I^R = \hat{Y}_I - \sum_{i\in s_r} \hat{B}_{qi}^I(I_i=1,r_i=1) + \sum_{i\in s_r} \psi_c\left\{\hat{B}_{qi}^I(I_i=1,r_i=1)\right\}, \quad (4.5.2)$$

où $\hat{B}_{qi}^I(I_i=1,r_i=1)$ est un estimateur de $B_{qi}^I(I_i=1,r_i=1)$ donné par (4.4.3) obtenu en remplaçant les paramètres inconnus par des estimateurs robustes. Nous faisons les remarques suivantes : (i) En l'absence de non-réponse, $s_r = s$, l'estimateur (4.5.2) coïncide avec l'estimateur robuste (4.5.1). (ii) Lorsque $c \to \infty$, l'estimateur imputé robuste, $\hat{Y}_I^R$, tend vers l'estimateur imputé doublement robuste $\hat{Y}_I$.

4.5.2. Estimateur triplement robuste sous l'approche IM

De manière similaire, une version robuste de $\hat{Y}_I$ sous l'approche IM est donnée par :

$$\hat{Y}_I^R = \hat{Y}_I - \sum_{i\in s_r} \hat{B}_{mi}^I(y_i,I_i=1,r_i=1) + \sum_{i\in s_r} \psi_c\left\{y_i,\hat{B}_{mi}^I(I_i=1,r_i=1)\right\}, \quad (4.5.3)$$

où $\hat{B}_{mi}^I(y_i,I_i=1,r_i=1)$ est un estimateur de $B_{mi}^I(y_i,I_i=1,r_i=1)$ donné par (4.4.6) obtenu en remplaçant les paramètres inconnus par des estimateurs robustes. Les remarques faites à propos de l'estimateur robuste (4.5.2) s'appliquent également à l'estimateur robuste (4.5.3).

4.6. Estimation de l'erreur quadratique moyenne

Traditionnellement, il y a deux approches pour l'estimation de la variance en présence des données imputées : (i) l'approche deux phases (Särndal, 1992 ; Rao 1996) et l'approche renversée (Shao & Steel 1999 ; Kim & Rao 2009). Dans l'approche deux phases, la non-réponse est vue comme une seconde phase du plan d'échantillonnage. Il en résulte une décomposition de la variance comme une somme de la variance d'échantillonnage de la première phase et la variance due à la non-réponse (deuxième phase). Dans l'approche renversée la population peut être subdivisée en deux domaines : la population des répondants $\mathcal{U}_r$ et la population des non-répondants $\mathcal{U}_m$. Ici, le mécanisme de réponse n'est plus aléatoire.

Nous utiliserons l'approche deux phases pour illustrer la méthode d'estimation de l'erreur quadratique moyenne (EQM) dans le cas du modèle de non-réponse. Une démarche similaire basée sur l'approche renversée permet de dériver un estimateur de l'EQM dans le cas du modèle d'imputation.

L'estimateur de l'erreur quadratique moyenne proposé est fondé sur la méthode de Gwet & Rivest (1992) : nous déterminerons séparément un estimateur du carré du biais et un estimateur de la variance. L'estimateur de l'EQM sera la somme des deux estimateurs obtenus précédemment. Outre les hypothèses du théorème 4.3.1, nous formulons l'hypothèse suivante à propos de l'estimateur du biais conditionnel :

Hypothèse H.1

$$\pi_i \hat{B}_{qi}^I = \pi_i B_{qi}^I + O_p(n^{-1/2}).$$

L'hypothèse **H.1** est satisfaite par les estimateurs du biais conditionnel proposés par Beaumont *et al.* (2013) ; c'est donc une hypothèse réaliste. L'hypothèse **H.1** et le théorème 4.3.1 conduisent à l'expression (4.6.1) suivante :

$$\hat{Y}_I^R \;=\; \tilde{Y}_I^R + O_p(Nn^{-1/2}). \tag{4.6.1}$$

avec

$$\tilde{Y}_I^R = \sum_{i \in s} d_i \left[\eta_i - \pi_i r_i \left\{ B_{qi}^I - \psi_c \left(B_{qi}^I \right) \right\} \right], \tag{4.6.2}$$

où

$$\eta_i = m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right) + r_i g_i(p) \left\{ y_i - m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right) \right\} + \nu_i (r_i - p_i).$$

L'expression (4.6.1) montre que $\hat{Y}_I^R$ est asymptotiquement équivalent à $\tilde{Y}_I^R$ et les biais asymptotiques vérifient :

$$B_{pq}(\hat{Y}_I^R) = B_{pq}(\tilde{Y}_I^R) + O(Nn^{-1/2}).$$

Par ailleurs les variances $V_{pq}(\hat{Y}_I^R)$ et $V_{pq}(\tilde{Y}_I^R)$ sont asymptotiquement équivalentes, c'est-à-dire

$$V_{pq}(\hat{Y}_I^R) = V_{pq}(\tilde{Y}_I^R) + O\left(\frac{N^2}{n}\right).$$

Dans cette thèse, nous ne faisons pas la preuve de cette affirmation. Nous envisageons de le montrer dans nos travaux futurs en utilisant une démarche similaire à Kim & Rao (2009). Ces équivalences asymptotiques des biais et variances impliquent que les erreurs quadratiques moyennes $EQM_{pq}(\hat{Y}_I^R)$ et $EQM_{pq}(\tilde{Y}_I^R)$ sont également asymptotiquement équivalentes. De ce fait, nous procéderons à l'estimation de l'EQM du pseudo estimateur $\tilde{Y}_I^R$ et ensuite, nous obtiendrons un estimateur de type «plug-in» pour l'erreur quadratique moyenne $EQM_{pq}(\hat{Y}_I^R)$ en remplaçant dans l'expression de $EQM_{pq}(\tilde{Y}_I^R)$ les valeurs de η_i, B_{qi}^I et p_i par leur estimateur respectif $\hat{\eta}_i$, $\hat{B}_{qi}^I$ et $\hat{p}_i$.

4.6.1. Estimation du carré du biais

Le biais $B_{pq}(\tilde{Y}_I^R)$ de $\tilde{Y}_I^R$ est donné par :

$$\begin{aligned} B_{pq}(\tilde{Y}_I^R) &= E_{pq}\left[\sum_{i\in s} r_i\left\{-B_{qi}^I + \psi_c\left(B_{qi}^I\right)\right\}\right] \\ &= E_p\left(E_q\left[\sum_{i\in s} r_i\left\{-B_{qi}^I + \psi_c\left(B_{qi}^I\right)\right\}\right]\right). \end{aligned}$$

Il est à noter que le biais conditionnel donné par l'expression (4.4.3) ne dépend pas du vecteur des réponses $\mathbf{r} = (r_1, \ldots, r_n)^\top$. De ce fait, le biais $B_{pq}(\tilde{Y}_I^R)$ de $\tilde{Y}_I^R$ devient

$$B_{pq}(\tilde{Y}_I^R) = \sum_{i\in\mathcal{U}} p_i\left\{-B_{qi}^I + \psi_c\left(B_{qi}^I\right)\right\}.$$

Un estimateur du carré du biais est donné par :

$$\hat{B}_{pq}^2(\tilde{Y}_I^R) = \left[\sum_{i\in s} r_i\left\{-B_{qi}^I + \psi_c\left(B_{qi}^I\right)\right\}\right]^2. \tag{4.6.3}$$

Pour simplifier l'écriture, posons : $A_i = B_{qi}^I - \psi_c\left(B_{qi}^I\right)$, $\hat{A}_i = \hat{B}_{qi}^I - \psi_c\left(\hat{B}_{qi}^I\right)$ et $E_i(\mathbf{z}_i, \boldsymbol{\beta}) = y_i - m(\mathbf{z}_i, \boldsymbol{\beta})$.

L'estimateur du carré du biais (4.6.3) est biaisé. Un estimateur sans biais du carré du biais est donné par :

$$\tilde{B}_{pq}^2(\tilde{Y}_I^R) = \left(\sum_{i\in s} r_i A_i\right)^2 - \hat{V}_{pq}\left(\sum_{i\in s} r_i A_i\right), \tag{4.6.4}$$

où $\hat{V}_{pq}\left(\sum_{i\in s} r_i A_i\right)$ est donné par (4.6.5). En effet, on a :

$$\begin{aligned} V_{pq}\left(\sum_{i\in s} r_i A_i\right) &= V_p\left\{\sum_{i\in s} E_q\left(r_i A_i\right)\Big| s\right\} + E_p\left\{\sum_{i\in s} V_q\left(r_i A_i\right)\Big| s\right\} \\ &= \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}\left(\pi_{ij} - \pi_i\pi_j\right) p_i A_i p_j A_j + \sum_{i\in\mathcal{U}} p_i(1-p_i)\pi_i A_i^2. \end{aligned}$$

Ainsi, un estimateur sans biais de $V_{pq}\left(\sum_{i\in s} r_i A_i\right)$ est donné par :

$$\hat{V}_{pq}\left(\sum_{i\in s} r_i A_i\right) = \sum_{i\in s_r}\sum_{j\in s_r}\frac{\left(\pi_{ij} - \pi_i\pi_j\right)}{\pi_{ij}} A_i A_j + \sum_{i\in s_r}(1-p_i)A_i^2. \tag{4.6.5}$$

L'estimateur sans biais du carré du biais donné par (4.6.4) peut être négatif. De ce fait, on considérera comme Gwet et Rivest (1992) celui donné par :

$$\hat{B}_{pq}^2(\tilde{Y}_I^R) = max\left\{0, \tilde{B}_{pq}^2(\tilde{Y}_I^R)\right\}, \tag{4.6.6}$$

où $\tilde{B}_{pq}^2(\tilde{Y}_I^R)$ est donné par (4.6.4).

