University of Wollongong

Research Online

Faculty of Engineering and Information Sciences - Papers: Part A

Faculty of Engineering and Information Sciences

1-1-2014

Outlier robust small area estimation

Ray Chambers
University of Wollongong, ray@uow.edu.au

Hukum Chandra
Indian Agricultural Statistics Research Institute, hchandra@uow.edu.au

Nicola Salvati
University of Pisa

Nikos Tzavidis
University of Southampton

Follow this and additional works at: https://ro.uow.edu.au/eispapers

Part of the Engineering Commons, and the Science and Technology Studies Commons

Recommended Citation

Chambers, Ray; Chandra, Hukum; Salvati, Nicola; and Tzavidis, Nikos, "Outlier robust small area estimation" (2014). *Faculty of Engineering and Information Sciences - Papers: Part A*. 1895.
https://ro.uow.edu.au/eispapers/1895

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Outlier robust small area estimation

Abstract
Recently proposed outlier robust small area estimators can be substantially biased when outliers are drawn from a distribution that has a different mean from that of the rest of the survey data. This naturally leads one to consider an outlier robust bias correction for these estimators. We develop this idea, proposing two different analytical mean-squared error estimators for the ensuing bias-corrected outlier robust estimators. Simulations based on realistic outlier-contaminated data show that the bias correction proposed often leads to more efficient estimators. Furthermore, the mean-squared error estimation methods proposed appear to perform well with a variety of outlier robust small area estimators.

Keywords
robust, small, outlier, area, estimation

Disciplines
Engineering | Science and Technology Studies

Publication Details
Chambers, R., Chandra, H., Salvati, N. & Tzavidis, N. (2014). Outlier robust small area estimation. Journal of the Royal Statistical Society Series B: Statistical Methodology, 76 (1), 47-69.

This journal article is available at Research Online: https://ro.uow.edu.au/eispapers/1895

Outlier Robust Small Area Estimation

Ray Chambers
University of Wollongong, Australia
Hukum Chandra
Indian Agricultural Statistics Research Institute, New Delhi
Nicola Salvati
University of Pisa, Italy
and Nikos Tzavidis
University of Southampton, United Kingdom.

Summary. Recently proposed outlier robust small area estimators can be substantially biased when outliers are drawn from a distribution that has a different mean from that of the rest of the survey data. This naturally leads one to consider an outlier robust bias correction for these estimators. In this paper we develop this idea, proposing two different analytical mean squared error estimators for the ensuing bias corrected outlier robust estimators. Simulations based on realistic outlier contaminated data show that the proposed bias correction often leads to more efficient estimators. Furthermore, the proposed mean squared error estimation methods appear to perform well with a variety of outlier robust small area estimators.

Keywords: Bias-variance trade-off; Linear mixed model; M-estimation; M-quantile model; Robust prediction; Robust bias correction.

1. Introduction

Outliers are a fact of life for any survey and as a result, a variety of methods have been devised to mitigate the effects of outlier values on survey estimates. Some of these, for example identification and removal of outlier data values by experienced data experts during survey processing, can be effective in ensuring that the resulting survey estimates are unaffected by these values. However, being somewhat subjective, such methods are not amenable to scientific evaluation. As a consequence there are a number of objective methods for survey estimation that use statistical rules to decide whether an observation is a potential outlier, and to down-weight its contribution to the survey estimates if this is the case. Generally, an outlier robust estimator of this type is based on the assumption that the non-sample data values all follow the assumed working model, and so these estimators aim to robustly estimating the expected value of the non-sample sum (or mean) of the study variable under this working model on the basis of the outlier-contaminated sample data. In practice, this often involves replacement of an outlying sample value by an estimate of what it should have been if in fact it had been generated under the working model. We refer to such methods as *Robust Projective* in what follows since they project sample non-outlier (i.e. working model) behaviour on to the non-sampled part of the survey population.

