Robust estimation for GLM and GLMM in finite population

Cyril Favre Martinoz[1] (*), David Haziza[2] (**), Nikos Tzavidis[3] (***)

(*)Crest-Ensai, Irmar

(**) University of Montreal

(***) University of Southampton

Résumé

En pratique, il est courant d'observer des unités influentes dans les échantillon collectées, plus particulièrement lorsque l'on collecte de l'information sur des variables économiques dont la distribution est très asymétrique. Le fait d'ajouter ou de retirer cette unité dite influente a un impact significatif sur les estimateurs classiquement utilisés pour inférer sur des paramètres de population finie. La présence d'unités influentes est d'autant plus dramatique que la taille de l'échantillon est petite, c'est pourquoi les méthodes d'estimation robuste sur petits domaines se sont développées de façon importante au cours de ses dernières années, voir par exemple Gosh et al. (2008), Sinha and Rao (2009), Dongmo Jiongo et al. (2013), Chambers et al. (2013) and Fabrizi et al. (2014). La majorité de ces travaux reposent sur l'utilisation de modèle mixte au niveau des unités et s'intéressent à des variables d'intérêt continues. Dans ce cadre, quelques estimateurs robustes de l'estimateur linéaire sans biais optimal empirique ont été proposés dans la littérature en utilisant des méthodes de type M-quantile ou une approche basée sur le biais conditionnel. En pratique, il est courant de s'intéresser à des variables d'intérêt binaires ou discrètes. On a alors recours à des modèles logistiques mixtes ou des modèles de Poisson mixtes. Nous proposons dans un premier temps un estimateur robuste dans le cas d'une approche sous le modèle avec utilisation d'un modèle GLM, puis nous proposons une approche unifiée pour l'estimation robuste dans les petits domaines dans le cadre des modèles GLMM.

Mots-clés : estimation robuste, approche modèle, petits domaines, biais conditionnel.

[1]cyril.favremartinoz@ensai.fr

[2]david.haziza@umontreal.ca

[3]n.tzavidis@soton.ac.uk

Abstract

Influential units occur frequently in surveys, especially in the context of business surveys that collect economic variables whose distribution are highly skewed. A unit is said to be influential when its inclusion or exclusion from the sample has an important impact on the magnitude of survey statistics. Robust small area estimation has received a lot of attention in recent years; see Gosh et al. (2008), Sinha and Rao (2009), Dongmo Jiongo et al. (2013), Chambers et al. (2013) and Fabrizi et al. (2014), among others. So far, researchers have mainly focused on unit level models and continuous characteristics of interest. Several robust versions of the empirical best linear unbiased predictor based on linear mixed models (LMM) have been proposed in the literature, including an M-quantile regression approach and an approach based on the concept of conditional bias of a unit. In practice, one must often face binary and count data. In this case, methods based on LMMs are not suited. We first propose a robust estimator in a general model-based framework with the use of generalized linear models and then we propose a unified framework for robust small area estimation in the context of generalized LMMs. We construct a general robust estimator based on the concept of conditional bias.

Keywords: robust estimation, model-based approach, small area, conditional bias.

Introduction

Influential units are common feature of many sample surveys, especially in the context of business surveys, those variables of interest have highly skewed distribution. A unit is said to be influential when its inclusion or exclusion from the sample has an important impact on the magnitude of survey statistics. In presence of influential units, the Best Linear Unbiased Predictor is still unbiased but its variance can be very large. For continuous variables with the use of a linear model, some robust estimators have already been proposed, see Chambers (1986), Beaumont et al. (2013). In a small area context, some robust estimators have already been proposed for unit-level model using a linear mixed model, see Gosh et al. (2008), Sinha and Rao (2009), Dongmo Jiongo et al. (2013), Chambers et al. (2013) and Fabrizi et al. (2014), among others. So far, researchers have mainly focused on unit level models and continuous characteristics of interest. Several robust versions of the empirical best linear unbiased predictor based on linear mixed models (LMM) have been proposed in the literature, including an M-quantile regression approach and an approach based on the concept of conditional bias of a unit. In practice, one must often face binary and count data, which requires the use of generalized linear mixed models. For example, for a binary outcome in a frequentist approach, one can follow Jiang and Lahiri (2001), or Jiang (2003) who propose an empirical best predictor for generalized linear mixed models. These estimators are still very sensitive in presence of influential units that's why robust estimation is required. Some robust estimators have been proposed by Tzavidis et al. (2013) for count data and Chambers et al.(2014) for binary data using M-quantile regression approach. In this paper we propose in the first part an extension of the work of Beaumont et al. (2013) to a generalized linear model and we compare empirically the efficiency of the proposed robust estimator to the Empirical Best Predictor, and in the second part of this paper, we provide a robust estimator in a generalized linear mixed model in the special case of small area estimation.

1 Robust estimation for GLM in finite population

1.1 Model-based approach using GLM

In model-based inference for finite population sampling (e.g., Valliant, Dorfman and Royall, 2000), the y-values of the N population units are assumed to be generated from some model. We denote by X, the known N-row matrix containing the vector of explanatory variables x_i^T in its ith row. A non-informative sample s is selected from the finite population U and is treated as fixed when making inferences. The interest lies in the prediction of a function of the population Y-values through the sample Y -values. To fix ideas, we assume that we are interested in predicting the random population total $\theta = \sum_{i \in U} Y_i$. We consider a set of n i.i.d random variable Y_i whose expected value is denoted by $E(Y_i) = \mu_i$ and whose variance is denoted by $Var(Y_i) = \sigma_i^2$. We assume the distribution of Y_i is a member of the exponential family, so its probability density function can be written as

$$f(y_i) = exp\left(\frac{y_i\gamma_i - b(\gamma_i)}{a(\phi)} + c(y_i, a(\phi))\right)$$

where ϕ denotes the scale parameter.

The exponential family have the following well-known properties which are crucial for estimation and inference.

For the exponential family, we have the following results :

$$\mu_i = E(y_i) = \frac{\partial b(\gamma_i)}{\partial \theta_i},\ V(Y_i) = a(\phi)\frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2}.$$

The score function, denoted $S(\gamma)$, is defined as the derivative of the log-likelihood, $S(\theta) = \partial l(\gamma, y, \phi)/\partial\gamma$. The variance of the score function is called the Fisher information matrix, denoted by $Var(S(\gamma)) = I(\gamma)$.