4.6.2. Estimation de la variance

Posons

$$\tilde{\eta}_i \;=\; \eta_i - \pi_i r_i A_i. \tag{4.6.7}$$

La variance $V_{pq}(\tilde{Y}_I^R)$ de $\tilde{Y}_I^R$ est alors donnée par :

$$\begin{aligned}
V_{pq}\left(\tilde{Y}_I^R\right) &= V_p\left\{E_q\left(\sum_{i\in s} d_i\tilde{\eta}_i\Big|s\right)\right\} + E_p\left\{V_q\left(\sum_{i\in s} d_i\tilde{\eta}_i\Big|s\right)\right\} \\
&= \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}(\pi_{ij}-\pi_i\pi_j)\,\frac{E_q\left(\tilde{\eta}_i\big|s\right)}{\pi_i}\,\frac{E_q\left(\tilde{\eta}_j\big|s\right)}{\pi_j} \\
&\quad + \sum_{i\in\mathcal{U}} d_i p_i(1-p_i)\left[\nu_i + g_i(\mathbf{p})E_i(\mathbf{z}_i,\mathbf{B}_{\pi p}) - \pi_i A_i\right]^2 \\
&= \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}(\pi_{ij}-\pi_i\pi_j)\,\frac{m\left(\mathbf{z}_i,\mathbf{B}_{\pi p}\right) + p_i g_i(\mathbf{p})E_i(\mathbf{z}_i,\mathbf{B}_{\pi p}) - p_i\pi_i A_i}{\pi_i} \\
&\quad \times\frac{m\left(\mathbf{z}_j,\mathbf{B}_{\pi p}\right) + p_j g_j(p)E_j(\mathbf{z}_j,\mathbf{B}_{\pi p}) - p_j\pi_j A_j}{\pi_j} \\
&\quad + \sum_{i\in\mathcal{U}} d_i p_i(1-p_i)\left[\nu_i + g_i(\mathbf{p})E_i(\mathbf{z}_i,\mathbf{B}_{\pi p}) - \pi_i A_i\right]^2.
\end{aligned} \tag{4.6.8}$$

Par la suite, un estimateur de la variance $V_{pq}(\tilde{Y}_I^R)$ est donné par :

$$\begin{aligned}
\hat{V}_{pq}\left(\tilde{Y}_I^R\right) &= \sum_{i\in s}\sum_{j\in s}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}}\,\frac{m\left(\mathbf{z}_i,\mathbf{B}_{\pi p}\right) + r_i g_i(\mathbf{p})E_i(\mathbf{z}_i,\mathbf{B}_{\pi p}) - r_i\pi_i A_i}{\pi_i} \\
&\quad \times\frac{m\left(\mathbf{z}_j,\mathbf{B}_{\pi p}\right) + r_j g_j(\mathbf{p})E_j(\mathbf{z}_j,\mathbf{B}_{\pi p}) - r_j\pi_j A_j}{\pi_j} \\
&\quad + \sum_{i\in s} r_i(1-p_i)\left[d_i\nu_i + d_i g_i(\mathbf{p})E_i(\mathbf{z}_i,\mathbf{B}_{\pi p}) - A_i\right]^2.
\end{aligned}$$

Finalement, un estimateur de la variance $V_{pq}(\hat{Y}_I^R)$ est donné par :

$$\begin{aligned}
\hat{V}_{pq}\left(\hat{Y}_I^R\right) &= \sum_{i\in s}\sum_{j\in s}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}}\,\frac{m\left(\mathbf{z}_i,\hat{\boldsymbol{\beta}}_r\right) + r_i\hat{g}_i(\hat{\mathbf{p}})E_i(\mathbf{z}_i,\hat{\boldsymbol{\beta}}_r) - r_i\pi_i A_i}{\pi_i} \\
&\quad \times\frac{m\left(\mathbf{z}_j,\hat{\boldsymbol{\beta}}_r\right) + r_j\hat{g}_j(\hat{\mathbf{p}})E_j(\mathbf{z}_j,\hat{\boldsymbol{\beta}}_r) - r_j\pi_j A_j}{\pi_j} \\
&\quad + \sum_{i\in s} r_i(1-\hat{p}_i)\left[d_i\hat{\nu}_i + d_i\hat{g}_i(\hat{\mathbf{p}})E_i(\mathbf{z}_i,\hat{\boldsymbol{\beta}}_r) - \hat{A}_i\right]^2.
\end{aligned} \tag{4.6.9}$$

Par la suite, un estimateur de l'erreur quadratique moyenne EQM est donné par :

$$\widehat{EQM}_{pq}\left(\hat{Y}_I^R\right) \;=\; \hat{B}_{pq}^2 + \hat{V}_{pq}\left(\hat{Y}_I^R\right), \tag{4.6.10}$$

où $\hat{B}_{pq}^2$ est donné par (4.6.6).

4.7. Étude par simulation

4.7.1. Description de la population et du processus utilisé

Dans cette section, les résultats en termes de biais et d'efficacité d'une étude par simulation dont le but était de comparer les estimateurs imputés robustes et non robustes à la présence de valeurs influentes sont présentés.

Une population de taille $N = 10000$ comprenant deux variables : une variable d'intérêt y et une variable auxilaire z a été générée. D'abord, la variable z a été générée d'une loi normale de moyenne 10 et de variance 25. Étant donné z, la variable y a été générée selon le modèle de mélange

$$y_i = (1 - A_i)y_{0i} + A_i y_{1i}, \quad i = 1, \ldots, N, \tag{4.7.1}$$

où A_i est une variable dichotomique telle que $A_i = 1$ avec probabilité $\lambda \in (0, 1)$ et $A_i = 0$ avec probabilité $1 - \lambda$ et

$$\begin{aligned} y_{0i} &= 3 + z_i + \epsilon_{0i}, \quad i = 1, \ldots, N, \\ y_{1i} &= 5z_i + \epsilon_{1i}, \quad i = 1, \ldots, N. \end{aligned}$$

Les erreurs ϵ_{0i} et ϵ_{1i} ont été générées d'une loi normale de moyenne 0 et de variance 1. La valeur de $\lambda = 0.05$ a été utilisée. De la population, $T = 10000$ échantillons ont été sélectionnés selon un plan aléatoire simple sans remise de taille $n = 100$. Dans chaque échantillon tiré, une probabilité de réponse p_i, a été assignée à l'unité i selon :

$$p_i = \frac{\exp^{(2.5-0.2z_i)}}{1 + \exp^{(2.5-0.2z_i)}}, \quad i = 1, \ldots, N. \tag{4.7.2}$$

Les paramètres en (4.7.2) ont été choisis de manière à obtenir un taux de réponse global approximativement égal à 65%. Finalement, une variable indicatrice r_i pour l'unité i a été générée aléatoirement d'une distribution de Bernoulli de paramètre p_i, $i = 1, \cdots, n$.

Afin de construire les valeurs imputées, 3 scénarios ont été considérés :

(1) Scénario 1 : le modèle de non-réponse et le modèle d'imputation sont bien spécifiés. On a d'abord obtenu $\hat{p}_i = p(\mathbf{z}_i, \hat{\boldsymbol{\alpha}})$ avec $\mathbf{z}_i = (1, z_i)'$. Les valeurs manquantes à la variable y ont été ensuite imputées selon (4.2.5) avec $\mathbf{z}_i = (1, z_i)'$ et $c_i = 1$.

(2) Scénario 2 : le modèle de non-réponse est mal spécifié alors que le modèle d'imputation est bien spécifié. On a d'abord obtenu $\hat{p}_i = p(\mathbf{z}_i, \hat{\boldsymbol{\alpha}})$ avec $\mathbf{z}_i = 1$. Les valeurs manquantes à la variable y ont été ensuite imputées selon (4.2.5) avec $\mathbf{z}_i = (1, z_i)'$ et $c_i = 1$.

(3) Scénario 3 : le modèle de non-réponse est bien spécifié alors que le modèle d'imputation est mal spécifié. On a obtenu $\hat{p}_i = p(\mathbf{z}_i, \hat{\boldsymbol{\alpha}})$ avec $\mathbf{z}_i = (1, z_i)'$. Les valeurs manquantes à la variable y ont été ensuite imputées selon (4.2.5) avec $\mathbf{z}_i = 1$ et $c_i = 1$.

Pour chaque scénario, quatre estimateurs ont été calculés :

(i) l'estimateur doublement robuste $\hat{Y}_I$ (non-robuste à la présence de valeurs influentes) donné par (4.2.2) avec les valeurs imputées (4.2.5).

(ii) L'estimateur robuste $\hat{Y}_I^{eqm}$ donné par (4.5.3), où la constante c est celle qui minimise son erreur quadratique moyenne estimée donnée par (4.6.10). En fait, (4.6.10) a été calculé pour chaque valeur de $c = |r_i \hat{B}_i, i \in s|$ et la valeur de c qui minimise (4.6.10) a été retenue.

(iii) L'estimateur robuste $\hat{Y}_I^{MCEQM}$ donné par (4.5.3), où la constante c est celle qui minimise l'erreur quadratique moyenne Monte Carlo. En fait, 501 valeurs de c allant de 0 à 25000 ont été considérées : $c = 50k, k = 0, 1, \ldots, 500$ et la valeur de c qui minimise l'EQM Monte Carlo a été retenue.

Il est à noter que cet estimateur est difficilement calculable en pratique car la valeur de c optimale n'est pas connue en général. Cependant, l'estimateur $\hat{Y}_I^{MCEQM}$ est un standard auquel on compare $\hat{Y}_I^{eqm}$ et $\hat{Y}_I^{CB}$ défini ci-dessous.

(iv) L'estimateur robuste $\hat{Y}_I^{CB}$ donné par (4.5.3), où la constante c est la valeur du minimax de la valeur absolue du biais conditionnel de l'estimateur robuste, voir Beaumont *et al.* (2013) ; Cet estimateur robuste s'écrit $\hat{Y}_I^{CB} = \hat{Y}_I - \frac{1}{2}(B_{min} + B_{max})$, voir Beaumont *et al.* (2013).

Pour les estimateurs robustes $\hat{Y}_I^R$, le biais conditionnel (4.4.6) a été estimé selon l'approche basée sur le modèle d'imputation par

$$\begin{aligned}\hat{B}_{mi}^I(y_i, I_i = 1, r_i = 1) &= \frac{N}{N-1}\left(\frac{N}{n} - 1\right)\{y_i - med\,(\tilde{y}_i, i \in s_r)\} \\ &+ \frac{N}{n}\left(\hat{g}_i(r) - 1\right)\left(y_i - \mathbf{z}_i^\top \hat{\boldsymbol{\beta}}_r^{rob}\right),\end{aligned}$$

où

$$\hat{g}_i(r) = 1 + \left\{\frac{N}{n}\sum_{l\in s}(1 - r_l)\mathbf{z}_l^\top\right\}\left\{\frac{N}{n}\sum_{l\in s} r_l \frac{1-\hat{p}_l}{\hat{p}_l}\mathbf{z}_l\mathbf{z}_l^\top\right\}^{-1}\frac{1-\hat{p}_i}{\hat{p}_i}\mathbf{z}_i,$$

$\tilde{y}_i = r_i y_i + (1 - r_i)y_i^*$, et $med(\cdot)$ désigne la médiane.