Robust Projective methods essentially emulate the subjective approach described earlier, and typically lead to biased estimators with lower variances than would otherwise be the case. The reason for the bias is not difficult to find – it is extremely unlikely that all the non-sampled values in the target population are drawn from the same distribution as the sample non-outliers, and yet these methods are built on precisely this assumption. Chambers (1986) recognised this dilemma and coined the concept of a 'representative outlier', i.e. a sample outlier that is potentially drawn from a group of population outliers and hence cannot be unit-weighted in estimation. He noted that representative outliers cannot be treated on the same basis in estimation as other sample data more consistent with the working model for the population, since such outlier values can seriously destabilise the survey estimates, and suggested the addition of an outlier robust bias correction term to a Robust Projective survey estimator, e.g. one based on outlier-robust estimates of the model parameters. Welsh and Ronchetti (1998) expand on this idea, applying it more generally to estimation of the finite population distribution of a survey variable in the presence of representative outliers. A similar idea is implicit in the approach described in Chambers *et al*. (1993), where a nonparametric bias correction is suggested. In what follows, we refer to methods that allow for contributions from representative sample outliers as *Robust Predictive* since they attempt to predict the contribution of the population outliers to the population quantity of interest.

If outliers are a concern for estimation of population quantities, it is safe to say that they are even more of a concern in small area estimation (SAE), where sample sizes are considerably smaller and model-dependent estimation is the norm. It is easy to see that an outlier that destabilises a population estimate based on a large survey sample will almost certainly destroy the validity of the corresponding direct estimate for the small area from which the outlier is sourced, since this estimate will be based on a much smaller sample. This problem does not disappear when the small area estimator is an indirect one, e.g. an Empirical Best Linear Unbiased Predictor (EBLUP), since the weights underpinning this estimator will still put most emphasis on data from the small area of interest, and the estimates of the model parameters underpinning the estimator will themselves be destabilised by the sample outliers. Consequently, it is of some interest to see how outlier robust survey estimation can be adapted to this situation.

Chambers and Tzavidis (2006) explicitly address this issue of outlier robustness in SAE, using an approach based on fitting outlier robust M-quantile models to the survey data. More recently, Sinha and Rao (2009) also address this issue from the perspective of linear mixed models. Both these approaches, however, use plug-in robust prediction. That is, they replace parameter estimates in optimal, but outlier sensitive, predictors by outlier robust versions (a Robust Projective approach). Unfortunately, although this approach typically leads to a low prediction variance, it can involve an unacceptable prediction bias in situations where the outliers are drawn from a distribution that has a different mean to the rest of the survey data.

After discussing Robust Projective estimators for small areas in Section 2, we explore the extension of the Robust Predictive approach to the SAE situation in Section 3. In Section 4 we propose two different analytical mean squared error (MSE) estimators for outlier robust predictors of small area means. In particular, the first proposal is based on the bias-robust mean squared error estimation approach described in Chambers *et al*. (2011) and represents an extension of the ideas in Royall and Cumberland (1978). The second MSE estimator is based on first order approximations to the variances of solutions of outlier robust estimating equations. We show how these two approaches can be useful for estimating the MSE of various small area predictors considered in this paper. In Sections 5 and 6 we use model-based simulations based on realistic outlier contaminated data scenarios as well as design-based simulations to evaluate how these two different approaches compare, both in terms of point estimation performance as well as in terms of MSE estimation performance. Section 7 concludes the paper with some final remarks, and a discussion of future research aimed at outlier robust small area inference.

2. Robust Projective Estimation for Small Areas

In what follows we assume that unit record data are available at small area level. For the sampled units in the population this consists of indicators of small area affiliation, values y_j of the variable of interest, values $\mathbf{x}_j$ of a $p \times 1$ vector of individual level covariates, and values $\mathbf{z}_j$ of a vector of area level covariates. For the non-sampled population units we do not know the values of y_j. However, it is assumed that all areas are sampled and that we know the numbers of such units in each small area and the respective small area averages of $\mathbf{x}_j$ and $\mathbf{z}_j$. We also assume that there is a linear relationship between y_j and $\mathbf{x}_j$ and that sampling is non-informative for the small area distribution of y_j given $\mathbf{x}_j$, allowing us to use population level models with the sample data.