We define the Log-likelihood as

$$l(\gamma, \mathbf{y}, \mathbf{X}) = log \prod_{i \in U} f(y_i|\gamma).$$

It can been shown that $\tilde{\beta}$, the solution of the maximum likelihood, is also the solution of this estimating equation in matrix form on the population :

$$\sum_{i \in U} \frac{Y_i - \mu_i}{a(\phi)} w_i g_\mu(\mu_i) x_i^T = 0 \tag{1}$$

where $w_i = [V(\mu_i)g_\mu(\mu_i)]^{-1}$and $g_\mu(\mu_i) = \partial g(\mu_i)/\partial\mu_i$

Assuming that we use a non-information design, we assume that $\hat{\beta}$ is the solution of the sample estimating equation :

$$\sum_{i \in S} \frac{Y_i - \mu_i}{a(\phi)} w_i g_\mu(\mu_i) x_i^T = 0 \tag{2}$$

We note $\boldsymbol{t}(y_i, \beta) = \frac{Y_i - \mu_i}{a(\phi)} w_i g_\mu(\mu_i) x_i^T$ and $\boldsymbol{H}(y_i, \boldsymbol{\beta}) = \frac{\partial t(y_i, u)}{\partial \beta}(y_i, \beta)$.

In this context, the empirical best predictor (EBP) of $\theta = \sum_{i \in U} Y_i$, can be expressed as

$$\hat{\theta}^{EBP} = \sum_{i \in S} Y_i + \sum_{i \in U \backslash S} h(\mathbf{x_i^T}\hat{\beta}).$$

where $h = g^{-1}$.

1.2 The conditional bias in GLM

Now we want to express the conditional bias of $\hat{\theta}^{EBP}$ under the model. In a model based approach, the conditional bias attached to unit i is $B_i(y_i; \beta) = E(\hat{\theta} - \theta|s; Y_i = y_i)$. The main problem here is that the EBP is no longer linear in the Y-values. So we use a first-order Taylor expansion so that :

$$h(x_i^T\hat{\beta}) = h(\mathbf{x_i^T}\beta) + \frac{dh(u)}{du}(\mathbf{x_i^T}\beta)\mathbf{x_i^T}(\hat{\beta} - \beta) + O_p\left(\frac{1}{n^{1/2}}\right). \tag{3}$$

Following Fuller (2011, page 65), we have :

$$\hat{\beta} - \beta = \left(\sum_{j \in S} \boldsymbol{H}(\boldsymbol{y_j}, \boldsymbol{\beta})\right)^{-1} \sum_{k \in S} \boldsymbol{t}(y_k, \boldsymbol{\beta}) + o_p\left(\frac{1}{n^{1/2}}\right). \tag{4}$$

Combining (1) and (2), we obtain :

$$h(x_i^T\hat{\beta}) = h(x_i^T\beta) - \frac{dh(u)}{du}(\mathbf{x_i^T}\beta)\mathbf{x_i^T}\left(\sum_{j\in S}\boldsymbol{H}(\boldsymbol{y_j},\boldsymbol{\beta})\right)^{-1}\sum_{k\in S}\boldsymbol{t}(y_k,\boldsymbol{\beta}) + O_p\left(\frac{1}{n^{1/2}}\right)$$

Now, using the following decomposition of the prediction error, we will be able to give an approximation of the conditional bias.

$$\begin{aligned}\hat{\theta}^{EBP} - \theta &= \sum_{j\in S} Y_j + \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\hat{\beta}) - \sum_{j\in U} Y_j \\ \hat{\theta}^{EBP} - \theta &= \sum_{j\in U\setminus S}\left(h(\mathbf{x_j^T}\hat{\beta}) - h(\mathbf{x_j^T}\beta)\right) + \sum_{j\in U\setminus S}\left(h(\mathbf{x_j^T}\beta) - Y_j\right)\end{aligned}$$

To determine the conditional bias, we need to distinguish two cases, whether the unit i belongs to the sample or not.

1.2.1 Conditional bias for a selected unit

We start by determining an approximation of the conditional bias for a selected unit.

$$\begin{aligned}B_i^{EBP}(I_i=1) &= E_m(\hat{\theta}^{EBP} - \theta | s, Y_i = y_i) \\ &\approx E_m(\sum_{j\in S} Y_j + \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\beta) - \sum_{j\in U} Y_j \,|\, s, Y_i = y_i) \\ &+ E_m(\sum_{j\in U\setminus S}\frac{dh(u)}{du}(\mathbf{x_j^T}\beta)\mathbf{x_j^T}\left(\sum_{j\in S}\boldsymbol{H}(\boldsymbol{y_j},\boldsymbol{\beta})\right)^{-1}\sum_{k\in S}\boldsymbol{t}(y_k,\boldsymbol{\beta}) \,|\, s, Y_i = y_i) \\ &\approx y_i + \sum_{j\in S,\, j\neq i} h(\mathbf{x_j^T}\beta) + \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\beta) - y_i - \sum_{j\in U,\, j\neq i} h(\mathbf{x_j^T}\beta) \\ &+ \sum_{j\in U\setminus S}\frac{dh(u)}{du}(\mathbf{x_j^T}\beta)\mathbf{x_j^T}\left(\sum_{j\in S}\boldsymbol{H}(\boldsymbol{y_j},\boldsymbol{\beta})\right)^{-1} E_m(\sum_{k\in S}\boldsymbol{t}(y_k,\boldsymbol{\beta}) \,|\, s, Y_i = y_i) \\ &\approx \sum_{j\in U\setminus S}\frac{dh(u)}{du}(\mathbf{x_j^T}\beta)\mathbf{x_j^T} E_m(\left(\sum_{j\in S}\boldsymbol{H}(\boldsymbol{y_j},\boldsymbol{\beta})\right)^{-1}\sum_{k\in S}\boldsymbol{t}(y_k,\boldsymbol{\beta}) \,|\, s, Y_i = y_i)\end{aligned}$$

In some cases, the hessian matrix still depends on the $Y-$random variables, so the conditional bias can not be expressed analytically. We suggest either to use the expectation of the hessian matrix for the Taylor approximation or to compute a Monte-Carlo approximation of the conditional bias.

We will now discuss in detail the conditional bias of the Empirical Best Predictor in three special cases, which are very useful in pratice.