De même, pour l'approche basée sur le modèle de non-réponse, le biais conditionnel (4.4.3) a été estimé par

$$\hat{B}_{qi}^I(I_i = 1, r_i = 1) = \frac{N}{N-1}\left(\frac{N}{n} - 1\right)\{y_i - med\,(\tilde{y}_i, i \in s_r)\}$$

$$+ \quad \frac{N}{n}(1-\hat{p}_i)\left\{\hat{\nu}_i + \hat{g}_i(\hat{\mathbf{p}})\left(y_i - \mathbf{z}_i^\top \hat{\boldsymbol{\beta}}_r^{rob}\right)\right\},$$

où

$$\hat{g}_i(\hat{\mathbf{p}}) = 1 + \left\{\frac{N}{n}\sum_{l\in s}(1-\hat{p}_l)\mathbf{z}_l^\top\right\}\left\{\frac{N}{n}\sum_{l\in s}(1-\hat{p}_l)\mathbf{z}_l\mathbf{z}_l^\top\right\}^{-1}\frac{1-\hat{p}_i}{\hat{p}_i}\mathbf{z}_i,$$

$$\begin{aligned}\hat{\nu}_i &= \sum_{l\in s}\frac{N}{n}(1-\hat{p}_l)\mathbf{z}_l^\top\left\{\frac{N}{n}\sum_{l\in s}(1-\hat{p}_l)\mathbf{z}_l\mathbf{z}_l^\top\right\}^{-1}\left\{-\frac{N}{n}\sum_{l\in s} r_l\frac{1-\hat{p}_l}{\hat{p}_l}\left(y_l - \mathbf{z}_l^\top\hat{\boldsymbol{\beta}}_r^{rob}\right)\mathbf{z}_l\mathbf{z}_l^\top\right\}\\ &\quad\times\left\{\frac{N}{n}\sum_{l\in s}\hat{p}_l(1-\hat{p}_l)\mathbf{z}_l\mathbf{z}_l^\top\right\}^{-1}\mathbf{z}_i,\end{aligned}$$

et l'estimateur $\hat{\boldsymbol{\beta}}_r^{rob}$ est solution de

$$N^{-1}\sum_{i\in s} r_i d_i\left(\hat{p}_i^{-1}-1\right)\psi_c\left(\frac{y_i - \mathbf{z}_i^\top\boldsymbol{\beta}}{\sigma c_i^{1/2}}\right)\frac{\mathbf{z}_i}{\sigma c_i^{1/2}} = 0.$$

La constante de Huber c a été fixée à 1.345.

Deux mesures Monte Carlo ont été calculées :

(i) le biais relatif Monte Carlo (en %) donné par

$$BR(\hat{Y}) = T^{-1}\sum_{t=1}^{T}\frac{\left(\hat{Y}_t - Y\right)}{Y}\times 100,$$

où $\hat{Y}$ est une notation générique pouvant désigner soit $\hat{Y}_I$, soit $\hat{Y}_I^R$ ou alors l'estimateur de l'erreur quadratique moyenne $eqm(\hat{Y})$ de $EQM(\hat{Y})$ donnée par (4.6.10) ; et $\hat{Y}_t$ désigne l'estimateur $\hat{Y}$ pour l'échantillon t, $t = 1, \cdots, T$;

(ii) l'efficacité relative Monte Carlo (en %), définie par

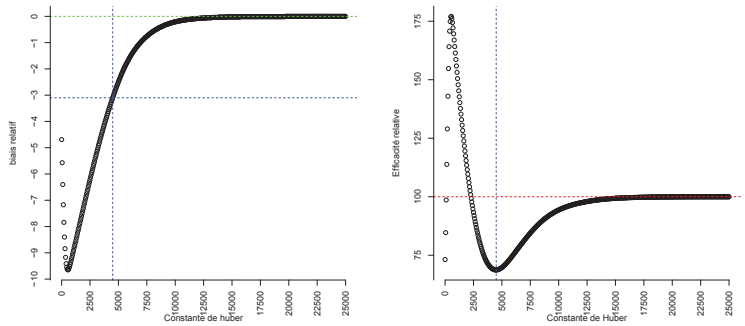
$$ER(\hat{Y}_I^R) = \frac{MSE(\hat{Y}_I^R)}{MSE(\hat{Y}_I)}\times 100,$$

où

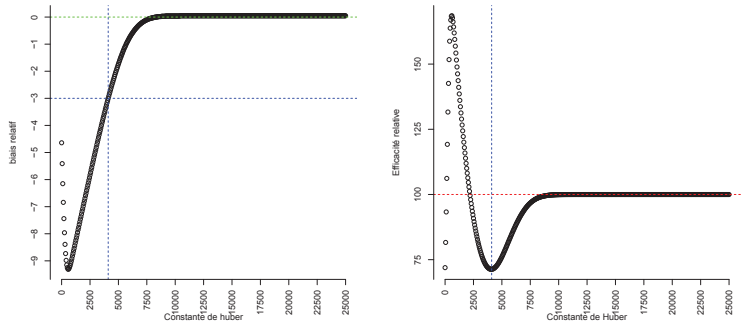
$$MSE(\hat{Y}) = T^{-1}\sum_{t=1}^{T}\left(\hat{Y}_t - Y\right)^2.$$

4.7.2. Résultats

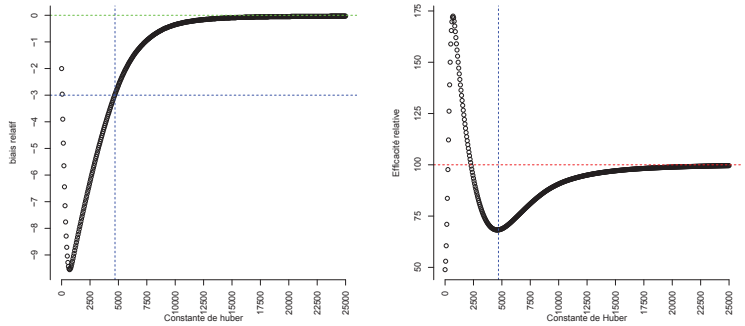
L'évolution de l'EQM Monte Carlo en fonction de la constante de rupture est présentée dans les figures (4.1) et (4.2) pour les modèles de non-réponse et d'imputation respectivement. Pour le modèle d'imputation, la valeur $c = 5000$ minimise l'EQM quel que soit le scénario considéré. Pour le modèle de non-réponse, la valeur de c qui minimise l'EQM Monte Carlo est de $c = 4500$ pour le scénario 1, $c = 4100$ pour le scénario 2 et $c = 0$ pour le scénario 3. Il apparaît clairement que lorsque la constante $c = \infty$, on obtient l'estimateur doublement robuste $\hat{Y}_I$.



(a) Scénario 1 (b) Scénario 1



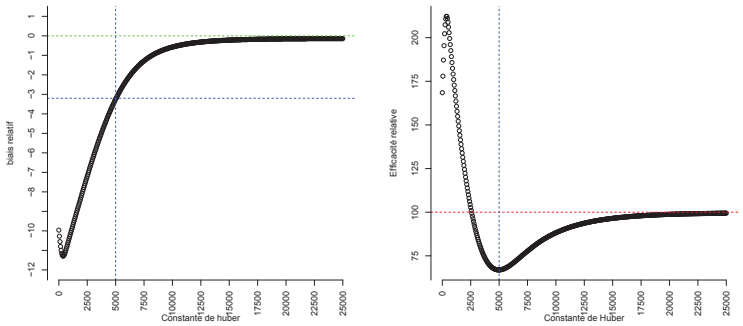
(c) Scénario 2 (d) Scénario 2



(e) Scénario 3 (f) Scénario 3

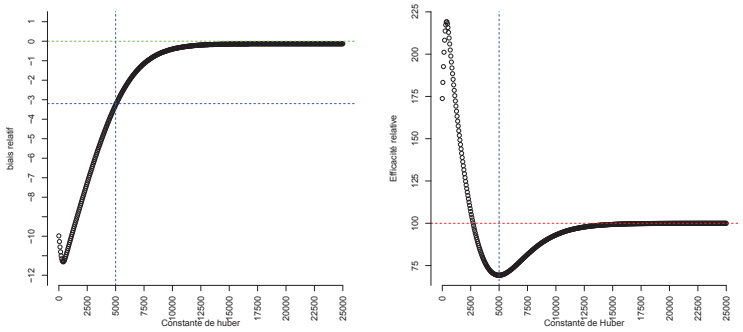
FIGURE 4.1. Biais absolu relatif et efficacité relative en fonction de la constante de rupture sous l'approche du modèle de non réponse.

Le tableau (4.1) présente les résultats des estimateurs ponctuels en termes de biais et d'efficacité relative pour le modèle de non-réponse. Il est clair que l'estimateur $\hat{Y}_I$ est doublement robuste puisque son biais est négligeable pour les 3 scénarios. L'estimateur obtenu avec la constante c qui minimise l'EQM estimée $\hat{Y}_I^{eqm}$ est plus efficace que celui obtenu avec la constante c égale au minimax du



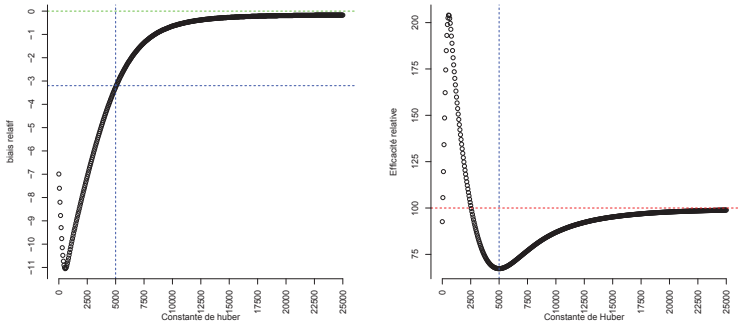
(a) Scénario 1

(b) Scénario 1



(c) Scénario 2

(d) Scénario 2



(e) Scénario 3

(f) Scénario 3

FIGURE 4.2. Biais absolu relatif et efficacité relative en fonction de la constante de rupture sous l'approche du modèle d'imputation.

biais conditionnel de l'estimateur robuste (Beaumont *et al.* 2013) ; cependant, son bias est plus élevé pour les scénario 1 et 2. L'estimateur $\hat{Y}_I^{eqm}$ a un biais apparent, il se situe à -4.5% pour les scénario 1 et 2, et à -2.2% pour le scénario 3. En revanche, son efficacité relative varie de 54% à 79%, ce qui illustre sa résistance à la présence d'unités influentes confirmant ainsi sa dénomination d'estimateur triplement robuste. Quand à l'estimateur obtenu avec le c qui minimise le l'EQM,

TABLE 4.1. Biais relatifs Monte Carlo (%) et efficacité relative Monte Carlo (%) des estimateurs avec $n = 100$. Approche basée sur le modèle de non-réponse.