Battese *et al.* (1988) introduced the use of linear mixed models for SAE, with random effects for the small areas of interest. See Rao (2003) for a comprehensive review of SAE based on these models. A more recent, and more compact, review is Chambers and Clark (2012, Chapter 15). Let $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{Z}$ denote the population level vector and matrices defined by y_j, $\mathbf{x}_j$ and $\mathbf{z}_j$ respectively. Then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{1}$$

where $\mathbf{u} = \left(\mathbf{u}_1^T, \cdots, \mathbf{u}_m^T\right)^T$ is a vector of dimension mq made up of m independent realisations $\{\mathbf{u}_i; i = 1, \cdots, m\}$ of a q-dimensional random area effect with $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ and $\mathbf{e} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_e)$ is a vector of N individual specific random effects. It is also assumed that $\mathbf{u}$ is distributed independently of $\mathbf{e}$. Here m is the total number of small areas that make up the population and q is the dimension of $\mathbf{z}_j$ so that $\mathbf{Z}$ is a $N \times mq$ matrix of fixed known constants. We assume that the covariance matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_e$ are defined in terms of a lower dimensional set of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$, which are typically referred to as the variance components of (1), while the vector $\boldsymbol{\beta}$ is usually referred to as its fixed effects parameter.

Let $\hat{\boldsymbol{\beta}}$ denote the estimate of the fixed effects parameter in (1) and let $\hat{\mathbf{u}} = \left(\hat{\mathbf{u}}_1^T, \cdots, \hat{\mathbf{u}}_m^T\right)^T$ denote the vector of predicted values of the random area effects in (1). The EBLUP of the area i mean of the y_j under (1) is then

$$\hat{\bar{y}}_i^{EBLUP} = N_i^{-1}\left\{n_i \bar{y}_{si} + \left(N_i - n_i\right)\left(\bar{\mathbf{x}}_{ri}^T \hat{\boldsymbol{\beta}} + \bar{\mathbf{z}}_{ri}^T \hat{\mathbf{u}}\right)\right\} \tag{2}$$

where we use indices of s and r to denote sample and non-sample quantities respectively. Thus, $\bar{y}_{si}$ is the average of the n_i sample values of y_j from area i and $\bar{\mathbf{x}}_{ri}$ and $\bar{\mathbf{z}}_{ri}$ denoting the vectors of average values of $\mathbf{x}_j$ and $\mathbf{z}_j$ respectively for the $N_i - n_i$ non-sampled units in the same area.

From a Robust Projective viewpoint, (2) can be made insensitive to sample outliers by replacing $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ by outlier robust alternatives. To motivate this approach, we initially assume the variance components $\boldsymbol{\theta}$

are known, so the covariance matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_e$ in (1) are known. Put $\mathbf{V}_s = \boldsymbol{\Sigma}_{es} + \mathbf{Z}_s \boldsymbol{\Sigma}_u \mathbf{Z}_s^T$ where $\boldsymbol{\Sigma}_{es}$ denotes the sample component of $\boldsymbol{\Sigma}_e$. Then the Best Linear Unbiased Estimator (BLUE) of the fixed effects parameter $\boldsymbol{\beta}$ and the Best Linear Unbiased Predictor (BLUP) of the random effects vector $\mathbf{u}$ are solutions to

$$\mathbf{X}_s^T \mathbf{V}_s^{-1} \left(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} \right) = \mathbf{0} \tag{3}$$

and

$$\boldsymbol{\Sigma}_u \mathbf{Z}_s^T \mathbf{V}_s^{-1} \left(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} \right) - \mathbf{u} = \mathbf{0} . \tag{4}$$

A straightforward way to make the solutions to (3) and (4) robust to sample outliers is therefore to replace them by

$$\mathbf{X}_s^T \mathbf{V}_s^{-1/2} \psi \left(\mathbf{V}_s^{-1/2} \left\{ \mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} \right\} \right) = \mathbf{0} \tag{5}$$

and

$$\boldsymbol{\Sigma}_u \mathbf{Z}_s^T \mathbf{V}_s^{-1/2} \psi \left(\mathbf{V}_s^{-1/2} \left\{ \mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} \right\} \right) - \boldsymbol{\Sigma}_u^{1/2} \psi \left(\boldsymbol{\Sigma}_u^{-1/2} \mathbf{u} \right) = \mathbf{0} . \tag{6}$$