1) linear case

We have $h = Id$ and $\boldsymbol{t}(y_i,\beta) = (y_i - x_i^T\beta)x_i^T$. We obtain :

$$\begin{aligned}B_i^{EBP}(I_i=1) &= \sum_{j\in U\setminus S}\frac{dh(u)}{du}(\mathbf{x_j^T}\beta)\mathbf{x_j^T}(\sum_{k\in S}\mathbf{x_k}\mathbf{x_k^T})^{-1}E_m(\sum_{j\in S}(y_j - \mathbf{x_j^T}\beta)\mathbf{x_j} \,|\, s, Y_i = y_i) \\ &= \sum_{j\in U\setminus S}\mathbf{x_j^T}(\sum_{k\in S}\mathbf{x_k}\mathbf{x_k^T})^{-1}\mathbf{x_i}\left(y_i - \mathbf{x_i^T}\beta\right)\end{aligned}$$

2) logistic case

We have $\boldsymbol{t}(y_i,\beta) = \big((y_i - h(x_i^T\beta)\big)\,\mathbf{x_i}$ we obtain :

$$\begin{aligned} B_i^{EBP}(I_i=1) &= \sum_{j\in U\setminus S}\frac{dh(u)}{du}(\mathbf{x_j^T}\beta)\mathbf{x_j^T}(\sum_{k\in S}\boldsymbol{H}(\boldsymbol{x_j},\boldsymbol{\beta}))^{-1}E_m(\sum_{j\in S}(y_j-\mathbf{h}(\mathbf{x_j^T}\beta)\mathbf{x_j}\,|\,s,Y_i=y_i) \\ &= \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\beta)(1-h(\mathbf{x_j^T}\beta))\mathbf{x_j^T}\left(\sum_{k\in S}h(\mathbf{x_k^T}\beta)(1-h(\mathbf{x_k^T}\beta))\mathbf{x_k}\mathbf{x_k^T}\right)^{-1}\mathbf{x_i}\left((y_i-h(\mathbf{x_i^T}\beta)\right) \end{aligned}$$

where $h(\mathbf{x_i^T}\beta) = \frac{exp(\mathbf{x_i^T}\beta)}{1+exp(\mathbf{x_i^T}\beta)}$.

For example, the conditional bias can be estimated by

$$\hat{B}_i^{EBP}(I_i=1) = \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\hat{\beta})(1-h(\mathbf{x_j^T}\hat{\beta}))\mathbf{x_j^T}\left(\sum_{k\in S}h(\mathbf{x_k^T}\hat{\beta})\mathbf{x_k}\mathbf{x_k^T}\right)^{-1}\mathbf{x_i}\left((y_i-h(\mathbf{x_i^T}\hat{\beta})\right)$$

where $\hat{\beta}$ is the regression coefficient of the sample-fitted logistic regression.

3)Poisson case

$$\begin{aligned} B_i^{EBP}(I_i=1) &= \sum_{j\in U\setminus S}\frac{dh(u)}{du}(\mathbf{x_j^T}\beta)\mathbf{x_j^T}(\sum_{k\in S}\boldsymbol{H}(\boldsymbol{x_j},\boldsymbol{\beta}))^{-1}E_m(\sum_{j\in S}(y_j-\mathbf{x_j^T}\beta)\mathbf{x_j}\,|\,s,Y_i=y_i) \\ &= \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\beta)\mathbf{x_j^T}\left(\sum_{k\in S}h(\mathbf{x_k^T}\beta)\mathbf{x_k}\mathbf{x_k^T}\right)^{-1}\mathbf{x_i}\left((y_i-F(\mathbf{x_i^T}\beta)\right) \end{aligned}$$

where $h(x_i^T\beta) = exp(x_i^T\beta)$.

1.2.2 Conditional bias for a non-selected unit

The conditional bias of a non-sample unit can be expressed as

$$B_i^{EBP}(I_i=0) = -(y_i - h(\mathbf{x_i^T}\beta)).$$

Proof:

$$\begin{aligned} E_m(\sum_{j\in S}Y_j + \sum_{j\in U\setminus S}h(\mathbf{x_j^T}\beta) - \sum_{j\in U}Y_j\,|\,s,Y_i=y_i) &= E_m\left(\sum_{j\in U\setminus S}h(\mathbf{x_j^T}\beta)-y_j|\,s,Y_i=y_i\right) \\ &= -\left(y_i - h(\mathbf{x_i^T}\beta)\right) \end{aligned}$$

The main problem is that the conditional bias of a non-sampled unit can't be estimated since it depends on the Y-values on the non-sample units, which are by definition not observed.

1.3 Construction of the robust estimator

The prediction error of the EBP can be approximately written as :

$$\hat{\theta}^{BLUP} - \theta \approx \sum_{i\in U\setminus S} B_i^{EBP}(I_i = 0) + \sum_{i\in S} B_i^{EBP}(I_i = 1).$$

Proof.

$$\begin{aligned}
& \sum_{i\in U\setminus S} B_i^{EBP}(I_i = 0) + \sum_{i\in S} B_i^{EBP}(I_i = 1) \\
= & -\sum_{i\in U\setminus S} \left(y_i - h(\mathbf{x_i^T}\beta)\right) + \sum_{i\in S}\sum_{j\in U\setminus S} h(\mathbf{x_j^T}\beta)\mathbf{x_j^T} \left(\sum_{k\in S} h(\mathbf{x_k^T}\beta)\mathbf{x_k}\mathbf{x_k^T}\right)^{-1} \mathbf{x_i}\left((y_i - h(\mathbf{x_i^T}\beta)\right) \\
= & -\sum_{i\in U\setminus S} \left(y_i - \left(h(\mathbf{x_i^T}\beta) - h(\mathbf{x_i^T}\hat{\beta})\right) - h(\mathbf{x_i^T}\hat{\beta})\right) \\
+ & \sum_{i\in S}\sum_{j\in U\setminus S} h(\mathbf{x_j^T}\beta)\mathbf{x_j^T} \left(\sum_{k\in S} h(\mathbf{x_k^T}\beta)\mathbf{x_k}\mathbf{x_k^T}\right)^{-1} \mathbf{x_i}\left((y_i - h(\mathbf{x_i^T}\beta)\right) \\
= & -\sum_{j\in U\setminus S} Y_j + \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\hat{\beta}) - \sum_{i\in U\setminus S} \left(h(\mathbf{x_i^T}\beta) - h(\mathbf{x_i^T}\hat{\beta})\right) \\
+ & \sum_{i\in S}\sum_{j\in U\setminus S} h(\mathbf{x_j^T}\beta)\mathbf{x_j^T} \left(\sum_{k\in S} h(\mathbf{x_k^T}\beta)\mathbf{x_k}\mathbf{x_k^T}\right)^{-1} \mathbf{x_i}\left((y_i - h(\mathbf{x_i^T}\beta)\right) \\
= & -\sum_{j\in U\setminus S} Y_j + \sum_{j\in U\setminus S} h(\mathbf{x_j^T}\hat{\beta}) + O_p\left(\frac{1}{n^{1/2}}\right) \\
= & \ \hat{\theta}^{BLUP} - \theta + O_p\left(\frac{1}{n^{1/2}}\right)
\end{aligned}$$

□

Following Beaumont et al.(2013) to construct a robust version of the EBP, we express it as :

$$\hat{\theta}^{REBP} = \hat{\theta}^{EBP} - \sum_{i\in S} B_i^{EBP}(I_i = 1) + \sum_{i\in S} \psi\left(B_i^{EBP}(I_i = 1)\right),$$

where $\psi(.)$ is the Huber function.