Estimateurs	BR (%)	ER (%)
Scénario 1		
$\hat{Y}_I$	0.00	100.00
$\hat{Y}_I^{CB}$	-2.18	93.10
$\hat{Y}_I^{eqm}$	-4.58	77.51
$\hat{Y}_I^{MCEQM}$	-3.09	68.74
Scénario 2		
$\hat{Y}_I$	0.05	100.00
$\hat{Y}_I^{CB}$	-1.85	95.92
$\hat{Y}_I^{eqm}$	-4.47	78.54
$\hat{Y}_I^{MCEQM}$	-3.08	71.33
Scénario 3		
$\hat{Y}_I$	-0.03	100.00
$\hat{Y}_I^{CB}$	-2.23	90.79
$\hat{Y}_I^{eqm}$	-2.15	53.72
$\hat{Y}_I^{MCEQM}$	-2.00	48.91

TABLE 4.2. Biais relatifs Monte Carlo (%) et efficacité relative Monte Carlo (%) des estimateurs avec $n = 100$. Approche basée sur le modèle d'imputation.

Estimateurs	BR (%)	ER (%)
Scénario 1		
$\hat{Y}_I$	0.14	100.00
$\hat{Y}_I^{CB}$	-2.54	91.26
$\hat{Y}_I^{eqm}$	-4.46	115.97
$\hat{Y}_I^{MCEQM}$	-3.30	66.96
Scénario 2		
$\hat{Y}_I$	-0.13	100.00
$\hat{Y}_I^{CB}$	-2.49	93.98
$\hat{Y}_I^{eqm}$	-4.43	120.14
$\hat{Y}_I^{MCEQM}$	-3.26	69.26
Scénario 3		
$\hat{Y}_I$	-0.15	100.00
$\hat{Y}_I^{CB}$	-2.51	90.41
$\hat{Y}_I^{eqm}$	-4.42	95.49
$\hat{Y}_I^{MCEQM}$	-3.26	67.32

il performe mieux que tous les autres estimateurs en termes d'efficacité relative quel que soit le scénario.

TABLE 4.3. Biais relatifs Monte Carlo (%) des estimateurs de l'erreur quadratique moyenne avec $n = 100$. Approche basée sur le modèle d'imputation.

	BR de l'estimateur de l'EQM (%)		
Estimateurs ponctuels	Scénario 1	Scénario 2	Scénario 3
Modèle de non-réponse		Approche deux phases	
$\hat{Y}_I$	-2.5	-14.2	3.1
$\hat{Y}_I^{CB}$	-40.5	-42.0	-55.9
$\hat{Y}_I^{eqm}$	-89.7	-88.1	-96.7
$\hat{Y}_I^{MCEQM}$	-14.7	-17.8	-92.8
Modèle d'imputation		Approche renversée	
$\hat{Y}_I$	-5.6	-3.8	-0.2
$\hat{Y}_I^{CB}$	-36.0	-34.5	-36.5
$\hat{Y}_I^{eqm}$	-55.9	-55.4	-62.7
$\hat{Y}_I^{MCEQM}$	-2.3	-2.8	-1.7

Le tableau (4.2) présente les résultats des estimateurs ponctuels en termes de biais et d'efficacité relative pour le modèle d'imputation. Il est également clair ici que l'estimateur $\hat{Y}_I$ est doublement robuste puisque son biais est négligeable pour les 3 scénarios. L'estimateur obtenu avec la constante c qui minimise l'EQM estimée $\hat{Y}_I^{eqm}$ performe moins bien que tous les autres estimateurs en termes de biais et d'efficacité. S'agissant de l'estimateur obtenu avec la constante de robustesse c qui minimise le l'EQM, il performe mieux que tous les autres estimateurs en termes d'efficacité relative quel que soit le scénario considéré. Son biais relatif se situe à -3.5% quel que soit le scénario et son efficacité relative se situe à 67.0%, 69.3% et 67.3% pour les scénarios 1,2 et 3 respectivement.

Le tableau (4.3) présente les résultats en terme de biais de l'estimation de l'EQM. Le biais de l'estimateur de l'EQM pour l'estimateur triplement robuste obtenu avec la constante qui minimise l'EQM estimée est le plus élevé quelle que soit l'approche pour l'inférence et le scénario considérés ; Il varie de -88.1% à -92.8% pour le modèle de non-réponse, et de -55.4% à -62.7% pour le modèle d'imputation. Le biais obtenu avec la constante égale au minimax du biais conditionnel de l'estimateur robuste (Beaumont *et al.* 2013) est également élevé quelle que soit l'approche pour l'inférence et le scénario considéré ; Il varie de -40.5% à -55.9% pour le modèle de non-réponse, et de -34.5% à -36.5% pour le modèle d'imputation. À l'inverse, le biais est très faible pour l'estimateur de l'EQM de l'estimateur triplement robuste obtenu avec la constante qui minimise l'EQM Monte Carlo dans le cas du modèle d'imputation. Ce biais est modéré pour les

scénarios 1 et 2 dans le cas du modèle de non-réponse. Mais il est très élevé dans le cas du scénario 3.

4.8. Conclusion et discussion

En ce qui concerne l'imputation, nous avons considéré le problème de l'élaboration d'un estimateur imputé triplement robuste, c'est à dire qu'il soit robuste à la présence des valeurs aberrantes et qu'il soit construit à partir d'un estimateur imputé doublement robuste. Un estimateur imputé est dit doublement robuste s'il est sans biais lorsque le modèle d'imputation ou le modèle de non-réponse est bien spécifié. Partant de deux différentes approches pour l'inférence, à savoir l'inférence basée sur le modèle d'imputation et l'inférence basée sur le modèle de non-réponse, nous avons construit un estimateur doublement robuste avec une démarche similaire à Haziza & Rao (2006). Par la suite, deux estimateurs triplement robustes ont été élaborés en utilisant une approche similaire à celle de Beaumont *et al.* (2013) basée sur le biais conditionnel. Le premier était construit en utilisant l'approche du modèle de non-réponse pour l'inférence et le deuxième en utlisant l'approche du modèle d'imputation pour l'inférence.

Un des défis rencontrés lors de l'utilisation de l'estimateur triplement robuste proposé était la détermination de la constante de robustesse. Les procédures existantes proposent entre autres d'utiliser le minimax de la valeur absolue du biais conditionnel (Beaumont *et al.* 2013) de l'estimateur robuste ou d'utiliser celle qui minimise l'erreur quadratique moyenne (Kokic & Bell 1994) ou un estimateur de l'erreur quadratique moyenne. En utilisant des simulations Monte Carlo, nous avons examiné les performances de l'estimateur triplement robuste en termes de biais et d'efficacité pour ces différents choix de la constante de robustesse. Nous avons trouvé que l'estimateur triplement robuste qui utilise le minimax de la valeur absolue du biais conditionnel de l'estimateur robuste a un biais modéré mais une faible performance en terme d'efficacité. Quant à l'estimateur basé sur la constante de robustesse qui minimise l'erreur quadratique moyenne estimée, il est biaisé et performe bien en terme d'efficacité pour ce qui est du modèle de non-réponse mais performe mal pour le modèle d'imputation. En revanche, s'agissant de l'estimateur triplement robuste qui utilise la valeur de la constante de robustesse qui minimise l'erreur quadratique moyenne, il a un biais modéré et performe bien en termes d'efficacité.

Dans les agences statistiques officielles, il est courant de faire un remaniement des enquêtes après chaque recensement. C'est l'occasion de fixer certains paramètres qui seront utilisés jusqu'au prochain recensement. La valeur de la constante de robustesse qui minimise l'erreur quadratique moyenne peut ainsi

être déterminée avec les données du recensement en utilisant les procédures développées dans cette thèse. Cette constante de robustesse pourra être fixée jusqu'au prochain recensement et/ou remaniement de l'enquête.

Finalement, nous avons proposé un estimateur de l'erreur quadratique moyenne fondé sur le fait que le biais conditionnel pondéré par le poids de sondage est approximativement constant. Les résultats des simulations pour le modèle d'imputation sont encourageants lorsque la constante de robustesse utilisée est celle qui minimise l'erreur quadratique moyenne. Mais nous notons que cet estimateur de l'erreur quadratique moyenne est biaisé en général et le biais peut être très important pour certains choix de la constante de robustesse. C'est le cas notamment lorsque l'on utilise la constante de robustesse qui minimise l'erreur quadratique moyenne estimée ou le minimax du biais conditionnel de l'estimateur robuste. Une alternative sera d'utiliser le bootstrap pour estimer l'erreur quadratique moyenne. Ce problème représente une question de recherche intéressante pour le futur.

Appendice

A.1. Démonstration du théorème 4.3.1

Soit $\mathcal{U}$ une population finie de taille N. Soit $\mathbf{z}_i, i \in \mathcal{U}$ un vecteur de variables auxiliaires de dimension $k \times 1$ disponible pour chaque élément de la population $\mathcal{U}$. On veut estimer le total $Y = \sum_{i \in \mathcal{U}} y_i$ pour une variable d'intérêt y définie selon le modèle d'imputation (4.2.4). Soit s un échantillon de taille n obtenu selon un plan $p(s)$. Soit $r_i, i \in \mathcal{U}$ l'indicateur de réponse à la variable aléatoire y_i défini par le modèle paramétrique (4.2.3).

Nous énonçons ci-dessous les hypothèses similaires à celles de Isaki & Fuller (1982). Ces hypothèses sont adaptées au cadre spécifique de l'imputation.

Hypothèse NM.1 Les réponses sont indépendantes

$$P(r_i = 1, r_j = 1) = p_i p_j, i \neq j.$$

Hypothèse NM.2 Les probabilités de réponses sont minorées

$$\exists \quad \lambda > 0, \quad \forall i \in \mathcal{U}, \quad \lambda < p_i.$$

Hypothèse NM.3 Le modèle d'imputation est avec constante

$$h(\mathbf{z}_i, \boldsymbol{\beta}) = \frac{1}{c_i} \frac{\partial m(\mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left(1, h_{1i}^{\top}(\mathbf{z}_i, \boldsymbol{\beta})\right)^{\top}.$$

Hypothèse NM.4 On suppose que le plan d'échantillonnage et le mécanisme de réponse sont indépendants.