Here ψ is a bounded influence function and $\psi(\mathbf{a})$ denotes the vector defined by applying ψ to every component of $\mathbf{a}$. Observe that the bounded influence function is applied separately to model residuals and to predicted area effects in (5) and (6), making the solutions to these estimating equations robust against individual as well as area outliers. Unfortunately, since $\mathbf{V}_s$ is not a diagonal matrix, the solution to (6) can be numerically unstable. An alternative approach was therefore suggested by Fellner (1986), who noted that any solution to (3) and (4) was also a solution to

$$\mathbf{X}_s^T \boldsymbol{\Sigma}_{es}^{-1} \left(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} - \mathbf{Z}_s \mathbf{u} \right) = \mathbf{0} \quad \text{and} \quad \mathbf{Z}_s^T \boldsymbol{\Sigma}_{es}^{-1} \left(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} - \mathbf{Z}_s \mathbf{u} \right) - \boldsymbol{\Sigma}_u^{-1} \mathbf{u} = \mathbf{0} .$$

Fellner (1986) suggested that these alternative estimating equations (and hence their solutions) can be made outlier robust by replacing them by

$$\mathbf{X}_s^T \boldsymbol{\Sigma}_{es}^{-1/2} \psi \left(\boldsymbol{\Sigma}_{es}^{-1/2} \left\{ \mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} - \mathbf{Z}_s \mathbf{u} \right\} \right) = \mathbf{0} \tag{7}$$

and

$$\mathbf{Z}_s^T \boldsymbol{\Sigma}_{es}^{-1/2} \psi \left(\boldsymbol{\Sigma}_{es}^{-1/2} \left\{ \mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta} - \mathbf{Z}_s \mathbf{u} \right\} \right) - \boldsymbol{\Sigma}_u^{-1/2} \psi \left(\boldsymbol{\Sigma}_u^{-1/2} \mathbf{u} \right) = \mathbf{0} . \tag{8}$$

Since (7) and (8) assume the variance components $\boldsymbol{\theta}$ are known, their usefulness is somewhat limited unless outlier robust estimators of these parameters can also be defined. Richardson and Welsh (1995) propose two outlier robust variations to the maximum likelihood estimating equations for $\boldsymbol{\theta}$. One of these (ML Proposal II) leads to an estimating equation for the variance component θ_k of $\boldsymbol{\theta}$ of the form

$$\psi\left\{\left(\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}\right)^T\mathbf{V}_s^{-1/2}\right\}\mathbf{V}_s^{-1/2}\left(\partial_{\theta_k}\mathbf{V}_s\right)\mathbf{V}_s^{-1/2}\psi\left\{\mathbf{V}_s^{-1/2}\left(\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}\right)\right\}=tr\left\{\mathbf{D}_n^{\psi}\left(\partial_{\theta_k}\mathbf{V}_s\right)\right\}, \tag{9}$$

where $\partial_{\theta_k}\mathbf{V}_s$ denotes the first order partial derivative of $\mathbf{V}_s$ with respect to the variance component θ_k and, for $Z \sim N(0,1)$, $\mathbf{D}_n^{\psi}=E\left\{\psi^2(Z)\right\}\mathbf{V}_s^{-1}$. Richardson and Welsh (1995) also proposed robust REML-type equations for $\boldsymbol{\theta}$ but unlike (9) these are not robust generalisations of REML estimating equations.

Sinha and Rao (2009) describe an approach to outlier robust estimation of $\boldsymbol{\beta}$ and $\mathbf{u}$ in (1) that builds on these results, substituting approximate solutions to both (5) and (9) into the Fellner estimating equation (8) to obtain an outlier robust predicted value of the area effect $\mathbf{u}$. In particular, their approach replaces (5) by

$$\mathbf{X}_s^T\mathbf{V}_s^{-1}\mathbf{U}_s^{1/2}\psi\left(\mathbf{U}_s^{-1/2}\left\{\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}\right\}\right)=\mathbf{0}, \tag{10}$$

where $\mathbf{U}_s=diag\left(\mathbf{V}_s\right)$, and replaces (9) by

$$\Psi(\boldsymbol{\theta})=\psi\left\{\left(\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}\right)^T\mathbf{U}_s^{-1/2}\right\}\mathbf{U}_s^{1/2}\mathbf{V}_s^{-1}\left(\partial_{\theta_k}\mathbf{V}_s\right)\mathbf{V}_s^{-1}\mathbf{U}_s^{1/2}\psi\left\{\mathbf{U}_s^{-1/2}\left(\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}\right)\right\}=tr\left\{\mathbf{D}_n^{\psi}\left(\partial_{\theta_k}\mathbf{V}_s\right)\right\}. \tag{11}$$