Now, we compute the conditional bias of the robust estimator define by

$$B_i^R(I_i = 1) = E_m(\hat{\theta}^{REBP} - \theta | s, Y_i = y_i).$$

We can prove that :

$$B_i^R(I_{1i} = 1) = B_i^{EBP}(I_i = 1) + E_m(n\bar{\Delta}(c)|s, Y_i = y_i),$$

where

$$\bar{\Delta}(c) = \frac{1}{n}\sum_{i\in S}\left[\psi\left(B_i^{EBP}(I_i = 1)\right) - B_i^{EBP}(I_i = 1)\right]$$

Then a conditional unbiased estimator of the conditional bias of the robust estimator can be expressed as

$$\hat{B}_i^R(I_{1i}=1) = \hat{B}_i^{EBP}(I_i=1) + \sum_{i\in S}\left[\psi\left(\hat{B}_i^{BLUP}(I_i=1)\right) - \hat{B}_i^{EBP}(I_i=1)\right]$$

where $\hat{B}_i^{EBP}(I_i=1)$ is a suitable estimator of $B_i^{EBP}(I_i=1)$.

Let $\hat{B}_{min} = \min\left(\hat{B}_i^{EBP}(I_i=1); c\right)$ and $\hat{B}_{max} = \max\left(\hat{B}_i^{EBP}(I_i=1); c\right)$, we can prove that the value of c that minimizes $\max\{\hat{B}_i^R(I_i=1|i\in S\}$ denoted by c_{minmax} , leads to the robust estimator :

$$\hat{\theta}^{REBP}(c_{minmax}) = \hat{\theta}^{EBP} - \frac{1}{2}(\hat{B}_{min} + \hat{B}_{max}). \tag{5}$$

Proof. We want to calibrate c to minimize $\max\{\hat{B}_i^R(I_{1i}=1|i\in S\}$. We have to resolve this optimization problem :

$$min_{c\in\mathbb{R}} max_{i\in S}\left(\hat{B}_i^{EBP}(I_i=1) + n\bar{\Delta}(c)\right)$$

The solution of this problem is the midrange $\frac{1}{2}(\hat{B}_{min} + \hat{B}_{max})$:

$$-n\bar{\Delta}(c) = \frac{1}{2}(\hat{B}_{min} + \hat{B}_{max})$$

so,

$$n\bar{\Delta}(c) = -\frac{1}{2}(\hat{B}_{min} + \hat{B}_{max}).$$

□

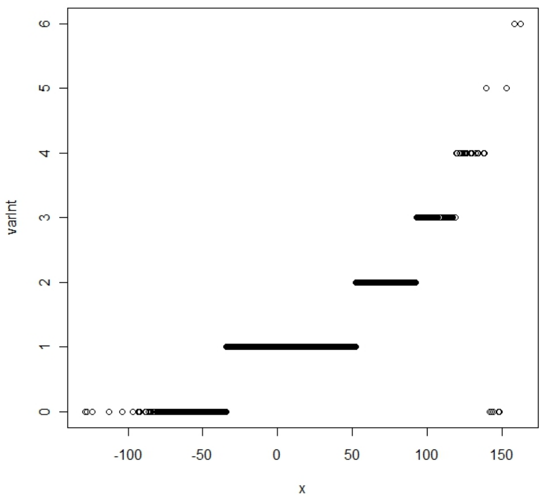
In some simulation studies, we are going to compare this robust estimator define by (5) to the EBP estimator with a robust estimator $\hat{\beta}^R$ given by Cantoni and Ronchetti (2001). More precisely, we are going to compare the robust estimator to :

$$\hat{\theta}^{RCantoni} = \sum_{i\in S} Y_i + \sum_{i\in U\setminus S} h(\mathbf{x_i^T}\hat{\beta}^R).$$

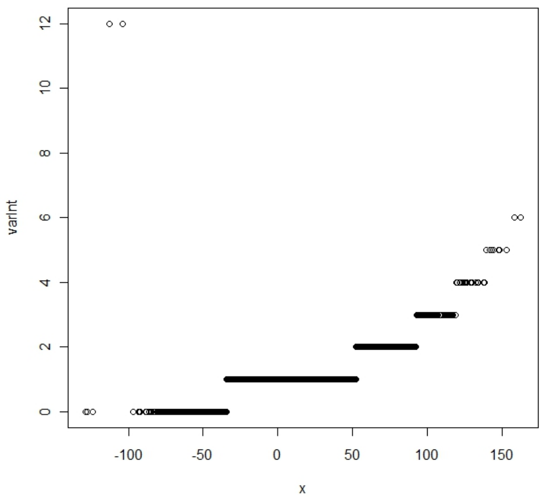
The robust estimator proposed by Cantoni and Ronchetti (2001) requires the determination of the tuning constant appearing in the Huber psy function. In the literature, the authors recommend using 1.345. This tuning constant allowed the robust methods to perform very well in a classical statistical context, but we know that in case of a finite population this choice of the tuning constant leads to robust estimators which are too biased and do not perform very well. This fact will be illustrated in the next simulation study. We had in the simulation a kind of oracle estimator with the tuning constant which minimize an estimation by Monte-Carlo of the mean square error.