Hypothèse NM.5 On suppose que le mécanisme de réponse est ignorable.

Hypothèse NM.6 On suppose que $m(\mathbf{z}_i, \boldsymbol{\beta})$ et $p(\mathbf{z}_i, \boldsymbol{\alpha})$ sont des fonctions de classe $\mathcal{C}^2$ sur un ensemble compact $\mathcal{K}$ (deux fois dérivables et avec des dérivées partielles continues) contenant respectivement $\boldsymbol{\beta}_0$ et $\boldsymbol{\alpha}_0$ comme points intérieurs.

Hypothèse NM.7 On suppose que les équations

$$\hat{\mathbf{U}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) = 0, \quad \text{et} \quad \mathbf{U}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) = E_{pq}\left\{\hat{\mathbf{U}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)\right\} = 0,$$

admettent chacune une unique solution notée respectivement $\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)$ et $\mathbf{B}_{\pi p}$, où E_{pq} désigne l'espérance par rapport à la distribution conjointe du modèle de non réponse et du plan d'échantillonnage.

Les facteurs d'ajustement $g_i(p)$ et ν_i donnés dans le théorème 4.3.1 sont définis ci-dessous :

$$g_i(\mathbf{p}) = 1 + \left\{\frac{\partial M(\mathbf{z}, \mathbf{B}_{\pi p})}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{-\mathbf{H}_{\pi p}(\mathbf{B}_{\pi p})\right\}^{-1} N^{-1} \frac{1-p_i}{p_i} h(\mathbf{z}_i, \mathbf{B}_{\pi p}), \qquad \text{(A.1)}$$

$$\nu_i = \left\{\frac{\partial M(\mathbf{z}, \mathbf{B}_{\pi p})}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{-\mathbf{H}_{\pi p}(\mathbf{B}_{\pi p})\right\}^{-1} \mathbf{T}_{\pi p}(\mathbf{B}_{\pi p}) \left\{-\mathbf{I}(\boldsymbol{\alpha})\right\}^{-1} N^{-1} \mathbf{l}_i(\boldsymbol{\alpha}), \quad \text{(A.2)}$$

où

$$\begin{aligned}
M(\mathbf{z}, \boldsymbol{\beta}) &= \sum_{i \in \mathcal{U}} (1-p_i) m(\mathbf{z}_i, \boldsymbol{\beta}), \\
\mathbf{T}_{\pi p}(\boldsymbol{\beta}) &= N^{-1} \sum_{i \in \mathcal{U}} p_i \left\{y_i - m(\mathbf{z}_i, \boldsymbol{\beta})\right\} \left(\frac{\partial p_i^{-1}}{\partial \boldsymbol{\alpha}}\right)^{\top} h(\mathbf{z}_i, \boldsymbol{\beta}), \\
\hat{\mathbf{U}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= N^{-1} \sum_{i \in s} d_i r_i \frac{1-p_i}{p_i} \left\{y_i - m(\mathbf{z}_i, \boldsymbol{\beta})\right\} h(\mathbf{z}_i, \boldsymbol{\beta}), \\
\hat{\mathbf{H}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \frac{\partial \hat{\mathbf{U}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \\
&= N^{-1} \sum_{i \in s} d_i r_i \frac{1-p_i}{p_i} \left[\left\{y_i - m(\mathbf{z}_i, \boldsymbol{\beta})\right\} \frac{\partial h(\mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{\partial m(\mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} h^{\top}(\mathbf{z}_i, \boldsymbol{\beta})\right], \\
\mathbf{H}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= E_{pq}\left\{\hat{\mathbf{H}}_r(\boldsymbol{\beta})\right\} \\
&= N^{-1} \sum_{i \in \mathcal{U}} (1-p_i) \left[\left\{y_i - m(\mathbf{z}_i, \boldsymbol{\beta})\right\} \frac{\partial h(\mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{\partial m(\mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} h^{\top}(\mathbf{z}_i, \boldsymbol{\beta})\right]
\end{aligned}$$

.

On définit les hypothèses similaires à celles de Izaki et Fuller (1982) dans un contexte d'absence de non-réponse.

Hypothèse NM.8

$$\lim_{N \to \infty} N^{-1} \mathbf{W}_N^{\top} (\mathbf{I}_N - \mathbf{P}_N) \mathbf{W}_N = \mathbf{G}_{\pi p}, \quad 0 < |\mathbf{G}_{\pi p}| < \infty, \qquad \text{(A.3)}$$

où

$$\mathbf{W}_N = \left(\mathbf{w}_1^\top, \ldots, \mathbf{w}_N^\top\right)^\top, \quad \mathbf{P}_N = diag(p_1, \ldots, p_N),$$

$$\mathbf{w}_i = \left[y_i, m(\mathbf{z}_i, \boldsymbol{\beta}), p_i, vect\left(\frac{\partial h_i}{\partial \boldsymbol{\beta}}\right), \left\{\frac{\partial m(\mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right\}^\top, h^\top(\mathbf{z}_i, \boldsymbol{\beta}) \right]^\top,$$

$vect\left(\frac{\partial h_i}{\partial \boldsymbol{\beta}}\right)$ est le vecteur de dimension $q^2 \times 1$ dont les éléments sont ceux de la matrice $\frac{\partial h_i(\mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$, et $\mathbf{I}_N$ est la matrice unité d'ordre N.

Hypothèse NM.9

$$\lim_{\substack{n\to\infty \\ N\to\infty}} nE_{pq}\left\{(\hat{\mathbf{V}}_r - \mathbf{V}_{\pi p})(\hat{\mathbf{V}}_r - \mathbf{V}_{\pi p})^\top\right\} = \boldsymbol{\Phi}_{vv\pi p}, \quad 0 < |\boldsymbol{\Phi}_{vv\pi p}| < \infty, \qquad (A.4)$$

où

$$\hat{\mathbf{V}}_r = N^{-1} \sum_{i\in s} d_i \left(p_i^{-1} - 1\right) \mathbf{w}_i \mathbf{w}_i^\top$$

$$\mathbf{V}_{\pi p} = N^{-1} \sum_{i\in\mathcal{U}} (1 - p_i)\, \mathbf{w}_i \mathbf{w}_i^\top.$$

Hypothèse NM.10 Les conditions de régularité sont suffisantes pour que l'estimateur $\hat{\boldsymbol{\alpha}}$, unique solution de (4.2.7) vérifie

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = -\mathbf{I}(\boldsymbol{\alpha})^{-1} \hat{\mathbf{S}}(\boldsymbol{\alpha}_0) + o_p(n^{-1/2}), \qquad (A.5)$$

où $\mathbf{I}(\boldsymbol{\alpha}) = -E_{pq}\left\{\frac{\partial \hat{\mathbf{S}}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right\}$.

Dans la suite, on énonce et démontre trois lemmes qui seront utiles pour la démontration du théorème 4.3.1. Soit Ω l'ensemble univers constitué des points de l'espace probabilisé utilisé dans l'approche NM ; et soit $E = E_p E_q$ l'opérateur espérance associé.

Lemme A.1. *Sous les hypothèses **NM.1** à **NM.9**, on a*

$$\hat{\boldsymbol{\beta}}_r\left(\boldsymbol{\alpha}_0\right) - \mathbf{B}_{\pi p} = o_p(1), \qquad (A.6)$$

où O_p et o_p dénotent les ordres de grandeurs par rapport à la distribution conjointe du mécanisme de réponse et du plan d'échantillonnage.

Démonstration. L'hypothèse **NM.9** implique que

$$\hat{\mathbf{U}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) = \mathbf{U}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) + O_p(n^{-1/2}), \quad \forall \boldsymbol{\beta} \in \mathcal{K}. \tag{A.7}$$

Par ailleurs, d'après l'hypothèse (**NM.6**), $|\hat{\mathbf{U}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) - \mathbf{U}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)|$ est continue sur le compact $\mathcal{K}$. Dès lors,

$$\sup_{\boldsymbol{\beta} \in \mathcal{K}} |\hat{\mathbf{U}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) - \mathbf{U}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)| = O_p(n^{-1/2}). \tag{A.8}$$

Pour démontrer le lemme A.1, il suffit maintenant de montrer que toutes les sous-suites de $\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)$ admettent une sous-suite qui converge presque sûrement vers $\mathbf{B}_{\pi p}$, voir Billingsley P.(1986), p. 286.

Soit $\hat{\boldsymbol{\beta}}_{r_n}$ une sous-suite de $\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)$. Alors $\hat{\mathbf{U}}_{r_n}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)$ est une sous-suite de $\hat{\mathbf{U}}_r(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)$ qui converge en probabilité vers $\mathbf{U}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)$. De ce fait, $\hat{\mathbf{U}}_{r_n}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)$ admet une sous-suite, dénotée $\hat{\mathbf{U}}_{r_{n_k}}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)$ qui converge presque sûrement vers $\mathbf{U}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)$. De l'expression (A.8), on déduit que l'on peut en extraire une sous-suite de $\hat{\mathbf{U}}_{r_{n_k}}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)$ qui converge uniformément en $\boldsymbol{\beta}$ vers $\mathbf{U}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)$ presque sûrement. Pour alléger les notations, nous allons désigner cette nouvelle sous-suite par $\hat{\mathbf{U}}_{r_{n_k}}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)$.