Since the solutions to (10) and (11) depend on the influence function ψ, we denote them by a superscript of ψ below. The Sinha and Rao (2009) Robust Projective alternative to (2) is then

$$\hat{\bar{y}}_i^{SR}=\bar{\mathbf{x}}_i^T\hat{\boldsymbol{\beta}}^{\psi}+\bar{\mathbf{z}}_i^T\hat{\mathbf{u}}^{\psi}. \tag{12}$$

Note that (12) estimates the area i mean under (1). A minor modification restricts this to the mean of the non-sampled units in area i, in which case (12) becomes

$$\hat{\bar{y}}_i^{REBLUP}=N_i^{-1}\left\{n_i\bar{y}_{si}+\left(N_i-n_i\right)\left(\bar{\mathbf{x}}_{ri}^T\hat{\boldsymbol{\beta}}^{\psi}+\bar{\mathbf{z}}_{ri}^T\hat{\mathbf{u}}^{\psi}\right)\right\}. \tag{13}$$

From now on, we refer to (13) as the Robust EBLUP or REBLUP.

An alternative methodology for outlier robust SAE is the M-quantile regression-based method described by Chambers and Tzavidis (2006). This is based on a linear model for the M-quantile regression of $\mathbf{y}$ on $\mathbf{X}$, i.e.

$$m_q(\mathbf{y}\,|\,\mathbf{X})=\mathbf{X}\boldsymbol{\beta}_q, \tag{14}$$

where $m_q(\mathbf{y}|\mathbf{X})$ denotes the M-quantile of order q of the conditional distribution of $\mathbf{y}$ given $\mathbf{X}$. An estimate $\hat{\boldsymbol{\beta}}_q$ of $\boldsymbol{\beta}_q$ can be calculated for any value of q in the interval (0,1), and for each unit in the sample we define its unique M-quantile coefficient under this fitted model as the value q_j such that $y_j = \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_{q_j}$, with the sample average of these coefficients in area *i* denoted by $\bar{q}_i$. The M-quantile estimate of the mean of y_j in area *i*, hereafter MQ, is then

$$\hat{\bar{y}}_i^{MQ} = N_i^{-1}\left\{ n_i \bar{y}_{si} + \left(N_i - n_i\right) \bar{\mathbf{x}}_{ri}^T \hat{\boldsymbol{\beta}}_{\bar{q}_i} \right\}. \tag{15}$$

Note that the regression M-quantile model (14) depends on the influence function ψ and so does MQ. When this function is bounded, sample outliers have limited impact on $\hat{\boldsymbol{\beta}}_q$. That is, (15) corresponds to assuming that all non-sample units in area *i* follow the working model (14) with $q = \bar{q}_i$, in the sense that one can write $y_j = \mathbf{x}_j^T \boldsymbol{\beta}_{\bar{q}_i} + \textit{individual level noise}$ for all such units.

3. Robust Predictive Estimation for Small Areas

A problem with the Robust Projective approach is that it assumes all non-sampled units follow the working model, or, in what essentially amounts to the same thing, that any deviations from this model are noise and so cancel out 'on average'. Thus, under the linear mixed model (1) one can see that provided the individual errors of the non-sampled units are symmetrically distributed about zero, the REBLUP (13) suggested by Sinha and Rao (2009) will perform well since it is based on the implicit assumption that the average of these errors over the non-sampled units in area *i* converges to zero. The M-quantile estimator MQ (15) of Chambers and Tzavidis (2006) is no different since it assumes that the errors $y_j - \mathbf{x}_j^T \boldsymbol{\beta}_{\bar{q}_i}$ from the area *i*-specific M-quantile regression model are 'noise' and hence also cancel out on average. Note that this does not mean that these non-sample units are not outliers. It is just that our best prediction of the corresponding small area average value of their model errors is zero.