1.4 Simulation study

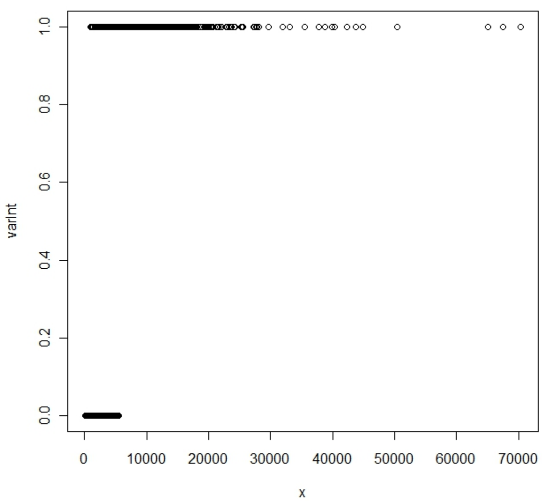
We are going to investigate the performance of the proposed robust estimator in terms of relative bias and relative efficiency. We generate some population which contains outliers and we limited our empirical study to logistic and Poisson cases. Here, $P = 5000$ populations are generated according to the four population models. The next 4 graphics represent one realisation of the population under the model. The population (1) and (3) are generated without outliers and the populations (2) and (4) are generated with 5% of outliers.



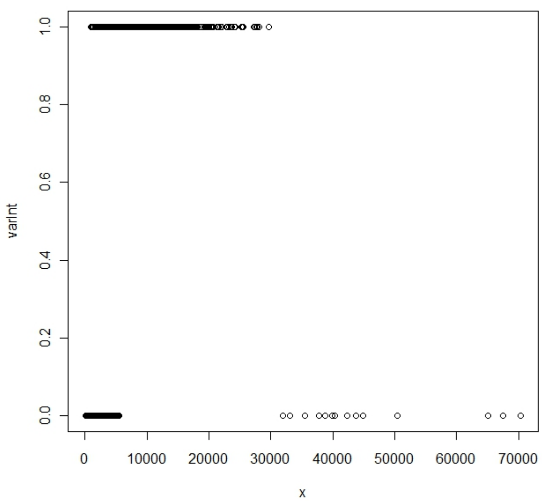
(a) Population 1



(b) Population 2



(c) Population 3



(d) Population 4

Figure 1: Representation of the four populations

For comparisons of estimators, we computed the Monte Carlo percent Relative Bias (RB), the Monte Carlo relative Variance and the Monte Carlo Relative Efficiency (RE) given by :

$$RB_{MC}\left(\hat{\theta}_p^R\right) = \frac{1}{P}\sum_{p=1}^{P}\frac{\left(\hat{\theta}_p^R - \theta_p\right)}{\theta_p} \times 100,$$

where

$$RV_{MC}\left(\hat{\theta}_p^R, \hat{\theta}\right) = \frac{\frac{1}{P}\sum_{p=1}^{P}\left(\hat{\theta}_p^R - E_{MC}\left(\hat{\theta}_p^R\right)\right)^2}{\frac{1}{P}\sum_{p=1}^{P}\left(\hat{\theta}_p - E_{MC}\left(\hat{\theta}_p\right)\right)^2} \times 100,$$

and

$$RE_{MC}\left(\hat{\theta}_p^R, \hat{\theta}\right) = \frac{\frac{1}{P}\sum_{p=1}^{P}\left(\hat{\theta}_p^R - \theta_p\right)^2}{\frac{1}{P}\sum_{p=1}^{P}\left(\hat{\theta}_p - \theta_p\right)^2} \times 100.$$

Population	Sample size	$\hat{\theta}^{RBLUP}$	$\hat{\theta}^{RCantoni}$	$\hat{\theta}^{RCantoni}(c_{opt})$
1	100	0.09(92)	6.35(684)	0.06(99)
	500	0.08(94)	5.9(279)	0.024(99)
2	100	−0.56(65)	5.1(203)	0.43(36)
	500	−0.38(79)	4.8(776)	0.38(41)
3	100	−0.10(101)	−0.52(177)	0.001(100)
	500	0.01(100)	0.22(115)	0.003(100)
4	100	0.18(93)	1.03(121)	0.0103(78)
	500	0.06(95)	1.5(99)	0.13(85)

Table 1: Bias and relative efficiency in brackets of the robust estimators

The results confirm our expectations regarding the behavior of the estimators : under the models corresponding to the populations (1) and (3), the robust estimator performs as well as the Empirical Best Predictor, which is the most efficient in these cases. Under the models (2) and (4) with some outliers, we can notice that the proposed robust estimator has a relative bias under 1% and there are significant improvements for the population 4 in terms of RMSE.

2 Robust Small area estimation using GLMMs

2.1 Small area estimation based on GLMMs

We focus now on the Generalized Linear Mixed Model (GLMM) and an extension in small area estimation. We adapt a little bit the notation introduce in the previous part, to extend the results in the case of small area estimation. Let U denote the finite population of size N, which is partitioned into k domains or small areas $U_1, ..U_k$ of sizes $N_1, ...N_k$, respectively. The domains sizes N_i are assumed to be known. Let y_{ij} be the values of y attached to the unit j in area i and let x_{ij} be a deterministic vector of dimension p containing the unit level covariates for the unit j in the area j. It is assumed that the values of x_{ij} are known for all units in the population. A sample s of size n is selected from U according to a non-informative sampling plan $p(s)$. Let $s_i = s \cap U_i$ be the sample of size n_i in the area i. The aim is to use the sample values of $y_{ij}$and the population values x_{ij} to infer the small area means $\theta_i = \frac{1}{N_i}\sum_{j\in U_i} y_{ij}$. Let $\mu = E\left(\boldsymbol{y}|\boldsymbol{u}\right)$ be the conditional mean vector with elements μ_{ij} and $\Sigma = Var\left(\boldsymbol{y}|\boldsymbol{u}\right)$ be the conditional covariance matrix which is diagonal with element σ_{ij}. Let us define the $N \times k$ matrix $Z = diag\left(1_{N_i}, i = 1, .., k\right)$ where 1_{N_i}corresponds to a vector of ones of dimension $N_i \times 1$.

In this paper, we assume a generalized linear mixed model for $\mu_{ij} = E\left[y_{ij}|\boldsymbol{u}_i\right]$ of the form

$$g(\mu_{ij}) = \eta_{ij} = x_{ij}^T\beta + u_i$$

where g is the link function, assumed to be known and invertible.