Considérons maintenant la sous-suite $\hat{\boldsymbol{\beta}}_{r_{n_k}}$ de $\hat{\boldsymbol{\beta}}_{r_n}$. Alors cette suite admet une sous-suite qui converge presque sûrement vers $\mathbf{B}_{\pi p}$. En effet, soit $\omega \in \Omega$, étant donné que $\mathcal{K}$ est compact, alors on peut en extraire de la suite des vecteurs $\hat{\boldsymbol{\beta}}_{r_{n_k}}(\omega) \in \mathcal{K}$ une sous-suite $\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}(\omega)$ convergente. Notons $\hat{\boldsymbol{\beta}}_\infty(\omega)$ la limite de cette sous-suite. Il résulte de la convergence uniforme presque sure de $\hat{\mathbf{U}}_{r_{n_{k_l}}}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)$ vers $\mathbf{U}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)$ et de la continuité de ces deux fonctions que

$$\lim_{l \to \infty} \hat{\mathbf{U}}_{r_{n_{k_l}}}\left(\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}, \boldsymbol{\alpha}_0\right) = \mathbf{U}_{\pi p}\left(\hat{\boldsymbol{\beta}}_\infty, \boldsymbol{\alpha}_0\right) \quad ps.$$

En effet soit Ω_0 l'ensemble négligeable pour lequel $\hat{\mathbf{U}}_{r_{n_{k_l}}}\{\boldsymbol{\beta}_\infty(\omega), \boldsymbol{\alpha}_0\}$ ne converge pas vers $\mathbf{U}_{\pi p}\{\boldsymbol{\beta}_\infty(\omega), \boldsymbol{\alpha}_0\}$, $\omega \in \Omega_0^c$. Pour simplifier l'écriture, $\hat{\mathbf{U}}_{r_{n_{k_l}}}\left(\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}, \boldsymbol{\alpha}_0\right)$ et $\mathbf{U}_{\pi p}\left(\hat{\boldsymbol{\beta}}_\infty, \boldsymbol{\alpha}_0\right)$ seront notés $\hat{\mathbf{U}}_{r_{n_{k_l}}}$ et $\mathbf{U}_{\pi p}$ respectivement. On a :

$$\begin{aligned}
\left|\hat{\mathbf{U}}_{r_{n_{k_l}}} - \mathbf{U}_{\pi p}\right| &\leq \left|\hat{\mathbf{U}}_{r_{n_{k_l}}}\left(\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}, \boldsymbol{\alpha}_0\right) - \mathbf{U}_{\pi p}\left(\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}, \boldsymbol{\alpha}_0\right)\right| \\
&\quad + \left|\mathbf{U}_{\pi p}\left(\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}, \boldsymbol{\alpha}_0\right) - \mathbf{U}_{\pi p}\left(\hat{\boldsymbol{\beta}}_\infty, \boldsymbol{\alpha}_0\right)\right| \\
&\leq \sup_{\boldsymbol{\beta} \in \mathcal{K}} \left|\hat{\mathbf{U}}_{r_{n_{k_l}}}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)(\omega) - \mathbf{U}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)(\omega)\right| \\
&\quad + \left|\mathbf{U}_{\pi p}\left\{\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}(\omega), \boldsymbol{\alpha}_0\right\} - \mathbf{U}_{\pi p}\left\{\hat{\boldsymbol{\beta}}_\infty(\omega), \boldsymbol{\alpha}_0\right\}\right|
\end{aligned}$$

$$= O\left(n_{k_l}^{-1/2}\right) + \left|\mathbf{U}_{\pi p}\left\{\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}(\omega), \boldsymbol{\alpha}_0\right\} - \mathbf{U}_{\pi p}\left\{\hat{\boldsymbol{\beta}}_{\infty}(\omega), \boldsymbol{\alpha}_0\right\}\right|. \quad \text{(A.9)}$$

Or

$$\forall l, \quad \hat{\mathbf{U}}_{r_{n_{k_l}}}\left\{\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}(\omega), \boldsymbol{\alpha}_0\right\} = 0 \quad \forall \omega \in \Omega_0^c,$$

donc

$$\mathbf{U}_{\pi p}\left\{\hat{\boldsymbol{\beta}}_{\infty}(\omega), \boldsymbol{\alpha}_0\right\} = \lim_{l\to\infty} \hat{\mathbf{U}}_{r_{n_{k_l}}}\left\{\hat{\boldsymbol{\beta}}_{r_{n_{k_l}}}(\omega), \boldsymbol{\alpha}_0\right\} = 0 \quad \forall \omega \in \Omega_0^c.$$

Ainsi, en raison de l'hypothèse **NM.7** (unicité de la solution), on en déduit que $\mathbf{B}_{\pi p} = \hat{\boldsymbol{\beta}}_{\infty}$ presque sûrement. Ce qui achève la démonstration du lemme A.1 □

Lemme A.2. *Sous les hypothèses **NM.1** à **NM.9** et en utilisant le lemme A.1, on a :*

$$\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0) - \mathbf{B}_{\pi p} = O_p(n^{-1/2}), \quad \text{(A.10)}$$

Démonstration. D'après le théorème des acroissements finis, il existe un point $\hat{\boldsymbol{\beta}}^*$ strictement compris entre $\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)$ et $\mathbf{B}_{\pi p}$ tel que :

$$\hat{\mathbf{U}}_r\left(\hat{\boldsymbol{\beta}}_r, \boldsymbol{\alpha}_0\right) - \hat{\mathbf{U}}_r\left(\hat{\boldsymbol{\beta}}_{\pi p}, \boldsymbol{\alpha}_0\right) = \hat{\mathbf{H}}_r\left(\hat{\boldsymbol{\beta}}^*, \boldsymbol{\alpha}_0\right)\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0) - \mathbf{B}_{\pi p}\right\}. \quad \text{(A.11)}$$

Comme dans le lemme A.1, les hypothèses **NM.6** et **NM.9** impliquent

$$\sup_{\boldsymbol{\beta}\in\mathcal{K}} |\hat{\mathbf{H}}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) - \mathbf{H}_{\pi p}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)| = O_p(n^{-1/2}). \quad \text{(A.12)}$$

Donc en particulier on a : $\hat{\mathbf{H}}_r(\boldsymbol{\beta}^*, \boldsymbol{\alpha}_0) = \mathbf{H}_{\pi p}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}_0) + O_p(n^{-1/2})$ et en tenant compte du lemme A.1, l'expression (A.11) devient

$$-\hat{\mathbf{U}}_r\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right) = \mathbf{H}_{\pi p}\left(\hat{\boldsymbol{\beta}}^*, \boldsymbol{\alpha}_0\right)\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0) - \mathbf{B}_{\pi p}\right\} + o_p(n^{-1/2}). \quad \text{(A.13)}$$

Par ailleurs, d'après l'hypothèse **NM.8**,

$$\mathbf{H}_{\pi p}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_0\right) = O_p(1), \quad \forall(\boldsymbol{\beta}).$$

Étant donné que $\mathcal{K}$ est un compact, et $\mathbf{H}_{\pi p}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_0\right)$ est continue alors $\mathbf{H}_{\pi p}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_0\right)$ est uniformément borné en $\boldsymbol{\beta}$:

$$\sup_{\boldsymbol{\beta}\in\mathcal{K}} \left|\mathbf{H}_{\pi p}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_0\right)\right| = O_p(1).$$

Par la suite, on déduit que

$$\mathbf{H}_{\pi p}\left(\hat{\boldsymbol{\beta}}^*, \boldsymbol{\alpha}_0\right) = O_p(1). \quad \text{(A.14)}$$

De plus l'hypothèse **NM.9** implique

$$\hat{\mathbf{U}}_r\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right) \;=\; \mathbf{U}_{\pi p}\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right) + O_p(n^{-1/2}) = O_p(n^{-1/2}). \tag{A.15}$$

En combinant les résultats (A.13), (A.14) et (A.15) on obtient :

$$\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0) - \mathbf{B}_{\pi p} = O_p(n^{-1/2}).$$

Ce qui achève la démonstration du lemme A.2 □

Lemme A.3. *Sous les hypothèses **NM.1** à **NM.9** et en utilisant le lemme A.2, on a :*

$$\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0) - \mathbf{B}_{\pi p} = -\left\{\mathbf{H}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{U}}_r(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0) + o_p(n^{-1/2}), \tag{A.16}$$

Démonstration. La continuité de la fonction $\mathbf{H}_{\pi p}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_0\right)$ et la convergence en probabilité de $\hat{\boldsymbol{\beta}}^*$ vers $\mathbf{B}_{\pi p}$ entraîne

$$\mathbf{H}_{\pi p}(\hat{\boldsymbol{\beta}}^*, \boldsymbol{\alpha}_0) = \mathbf{H}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0) + o_p(1), \tag{A.17}$$

En remplacant $\mathbf{H}_{\pi p}(\hat{\boldsymbol{\beta}}^*, \boldsymbol{\alpha}_0)$ donné par (A.17) dans l'expression (A.13) et en tenant compte du lemme A.2, on obtient :

$$-\hat{\mathbf{U}}_r\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right) \;=\; \mathbf{H}_{\pi p}\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right)\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0) - \mathbf{B}_{\pi p}\right\} + o_p(n^{-1/2}). \tag{A.18}$$

Le résultat du lemme A.3 est enfin obtenu en utilisant le fait que $\mathbf{H}_{\pi p}\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right)$ est inversible et également que $\mathbf{H}_{\pi p}\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right) = O_p(1)$.

□

Démonstration. **Preuve du théorème 4.3.1**

Nous commencons par faire un développement de Taylor à l'ordre un de $m(\mathbf{z}_i, \boldsymbol{\beta})$ au voisinage de $\mathbf{B}_{\pi p}$. Ensuite on applique le lemme A.3 pour obtenir :

$$\begin{aligned} m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\} \;=\; & m(\mathbf{z}_i, \mathbf{B}_{\pi p}) \\ & + \left\{\frac{\partial m(\mathbf{z}_i, \mathbf{B}_{\pi p})}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{-\mathbf{H}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{U}}_r(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0) \\ & + o_p(n^{-1/2}). \end{aligned} \tag{A.19}$$

Notons que l'on peut écrire l'estimateur donné par (4.2.2) comme fonction de $\hat{\boldsymbol{\beta}}_r$ et $\hat{\boldsymbol{\alpha}}$, solutions respectives de (4.2.6) et de (4.2.7) : $\hat{Y}_I(\hat{\boldsymbol{\beta}}_r, \hat{\boldsymbol{\alpha}})$. Il est également à noter que les hypothèses **NM.6** et **NM.7** assurent que $\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})$ est une fonction de classe $\mathcal{C}^1$ de $\hat{\boldsymbol{\alpha}}$ (théorème des fonctions implicites). Et par conséquent, $m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})\right\}$ est une fonction de classe $\mathcal{C}^1$ de $\hat{\boldsymbol{\alpha}}$ car c'est une composée de

fonctions de classe $\mathcal{C}^1$ de $\hat{\boldsymbol{\alpha}}$. De ce fait, en effectuant un développement de Taylor de $m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})\right\}$ autour du point $\boldsymbol{\alpha}_0$, on obtient :

$$\begin{aligned} m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})\right\} &= m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\} + \left[\frac{\partial m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\}}{\partial \boldsymbol{\beta}}\right]^\top \frac{\partial \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)}{\partial \hat{\boldsymbol{\alpha}}} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\ &\quad + (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)\, o_p\left(\left|\left|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right|\right|\right). \end{aligned} \tag{A.20}$$