Welsh and Ronchetti (1998) consider the issue of outlier robust prediction within the context of population level survey estimation. Starting with a working linear model linking the population values of y_j and $\mathbf{x}_j$, and sample data containing representative outliers, they extend the approach of Chambers (1986) to robust prediction of the empirical distribution function of the population values of y_j. Their argument immediately applies to robust prediction of the empirical distribution function of the area *i* values of y_j, and leads to a predictor of the form

$$\hat{F}_i^{\psi\phi}(t) = N_i^{-1}\left[\sum_{j\in s_i} I(y_j \le t) + n_i^{-1}\sum_{j\in s_i}\sum_{k\in r_i} I\left(\mathbf{x}_k^T\hat{\boldsymbol{\beta}}^{\psi} + \omega_{ij}^{\psi}\phi\left\{\left(y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}^{\psi}\right)\middle/\omega_{ij}^{\psi}\right\} \le t\right)\right]. \tag{16}$$

Here $\hat{\boldsymbol{\beta}}^{\psi}$ denotes an M-estimator of the regression parameter in the linear working model based on a bounded influence function ψ, ω_{ij}^{ψ} is a robust estimator of the scale of the residual $y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}^{\psi}$ in area *i* and ϕ denotes a bounded influence function that satisfies $|\phi| \ge |\psi|$. Tzavidis *et al.* (2010) note that the robust estimator of the area *i* mean of the y_j consistent with (16) is just the expected value functional defined by it, which is

$$\hat{\bar{y}}_i^{\psi\phi} = \int t d\hat{F}_i^{\psi\phi}(t) = N_i^{-1}\left[n_i\bar{y}_{si} + \left(N_i - n_i\right)\left(\bar{\mathbf{x}}_{ri}^T\hat{\boldsymbol{\beta}}^{\psi} + n_i^{-1}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left\{\left(y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}^{\psi}\right)\middle/\omega_{ij}^{\psi}\right\}\right)\right]. \tag{17}$$

These authors therefore suggest an extension to the M-quantile estimator (15) by replacing $\hat{\boldsymbol{\beta}}^{\psi}$ in (17) by $\hat{\boldsymbol{\beta}}_{\bar{q}_i}$, which leads to a bias-corrected version of (15), hereafter MQ-BC, given by

$$\hat{\bar{y}}_i^{MQ-BC} = N_i^{-1}\left[n_i\bar{y}_{si} + \left(N_i - n_i\right)\left(\bar{\mathbf{x}}_{ri}^T\hat{\boldsymbol{\beta}}_{\bar{q}_i} + n_i^{-1}\sum_{j\in s_i}\omega_{ij}^{MQ}\phi\left\{\left(y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i}\right)\middle/\omega_{ij}^{MQ}\right\}\right)\right]. \tag{18}$$

Here ω_{ij}^{MQ} is a robust estimator of the scale of the residual $y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i}$ in area *i*.

The use of the two influence functions ψ and ϕ in (18) is worthy of comment. The first, ψ, underpins $\hat{\boldsymbol{\beta}}_q$, and hence $\hat{\boldsymbol{\beta}}_{\bar{q}_i}$. Its purpose is to ensure that sample outliers have little or no influence on the fit of the working M-quantile model. As a consequence it is bounded and so downweights these outliers. The second, ϕ, is still bounded but 'less restrictive' than ψ (since $|\phi| \ge |\psi|$), and its purpose is to define an adjustment for the bias caused by the fact that the first two terms on the right hand side of (18) treat sample outliers as self-representing. A similar argument can be used to modify REBLUP (13). In particular, a Robust Predictive version of this estimator, hereafter REBLUP-BC, mimics the bias correction idea used in (18) and leads to

$$\hat{\bar{y}}_i^{REBLUP-BC} = \hat{\bar{y}}_i^{REBLUP} + \left(1 - n_i N_i^{-1}\right) n_i^{-1} \sum_{j \in s_i} \omega_{ij}^{\psi} \phi \left\{ \left(y_j - \mathbf{x}_j^T \hat{\boldsymbol{\beta}}^{\psi} - \mathbf{z}_j^T \hat{\mathbf{u}}^{\psi} \right) \Big/ \omega_{ij}^{\psi} \right\}, \tag{19}$$

where the ω_{ij}^{ψ} are now robust estimates of the scale of the area *i* residuals $y_j - \mathbf{x}_j^T \hat{\boldsymbol{\beta}}^{\psi} - \mathbf{z}_j^T \hat{\mathbf{u}}^{\psi}$.