Then the approximation to the minimum mean square error predictor of $\theta_i = \frac{1}{N_i}\sum_{j\in U_i} y_{ij}$ is

$$\frac{1}{N_i}\left(\sum_{j\in s_i} y_{ij} + \sum_{j\in U_i\setminus s_i} \mu_{ij}\right).$$

Since μ_{ij} depends on β and $\boldsymbol{u}_i$, we need to provide an estimation of β and $\boldsymbol{u}_i$, and this leads to a Empirical Plug-in Predictor of the $i-th$ area mean,

$$\hat{\theta}_i^{EPP} = \frac{1}{N_i}\left(\sum_{j\in S_i} y_{ij} + \sum_{j\in U_i\setminus S_i} \hat{\mu}_{ij}\right)$$

where $\hat{\mu}_{ij} = h\left(x_{ij}^T\hat{\beta} + \hat{u}_i\right)$, $\hat{\beta}$ is the vector of the estimated fixed effect, $\hat{\boldsymbol{u}}_i$ is the predicted area random effect for the area i and h is the inverse of the link function g.

In a small-area framework, the EPP are efficient under the correct model specification and distributional assumptions, but they might be very sensitive to the presence of outliers, especially when the sample size in small, as in small-area. The main objective here is to propose new robust estimators for small-area means using the conditional bias as measure of influence.

2.2 Conditional bias of the EPP based on GLMMs

To provide an estimation of the conditional bias, we are going to work on a linear mixed model which approximates the original generalized mixed model.

Following González-Manteiga et al. (2007), we consider the first order Taylor expansion of $g(y_{ij})$ around μ_{ij},

$$g(y_{ij}) \simeq \eta_{ij} + (y_{ij} - \mu_{ij})\, g^{'}(\mu_{ij}) \triangleq \xi_{ij}. \tag{6}$$

The conditional moments of the working variables ξ_{ij} are given by

$$E\left(\xi_{ij}|\boldsymbol{u}\right) = \eta_{ij}, \quad Var\left(\xi_{ij}|\boldsymbol{u}\right) = g^{'}\left(\mu_{ij}\right)^2 \sigma_{ij}^2$$

and

$$Cov\left(\xi_{ij}, \xi_{i^{'}j^{'}}|\boldsymbol{u}\right) = 0, \text{ for } i \neq i^{'} \text{ or } j \neq j^{'}.$$

The unconditional mean of ξ_{ij} is $x_{ij}^T\beta$, and the random effects u_i are assumed to be independent, normally distributed, with zero mean and constant variances equal to σ_u.

Now, we can approximate the original generalized mixed model (6) by the linear mixed model

$$\xi_{ij} = x_{ij}^T\beta + u_i + e_{ij}, \; j = 1, ..., N_i, \; i = 1, ..., k,$$

where e_{ij} are independent random variables, independent of $\boldsymbol{u}$, with zero means and variance $v_{ij} = g^{'}\left(\mu_{ij}\right)^2 \sigma_{ij}^2$ and

$$Var\left(\xi_{ij}\right) = \sigma_u Z_s Z_s^T + \Sigma_{es} \triangleq V_s$$

where Σ_{es} is a diagonal matrix whose elements are the variances v_{ij} of the residuals e_{ij}.

In matrix notation, this model can be compactly rewritten as

$$\xi_i = X_i\beta + u_i 1_{n_i} + e_i \;\left(i = 1, ..., k\right),$$

where $X_i = (x_{i1}, .., x_{in_i})$ is a matrix of dimension $n_i \times p$ and 1_{n_i} corresponds to a vector of ones of dimension $n_i \times 1$. In matrix notation, the variance-covariance matrix of ξ_i is $V_i = \sigma_u^2 1_{n_i} 1_{n_i}^T + \Sigma_{es_i}$ where Σ_{es_i} correspond to the $i-th$ block of the matrix Σ_{es}.

We assumed in a first approach that the variance matrix V_s is known, then the best linear unbiased estimator of β and the best linear predictor of $\boldsymbol{u}$ in the linear mixed model are given by

$$\hat{\beta} = \left(\sum_{h=1}^{k} X_h^T V_h^{-1} X_h\right)^{-1} \sum_{h=1}^{k} X_h^T V_h^{-1} \xi_h \;\; \hat{u}_i = \sigma_u^2 1_{n_h} V_h^{-1}\left(\xi_h - X_h\hat{\beta}\right)$$

Following Dongmo Jiongo, Haziza, Duchesne (2013), we compute the conditional bias of a unit j in the area h for the Empirical Plug-in Predictor in the domain i , but we use a conditioning on all area-random effect instead of the random effect in the area i.

$$B_{ihj}\left(y_{hj}, u_h; \beta\right) = E_m\left\{\hat{\theta}_i^{EPP} - \theta_i | s,\, y_{hj},\, u\right\} \tag{7}$$

First the EPP can be decomposed as

$$
\begin{aligned}
\hat{\theta}_i^{EPP} &= \tfrac{1}{N_i} \left(\sum_{j \in S_i} y_{ij} + \sum_{j \in U_i \setminus S_i} h\left(x_{ij}^T \hat{\beta} + \hat{u}_i\right) \right) \\
&= \tfrac{1}{N_i} \left\{ \sum_{j \in S_i} y_{ij} + \sum_{j \in U_i \setminus S_i} \left[h\left(x_{ij}^T \hat{\beta} + \hat{u}_i\right) - h\left(x_{ij}^T \beta + u_i\right) \right] + \sum_{j \in U_i \setminus S_i} h\left(x_{ij}^T \beta + u_i\right) \right\} \\
&\simeq \tfrac{1}{N_i} \left\{ \sum_{j \in S_i} y_{ij} + \sum_{j \in U_i \setminus S_i} h\left(x_{ij}^T \beta + u_i\right) + A \right\}
\end{aligned}
$$

where

$$
\begin{aligned}
A &= \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) \left(x_{ij}^T \hat{\beta} + \hat{u}_i - x_{ij}^T \beta - u_i \right) \\
&= \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) \left(x_{ij}^T \left\{ \left[\sum_{h=1}^k X_h^T V_h^{-1} X_h \right]^{-1} \sum_{h=1}^k X_h^T V_h^{-1} \xi_h - \beta \right\} \right) \\
&+ \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) \left(\sigma_u^2 1_{n_i}^T V_i^{-1} \left[\xi_i - X_i \left[\sum_{h=1}^k X_h^T V_h^{-1} X_h \right]^{-1} \sum_{h=1}^k X_h^T V_h^{-1} \xi_h \right] - u_i \right) \\
&= -\sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) u_i + \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) \left(x_{ij}^T \left\{ \left[\sum_{h=1}^k X_h^T V_h^{-1} X_h \right]^{-1} \sum_{h=1}^k X_h^T V_h^{-1} \xi_h - \beta \right\} \right) \\
&+ \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) \left(\sigma_u^2 1_{n_i}^T V_i^{-1} \left[\xi_i - X_i \left[\sum_{h=1}^k X_h^T V_h^{-1} X_h \right]^{-1} \sum_{h=1}^k X_h^T V_h^{-1} \xi_h \right] \right)
\end{aligned}
$$