En tenant compte de l'hypothès **NM.10**, on obtient :

$$\begin{aligned} m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})\right\} &= m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\} \\ &\quad + \left[\frac{\partial m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\}}{\partial \boldsymbol{\beta}}\right]^\top \frac{\partial \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)}{\partial \hat{\boldsymbol{\alpha}}_0} \left\{-\mathbf{I}(\boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{S}}(\boldsymbol{\alpha}_0) \\ &\quad + o_p\left(n^{-1/2}\right). \end{aligned} \tag{A.21}$$

En outre, la résolution de l'équation $\frac{\partial \hat{U}_r}{\partial \boldsymbol{\alpha}}\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0), \boldsymbol{\alpha}_0\right\} = 0$ permet d'obtenir

$$\frac{\partial \hat{\boldsymbol{\beta}}_r}{\partial \hat{\boldsymbol{\alpha}}}(\boldsymbol{\alpha}_0) = \left[-\hat{\mathbf{H}}_r\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\}\right]^{-1} \hat{\mathbf{T}}_r\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\}. \tag{A.22}$$

Par ailleurs, d'après les hypothèses **NM.8** et **NM.9**, on a :

$$\begin{aligned} \frac{\partial \hat{\boldsymbol{\beta}}_r}{\partial \hat{\boldsymbol{\alpha}}}(\boldsymbol{\alpha}_0) &= \left[-\hat{\mathbf{H}}_r\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\}\right]^{-1} \hat{\mathbf{T}}_r\left\{\hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\} \\ &= \left\{-\mathbf{H}_{\pi p}\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right)\right\}^{-1} \mathbf{T}_{\pi p}\left(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0\right) \\ &\quad + O_p(n^{-1/2}), \end{aligned} \tag{A.23}$$

$$\hat{\mathbf{S}}(\boldsymbol{\alpha}_0) = O_p(n^{-1/2}) \tag{A.24}$$

$$\frac{\partial m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\}}{\partial \boldsymbol{\beta}} = \frac{\partial m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)}{\partial \boldsymbol{\beta}} + O_p(n^{-1/2}). \tag{A.25}$$

En combinant (A.23), (A.24) et (A.25), alors (A.20) devient :

$$\begin{aligned} m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})\right\} &= m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\} \\ &+ \left\{\frac{\partial m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)}{\partial \boldsymbol{\beta}}\right\}^\top \left\{\frac{\partial \mathbf{B}_{\pi p}}{\partial \hat{\boldsymbol{\alpha}}}(\boldsymbol{\alpha}_0)\right\} \left\{-\mathbf{I}(\boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{S}}(\boldsymbol{\alpha}_0) \\ &+ o_p\left(n^{-1/2}\right). \end{aligned} \tag{A.26}$$

En substituant $m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\boldsymbol{\alpha}_0)\right\}$ par son expression donnée par (A.19), alors (A.26) devient

$$\begin{aligned} m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})\right\} &= m(\mathbf{z}_i, \mathbf{B}_{\pi p}) \\ &+ \left\{\frac{\partial m(\mathbf{z}_i, \mathbf{B}_{\pi p})}{\partial \boldsymbol{\beta}}\right\}^\top \left\{-\mathbf{H}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{U}}_r(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0) \end{aligned}$$

$$+ \quad \left\{\frac{\partial m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{\frac{\partial \mathbf{B}_{\pi p}}{\partial \hat{\boldsymbol{\alpha}}}(\boldsymbol{\alpha}_0)\right\} \left\{-\mathbf{I}(\boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{S}}(\boldsymbol{\alpha}_0)$$
$$+ \quad o_p\left(n^{-1/2}\right). \tag{A.27}$$

De même en substituant $m\left\{\mathbf{z}_i, \hat{\boldsymbol{\beta}}_r(\hat{\boldsymbol{\alpha}})\right\}$ par son expression donnée par (A.27), dans l'expression (4.2.2) on obtient :

$$\begin{aligned}
\hat{Y}_I &= \sum_{i\in U} d_i r_i I_i y_i + \sum_{i\in U} d_i (1-r_i) I_i m(\mathbf{z}_i, \mathbf{B}_{\pi p}) \\
&+ \sum_{i\in U} d_i (1-r_i) I_i \left\{\frac{\partial m(\mathbf{z}_i, \mathbf{B}_{\pi p})}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{-\mathbf{H}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{U}}_r(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0) \\
&+ \sum_{i\in U} d_i (1-r_i) I_i \left\{\frac{\partial m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{\frac{\partial \mathbf{B}_{\pi p}}{\partial \hat{\boldsymbol{\alpha}}}(\boldsymbol{\alpha}_0)\right\} \left\{-\mathbf{I}(\boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{S}}(\boldsymbol{\alpha}_0) \\
&+ \sum_{i\in U} d_i (1-r_i) I_i + o_p\left(n^{-1/2}\right).
\end{aligned} \tag{A.28}$$

Et d'après l'hypothèse **NM.9**,

$$\begin{aligned}
\sum_{i\in U} d_i (1-r_i) I_i \left\{\frac{\partial m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)}{\partial \boldsymbol{\beta}}\right\}^{\top} &= \sum_{i\in U} (1-p_i) \left\{\frac{\partial m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)}{\partial \boldsymbol{\beta}}\right\}^{\top} + O_p(Nn^{-1/2}) \\
\sum_{i\in U} d_i (1-r_i) I_i &= O_p(N) \\
\hat{\mathbf{U}}_r(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0) &= O_p(n^{-1/2}) \\
\hat{\mathbf{S}}(\boldsymbol{\alpha}_0) &= O_p(n^{-1/2}).
\end{aligned} \tag{A.29}$$

où l'ordre de grandeur est par rapport à la distribution conjointe du mécanisme de réponse et du plan d'échantillonnage. En tenant compe de (A.29), on obtient finalement :

$$\begin{aligned}
\hat{Y}_I &= \sum_{i\in U} d_i r_i I_i y_i + \sum_{i\in U} d_i (1-r_i) I_i m(\mathbf{z}_i, \mathbf{B}_{\pi p}) \\
&+ \sum_{i\in U} (1-p_i) \left\{\frac{\partial m(\mathbf{z}_i, \mathbf{B}_{\pi p})}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{-\mathbf{H}_{\pi p}(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{U}}_r(\mathbf{B}_{\pi p}, \boldsymbol{\alpha}_0) \\
&+ \sum_{i\in U} (1-p_i) \left\{\frac{\partial m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)}{\partial \boldsymbol{\beta}}\right\}^{\top} \left\{\frac{\partial \mathbf{B}_{\pi p}}{\partial \hat{\boldsymbol{\alpha}}}(\boldsymbol{\alpha}_0)\right\} \left\{-\mathbf{I}(\boldsymbol{\alpha}_0)\right\}^{-1} \hat{\mathbf{S}}(\boldsymbol{\alpha}_0) \\
&+ o_p\left(Nn^{-1/2}\right) \\
&= \hat{Y}_\pi + \sum_{i\in s} d_i \left\{r_i g_i(p) - 1\right\} \left\{y_i - m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)\right\} \\
&+ \sum_{i\in s} d_i \nu_i (r_i - p_i) + o_p\left(Nn^{-1/2}\right)
\end{aligned}$$

$$= \sum_{i \in s} d_i \eta_i + o_p\left(N n^{-1/2}\right). \tag{A.30}$$

où

$$\eta_i = m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right) + r_i g_i(p)\left\{y_i - m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)\right\} + \nu_i(r_i - p_i). \tag{A.31}$$

L'expression (A.31) est utile pour le calcul de la variance de $\hat{Y}_I$. Le terme $\nu_i(r_i - p_i)$ compte pour l'augmentation de la variance en raison du fait que les probabilités de réponse p_i sont estimées. Et le terme $\{r_i g_i(p) - 1\}\left\{y_i - m\left(\mathbf{z}_i, \mathbf{B}_{\pi p}\right)\right\}$ compte pour l'augmentation de la variance due au processus d'imputation. □

Bibliographie

[1] Beaumont, J.-F. (2005). Calibrated imputation in surveys under a quasi-model-assisted approach. *Journal of the Royal Statistical Society B*, **67**, 445–458.

[2] Beaumont, J.-F., Haziza, D. & Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, **100**, 555-569.

[3] Brick, J. M., Kalton, G. & Kim, J. K. (2004). Variance estimation with Hot-Deck imputation using a model. *Survey Methodology*, **30**, 57–66.

[4] Gwet, J.-P & Rivest, L.-P. (1992). Outlier resistant alternatives to the ratio estimator. *Journal of the American Statistical Association*, **87**, 736–739.

[5] Haziza, D. (2009). Imputation and inference in the presence of missing data. In C. R. Rao & D. Pfefferman (Editors), *Handbook of Statistics, Sample Surveys, Design Methods and Applications*, **29A**, pp. 215–246.

[6] Haziza, D. & Picard, F. (2012). On doubly robust point and variance estimation in the presence of imputed data. *The Canadian Journal of Statistics*, **40**, 259–281.

[7] Haziza, D. & Rao, J. N. K. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodology*, **32**, 53–64.

[8] Isaki, C.T. & Fuller W.A. (1982). Survey design under the regression superpopulation model. *Journal of American Statistical Association*, **77**, 89–96.

[9] Kim, J.K. & Haziza, D. (2014). Doubly robust inference with missing data. *Statistica Sinica*, **24**, 375–394.

[10] Kim, J.K. & Park, H. (2006). Imputation using response probability. *The Canadian Journal of Statistics*, **34**, 171–182.

[11] Kim, J.K. & Rao, J.N.K. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika*, **96**, 917–932.

[12] Kokic, P.N. & Bell, P.A. (1994). Optimal winzorizing cutoffs for a stratified finite population estimator. *Journal of Official Statistics*, **10**, 419–435.

[13] Kott, P.S. (1994). A note on handling nonresponse in sample surveys. *Journal of American Statistical Association*, **89**, 693–696.

[14] Moreno-Rebollo, J. L., Munoz-Reyes, A. et Munoz-Pichardo, J. (1999). Influence diagnostic in survey sampling : conditional bias. *Biometrika*, **86**, 923–928.

[15] Rao, J. N. K. (1996). On variance estimation with imputed survey data. *Journal of American Statistical Association*, **91**, 499–506.