and

$$
\xi_h = X_h \beta + u_h + e_h.
$$

$$
\begin{aligned}
A &= -\sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) u_i + \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) \left(x_{ij}^T \left\{ \left[\sum_{h=1}^k X_h^T V_h^{-1} X_h \right]^{-1} \sum_{h=1}^k X_h^T V_h^{-1} [u_h + e_h] \right\} \right) \\
&+ \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) \left(\sigma_u^2 1_{n_i}^T V_i^{-1} \left[u_i + e_i - X_i \left[\sum_{h=1}^k X_h^T V_h^{-1} X_h \right]^{-1} \sum_{h=1}^k X_h^T V_h^{-1} [u_h + e_h] \right] \right) \\
&= -\sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}(\eta_{ij}) u_i + B
\end{aligned}
$$

Noting that $\sum_{h=1}^{k} X_h^T V_h^{-1} (u_h + e_h) = \sum_{h=1}^{k} \sum_{j \in S_h} X_h^T C_h^{(j)} (u_h + e_{hj})$ where $C_h^{(j)}$ corresponding to the $j - th$ column of $C_h = V_h^{-1}$, it can be shown that the second term B can be expressed as :

$$B = \sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} (u_h + e_{hj})$$

where

$$w_{ihj} = \begin{cases} k^{-1} a_i X_h^T C_h^{(j)} & j \in s_h \\ k^{-1} a_i X_i^T C_i^{(j)} + \left[\sum_{j' \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij'})\right] \sigma_u^2 1_{n_i}^T C_i^{(j)} & j \in s_i, \end{cases}$$

and

$$a_i = \left\{ \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij}) \left[x_{ij}^T - \sigma_u^2 1_{n_i}^T V_i^{-1} X_i\right] \right\} \left\{ k^{-1} \sum_{i=1}^{k} X_i^T V_i^{-1} X_i \right\}^{-1}.$$

Then

$$\begin{aligned} \hat{\theta}_i^{EPP} - \theta_i &= \frac{1}{N_i} \left\{ - \sum_{j \in U_i \setminus S_i} \left(y_{ij} - h\left(x_{ij}^T \beta + u_i\right) + \frac{\partial h}{\partial \eta} (\eta_{ij}) u_i \right) + \sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} (u_h + e_{hj}) \right\} \\ &= \frac{1}{N_i} \left\{ - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij}) \left(\frac{\partial g}{\partial \mu} (\mu_{ij}) \left[y_{ij} - h\left(x_{ij}^T \beta + u_i\right)\right] + u_i \right) + \sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} (u_h + e_{hj}) \right\} \\ &= \frac{1}{N_i} \left\{ - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij}) (e_{ij} + u_i) + \sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} (u_h + e_{hj}) \right\} \end{aligned} \tag{8}$$

To determine this conditional bias, we need to distinguish four cases, whether the unit j belongs to the domain i or not and whether the unit j is sampled or not and we have to keep in mind that the weights w_{ihj} depends on all area random effects $\boldsymbol{u}$. Now using the decomposition of the EPP (8) and the definition of the conditional bias (7), we have

$$B_{ihj}(y_{hj}, u_h; \beta) = \begin{cases} N_i^{-1} \left(\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij}) u_i + w_{iij} e_{ij} \right) & j \in s_i \\ N_i^{-1} \left(\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij}) u_i + w_{ihj} e_{hj} \right) & j \in s_h, h \neq i \\ N_i^{-1} \left(\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij}) u_i - \frac{\partial h}{\partial \eta} (\eta_{ij}) e_{ij} \right) & j \in U_i \setminus s_i \\ N_i^{-1} \left(\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta} (\eta_{ij}) u_i \right) & j \in U_h \setminus s_h, h \neq i. \end{cases}$$

In these expressions of the conditional bias, we can notice that a unit outside the area $j \in s_h$ may have a large influence if its weight w_{ihj} is large and its model residual e_{hj} is large. It is important to notice that even if non-sampled units may have large influences, it is not possible to reduce their impact at the estimation stage, because their conditional bias cant be estimated.

In the case of a linear mixed model, we find the same conditional bias as Dongmo Jiongo et al. (2013) with a conditioning on all area effect.

2.3 Construction of the robust estimator

Now, we can show that the prediction error of the EPP can be approximately written as :

$$\hat{\theta}_i^{EPP} - \theta_i \approx \sum_{h=1}^{k} \sum_{j \in U_h} B_{ihj}\left(y_{hj}, u_h; \beta\right) - \frac{N-1}{N_i}\left[\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}\left(\eta_{ij}\right) u_i\right]. \quad (9)$$

This expression suggests that the conditional bias can be interpreted as a contribution to the prediction error of the area means in the domain i. Following Beaumont et al. (2013) and Dongmo Jiongo et al. (2014), we define a robust predictor of θ_i as

$$\begin{aligned} \hat{\theta}_i^{REPP} &= \theta_i + \sum_{h=1}^{k} \sum_{j \in S_h} \phi_{d_1,d_2}\left\{B_{ihj}\left(y_{hj}, u_h; \beta\right)\right\} \\ &+ \sum_{h=1}^{k} \sum_{j \in U_h \setminus S_h} B_{ihj}\left(y_{hj}, u_h; \beta\right) \\ &- \frac{N-1}{N_i}\left[\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}\left(\eta_{ij}\right) u_i\right] \end{aligned} \quad (10)$$

where

$$\phi_{d_1,d_2}\left\{B_{ihj}\left(y_{hj}, u_h; \beta\right)\right\} =$$

$$\begin{cases} N_i^{-1}\left(\psi_{d1}\left\{w_{iij} e_{ij}\right\} + \psi_{d_2}\left\{\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}\left(\eta_{ij}\right) u_i\right\}\right) & j \in s_i \\ N_i^{-1}\left(\psi_{d1}\left\{w_{ihj} e_{hj}\right\} + \psi_{d_2}\left\{\sum_{h=1}^{k} \sum_{j \in S_h} w_{ihj} u_h - \sum_{j \in U_i \setminus S_i} \frac{\partial h}{\partial \eta}\left(\eta_{ij}\right) u_i\right\}\right) & j \in s_h, h \neq i \end{cases}$$