[16] Särndal, C. E. (1992). Method for estimating the precision of survey estimates when imputation has been used. *Survey Methodology*, **18**, 241–252.

[17] Shao, J. et Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, **94**, 254–65.

CONCLUSION

Dans cette thèse, le problème du traitement des valeurs aberrantes dans deux aspects importants des enquêtes a été abordé : il s'agit de l'estimation des petits domaines et de l'imputation en présence de non-réponse partielle. La thèse a été subdivisée en quatre chapitres. Le chapitre un porte sur une revue de la littérature qui a pour objectif de définir les termes et les notions qui ont été utilisés dans les chapitres suivants. Les chapitres deux et trois sont consacrés à l'estimation robuste des petits domaines et font l'objet de deux articles. Le chapitre quatre porte sur l'imputation robuste et fait l'objet d'un troisième article.

S'agissant de l'estimation des petits domaines, deux classes d'estimateurs robustes dans le cadre des modèles au niveau des unités ont été proposées. Etant donné que l'estimateur robuste proposé par Sinha & Rao (2009) peut être biaisé, deux nouveaux estimateurs corrigés pour le biais ont été proposés. Le premier est fondé sur les travaux de Chambers (1986) et le deuxième sur le concept de biais conditionnel comme mesure d'influence. Chacque terme de correction pour le biais proposé est différent de celui de Chambers *et al.* (2014), et ils montrent un impact possible des valeurs aberrantes présentes dans les autres domaines sur le domaine d'intérêt. En utilisant les arguments asymptotiques dans le cadre d'un modèle linéaire mixte, nous avons montré que l'estimateur de Sinha & Rao (2009) peut souffrir d'un biais substantiel, alors que le biais des nouvelles méthodes peut être contrôlé au moyen d'un choix approprié de la constante de robustesse de la fonction d'influence. Chose intéressante, les nouveaux estimateurs proposés représentent un compromis entre l'estimateur non-robuste (EBLUP) et l'estimateur de Sinha & Rao (2009).

Une étude Monte Carlo a été effectuée, et les comparaisons ont été faites entre les nouveaux estimateurs robustes, l'estimateur de Sinha & Rao (2009) et l'estimateur corrigé pour le biais de Chambers *et al.* (2014). Sous un modèle de mélange, nous avons montré que Sinha & Rao (2009) peut avoir un biais substantiel, alors que les estimateurs proposés offrent souvent une meilleur performance en termes de biais et d'erreur quadratique moyenne. De ce fait, l'objectif principal

qui était de proposer un estimateur robuste avec de meilleures propriétés peut donc clairement être considéré comme atteint. Un objectif équivalent et important dans les applications était de fournir un estimateur performant de l'erreur quadratique moyenne.

Ce problème a été examiné sous l'angle des méthodes bootstrap. Il est à noter qu'une procédure bootstrap paramétrique a été étudiée par Sinha & Rao (2009) et que des estimateurs analytiques ont été proposés par Chambers *et al.* (2014). Étant donné que la variabilité des estimateurs robustes des paramètres de variance est typiquement plus petite que celle des données originelles (Field *et al.* 2010), le bootstrap de Sinha & Rao (2009) peut être biaisé négativement. Afin de lever cette contrainte, nous avons utilisé les estimateurs du maximum de vraisemblance (non-robustes) pour générer les échantillons bootstrap et par la suite nous avons appliqué les mêmes techniques sur ces échantillons pour obtenir une version bootstrap des estimateurs robustes. C'est à partir de ce point que les estimateurs de l'erreur quadratique moyenne des estimateurs robustes sont construits.

Par ailleurs, les propriétés asymptotiques des procédures existantes d'estimation de l'erreur quadratique moyenne des estimateurs robustes des petits domaines ne sont pas établies théoriquement. Nous avons proposé une nouvelle méthode bootstrap semi-paramétrique et nous avons montré formellement la validité théorique de cette nouvelle procédure. Nos résultats théoriques sont dérivés en utilisant une approche similaire à celle de Bickel & Freedman (1981) et Freedman (1981) ainsi que les résultats asymptotiques établis par Huggins (1993). Les preuves pour l'estimateur bootstrap de l'erreur quadratique moyenne ont été faites pour l'estimateur de Sinha & Rao (2009), et des arguments similaires peuvent facilement être utilisés pour faire une extension à d'autres estimateurs robustes. C'est à notre connaissance, la première fois que la validité théorique d'une procédure bootstrap est établie pour un estimateur robuste des petits domaines. Mieux encore, la nature semi-paramétrique de la méthode proposée la rend particulièrement attrayante car elle peut ainsi s'affranchir de l'hypothèse de normalité qui n'est pas souvent rencontrée en pratique.

Nous avons examiné les propriétés en termes de biais et d'erreur quadratique moyenne de la nouvelle méthode à travers les simulations Monte Carlo et ses performances ont été comparées avec cinq autres méthodes : le bootstrap de Sinha & Rao (2009), le bootstrap de Jiongo *et al.* (2013), l'estimateur analytique pseudo-linéaire ainsi que l'estimateur linéaire de l'erreur quadratique moyenne de Chambers *et al.* (2014), et l'estimateur de Prsad & Rao (1990). Les résultats montrent que pour tous les différents estimateurs robustes et tous les différents

types de contamination considérés, l'estimateur bootstrap proposé de l'erreur quadratique moyenne a la meilleure performance en termes de biais et d'efficacité. Une application aux données réelles tirées de l'agriculture américaine et des images satellitaires illustre comment le bootstrap proposé peut être utilisé en pratique.

Finalement, nous notons que les estimateurs robustes corrigés pour le biais des petits domaines proposés dépendent d'un choix approprié de la constante de robustesse. S'agissant de celui basé sur le biais conditionnel, il a été montré au chapitre 3 (deuxième article) que le choix proposé par Beaumont *et al.* (2013), est facile à mettre en œuvre en pratique et donne de très bons résultats et ce quel que soit le mode de contamination des valeurs aberrantes. En revanche, s'agissant de l'estimateur fondé sur la décomposition de Chambers (1986), il reste à trouver une manière simple de déterminer la constante de robustesse. Cette méthode devrait être simple à implémenter et donner de bons résultats que les valeurs aberrantes soient présentes dans les termes d'erreurs, les effets aléatoires ou les effets fixes. Par ailleurs, l'estimateur bootstrap proposé se comporte bien dans le cadre d'un modèle linéaire mixte au niveau des unités. Il serait possible de développer une version de cet estimateur bootstrap pour un modèle linéaire mixte semi-paramétrique qui stipule une forme non-paramétrique pour les variables dépendantes. Ces problèmes représentent des possibilités de recherche future.

S'agissant de l'imputation, une approche unifiée de trois concepts de la robustesse a été proposée. Il s'agit précisément de l'élaboration d'un estimateur imputé ayant les caractéristiques suivantes :

- robuste à la mauvaise spécification du modèle de non-réponse lorsque le modèle d'imputation est bien spécifié ;
- robuste à la mauvaise spécification du modèle d'imputation lorsque le modèle de non-réponse est bien spécifié ;
- robuste aux valeurs aberrantes présentes dans l'échantillon.

La double robustesse par rapport au modèle de non-réponse ou au modèle d'imputation a été établie en utilisant une démarche similaire à Haziza & Rao (2006) et la robuste par rapport aux valeurs aberrantes est basée sur l'approche de Beaumont *et al.* (2013). Étant donné ces trois propriétés de robustesse, l'estimateur obtenu a été qualifié d'estimateur imputé triplement robuste.

Les résultats des simulations Monté Carlo ont montré que l'estimateur imputé triplement robuste peut être très efficace si la constante de robustesse est choisie de manière appropriée. C'est par exemple le cas lorsque la constante de robustesse est celle qui minimise l'erreur quadratique moyenne. En pratique, c'est estimateur peut uniquement être calculé si l'on dispose d'une base de sondage sur laquelle des

simulations sont au préalable effectuées. C'est le cas par exemple dans les agences officielles de statistique où il est courant d'effectuer un remaniement après chaque recensement. À c'est occasion, certains paramètres sont fixés et utilisés jusqu'au prochain recensement et/ou remaniement de l'enquête.

Deux autres méthodes de détermination de la constante de robustesse ont été proposées. La première utilise le minimax de la valeur absolue du biais conditionnel (Beaumont *et al.* 2013) de l'estimateur robuste, et la deuxième utilise celle qui minimise un estimateur de l'erreur quadratique moyenne. Les résultats obtenus avec cette dernière approche sont surprenant car l'estimateur triplement robuste performe bien dans le cas du modèle de non-réponse et a une mauvaise performance dans le cas du modèle d'imputation. Ce résultat semble remettre en question l'estimateur de l'erreur quadratique moyenne proposé. En particulier l'approximation de la variance de l'estimateur triplement robuste devrait être validée. En général, l'estimation de l'erreur quadratique moyenne n'est pas un problème facile. Une alternative à la méthode proposée est l'utilisation du bootstrap. Dans ce cadre, on pourrait adapter l'approche de Booth, Butler et Hall (1994) à l'imputation en présence de non réponse partielle.

Bibliographie

[1] American Mathematical Society, *$\mathcal{A}\mathcal{M}\mathcal{S}$-LaTeX Version 1.1 User's Guide*, Amer. Math. Soc., Providence, R. I., 1991.

[2] Beaumont, J.-F., Haziza, D. & Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, **100**, 555–569.

[3] Bickel, P.-J., Freedman, D.-A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196–1217.

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

[5] Chambers, R. L. (1986). Outliers robust finite population estimation. *Journal of the American Statistical Association*, **81**, 1063–1069.

[6] Chambers, R. L., Chandra, H., Salvati, N. & Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society. Series B*, **76**, 47–69.

[7] Field, C. A., Pang, Z. and Welsh, A. H. (2010). Bootstrapping robust estimates for clustered data. *Journal of American Statistical Association*, **105**, 1606–1616.

[8] Freedman, D.-A. (1981). Bootstrapping regression models. *The Annals of Statistics*, **9**, 1218-1228.

[9] Haziza, D. et Rao, J. N. K. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodology*, **32**, 53–64.

[10] Huggins, R.M. (1993). On the robust analysis of variance components models for pedigree data. *Australian Journal of Statistics*, **35**, 43–57.

[11] Jiongo, V., D., Haziza, D. & Duchesne, P. (2013). Controlling the bias of robust small-area estimators. *Biometrika*, **100**, 843–858.

[12] Prasad, N. G. N. and 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.

[13] Sinha, S. K. & Rao, J. N. K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, **37**, 381–399.