Using the expression (9) and (10) and choosing $d_2 = +\infty$, we have

$$\begin{aligned} \hat{\theta}^{EPP} &= \hat{\theta}_i^{EPP} - \sum_{h=1}^{k} \sum_{j \in S_h} B_{ihj}\left(y_{hj}, u_h; \beta\right) + \sum_{h=1}^{k} \sum_{j \in S_h} \phi_{d_1,+\infty}\left\{B_{ihj}\left(y_{hj}, u_h; \beta\right)\right\} \\ &= \hat{\theta}_i^{EPP} - \sum_{h=1}^{k} \sum_{j \in S_h} B'_{ihj}\left(y_{hj}, u_h; \beta\right) + \sum_{h=1}^{k} \sum_{j \in S_h} \phi_{d_1,+\infty}\left\{B'_{ihj}\left(y_{hj}, u_h; \beta\right)\right\} \end{aligned}$$

where

$$B'_{ihj}\left(y_{hj}, u_h; \beta\right) = \begin{cases} N_i^{-1} \psi_{d1}\left\{w_{iij} e_{ij}\right\} & j \in s_i \\ N_i^{-1} \psi_{d1}\left\{w_{ihj} e_{hj}\right\} & j \in s_h, h \neq i \end{cases}$$

if the influence of all the sample units is small we have

$$\phi_{d_1,+\infty}\left\{B'_{ihj}\left(y_{hj}, u_h; \beta\right)\right\} = B'_{ihj}\left(y_{hj}, u_h; \beta\right), \forall j \in S,$$

so the summation of the second and third term is close to zero, therefore the robust estimator is close to the non-robust one, i.e the EPP.

We have to determine the tuning constant d_1 which adjust the trade-off between bias and variance. Since, it is not possible to provide an analytic expression for the mean square error of the robust estimator, we choose the constant d_1 which verify a minmax criterion. We can use the same proof as in part 1 to show that the robust estimator construct with the minmax constant can be written as

$$\hat{\theta}^{EPP} = \hat{\theta}_i^{EPP} - \frac{1}{2}\left(\hat{B}'_{max} + \hat{B}'_{min}\right)$$

where $\hat{B}'_{max} = \max_{j\in S}\left\{\hat{B}'_{ihj}\left(y_{hj}, u_h; \beta\right)\right\}$ and $\hat{B}'_{min} = \min_{j\in S}\left\{\hat{B}'_{ihj}\left(y_{hj}, u_h; \beta\right)\right\}$, where $\hat{B}'_{ihj}\left(y_{hj}, u_h; \beta\right)$ is a suitable estimator of $B'_{ihj}\left(y_{hj}, u_h; \beta\right)$.

The conditional biases $B'_{ihj}\left(y_{hj}, u_h; \beta\right)$ are unknown since they depend on the model parameters (β, Σ_{es}), the random small-area effects $\boldsymbol{u}$, and the variance of the small area effects σ_u^2. The estimation of these parameters can be carried out by using a combination of Maximum Penalized Quasi-Likelihood (MPQL) for the estimation of β and $\boldsymbol{u}$, and REML for the estimation of the variance components (Saei and Chambers, 2003).

3 Final remarks

In this paper, we proposed an extension of Beaumont et al. (2013) to a model-based approach with the use of GLM, and we show empirically that the robust estimator perform very well in terms of MSE. We also focus on small area prediction for binary and count data, since it is an challenging problem. We proposed an extension of the results of Dongmo Jiongo et al. (2013) involving the conditional bias to binary and count data. The last part of the work, not presented here, is to test at least empirically the efficiency of the robust estimator compared to the non robust one and the M-quantile estimator proposed by Tzavidis et al. (2013) for count data and Chambers et al. (2014) for binary data. The estimation of the MSE of the proposed robust estimator is still an area of current research.

References

[1] Beaumont, J.-F., Haziza, D., and Ruiz-Gazen, A. A unified approach to robust estimation in finite population sampling. *Biometrika* (2013), ast010.

[2] Cantoni, E., and Ronchetti, E. Robust inference for generalized linear models. *Journal of the American Statistical Association 96*, 455 (2001).

[3] Chambers, R., Chandra, H., Salvati, N., and Tzavidis, N. Outlier robust small area estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*, 1 (2014), 47–69.

[4] Chambers, R., and Clark, R. *An introduction to model-based survey sampling with applications*. Oxford University Press, 2012.

[5] Chambers, R. L. Outlier robust finite population estimation. *Journal of the American Statistical Association 81*, 396 (1986), 1063–1069.

[6] Fabrizi, E., Salvati, N., Pratesi, M., and Tzavidis, N. Outlier robust model-assisted small area estimation. *Biometrical Journal 56*, 1 (2014), 157–175.

[7] Fuller, W. A. *Sampling statistics*, vol. 560. John Wiley & Sons, 2011.

[8] Ghosh, M., Natarajan, K., Stroud, T., and Carlin, B. P. Generalized linear models for small-area estimation. *Journal of the American Statistical Association 93*, 441 (1998), 273–282.

[9] González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Computational statistics & data analysis 51*, 5 (2007), 2720–2733.

[10] Jiang, J. Empirical best prediction for small-area inference based on generalized linear mixed models. *Journal of Statistical Planning and Inference 111*, 1 (2003), 117–127.

[11] Jiang, J., and Lahiri, P. Empirical best prediction for small area inference with binary data. *Annals of the Institute of Statistical Mathematics 53*, 2 (2001), 217–243.

[12] Jiongo, V. D., Haziza, D., and Duchesne, P. Controlling the bias of robust small-area estimators. *Biometrika* (2013), ast030.

[13] McCulloch, C. E., and Neuhaus, J. M. *Generalized linear mixed models*. Wiley Online Library, 2001.

[14] Rao, J. N. *Small area estimation*, vol. 331. John Wiley & Sons, 2005.

[15] Saei, A., and Chambers, R. Small area estimation under linear and generalized linear mixed models with time and area effects.

[16] Schall, R. Estimation in generalized linear models with random effects. *Biometrika 78*, 4 (1991), 719–727.

[17] Sinha, S. K., and Rao, J. Robust small area estimation. *Canadian Journal of Statistics 37*, 3 (2009), 381–399.

[18] Tzavidis, N., Ranalli, M. G., Salvati, N., Dreassi, E., and Chambers, R. Robust small area prediction for counts. *Statistical methods in medical research* (2014), 0962280214520731.

[19] Valliant, R., Dorfman, A. H., and Royall, R. M. Finite population sampling and inference: a prediction approach.