4. MSE Estimation for Robust Predictors

In this Section we propose two different analytic methods of MSE estimation for robust predictors of small area means under the Robust Projective and Robust Predictive approaches. Both are developed on the assumption that the working model for inference conditions on the realised values of the area effects, and so the proposed MSE estimators are conditional estimators. In Section 4.1 we apply the ideas set out by Chambers *et al*. (2011) to define a pseudo-linearization estimator of the conditional MSE of REBLUP (13). Similar conditional MSE estimators for REBLUP-BC (19), MQ (15) and MQ-BC (18) follow directly. In Section 4.2 we use first order approximations to the variances of solutions of estimating equations to develop conditional MSE estimators for REBLUP (13) and REBLUP-BC (19). Analogous MSE estimators for MQ (15) and MQ-BC (18) based on this approach are described in the Appendix.

4.1 Pseudo-linearization approach to MSE estimation for robust small area predictors

Sinha and Rao (2009) propose a parametric bootstrap-based estimator for the MSE of REBLUP. Here we describe an analytical estimator of the conditional MSE of REBLUP that is less computationally demanding. The proposed estimator is based on the pseudo-linearization approach to MSE estimation described by Chambers *et al*. (2011), which can be used for predictors that can be expressed as weighted sums of the sample values. Since REBLUP can be expressed in a pseudo-linear form, i.e. as a weighted sum of the sample values of *y*, this approach is immediately applicable. To start, we note that under (1), and assuming that the variance components are known, the Robust BLUP or RBLUP of $\bar{y}_i$ can be expressed as

$$\hat{\bar{y}}_i^{RBLUP} = \sum_{j \in s} w_{ij}^{RBLUP} y_j = \left(\mathbf{w}_{is}^{RBLUP} \right)^T \mathbf{y}_s, \tag{20}$$

where

$$\left(\mathbf{w}_{is}^{RBLUP} \right)^T = N_i^{-1} \left\{ \mathbf{1}_s^T + \left(N_i - n_i \right) \left[\bar{\mathbf{x}}_{ir}^T \mathbf{A}_s + \bar{\mathbf{z}}_{ir}^T \mathbf{B}_s \left(\mathbf{I}_s - \mathbf{X}_s \mathbf{A}_s \right) \right] \right\}.$$

Here

- $\mathbf{A}_s = \left(\mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{U}_s^{1/2} \mathbf{W}_{1s} \mathbf{U}_s^{-1/2} \mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{U}_s^{1/2} \mathbf{W}_{1s} \mathbf{U}_s^{-1/2}$, with $\mathbf{W}_{1s}$ a $n \times n$ diagonal matrix of weights with j-th component $w_{1j} = \psi\left(U_j^{-1/2}\left\{y_j - \mathbf{x}_j^T \tilde{\boldsymbol{\beta}}^\psi\right\}\right) \Big/ U_j^{-1/2}\left\{y_j - \mathbf{x}_j^T \tilde{\boldsymbol{\beta}}^\psi\right\}$;
- $\mathbf{B}_s = \left(\mathbf{Z}_s^T \boldsymbol{\Sigma}_{es}^{-1/2} \mathbf{W}_{2s} \boldsymbol{\Sigma}_{es}^{-1/2} \mathbf{Z}_s + \boldsymbol{\Sigma}_u^{-1/2} \mathbf{W}_{3s} \boldsymbol{\Sigma}_u^{-1/2}\right)^{-1} \left(\mathbf{Z}_s^T \boldsymbol{\Sigma}_{es}^{-1/2} \mathbf{W}_{2s} \boldsymbol{\Sigma}_{es}^{-1/2}\right)$, with $\mathbf{W}_{2s}$ a $n \times n$ diagonal matrix of weights with j-th component $w_{2j} = \psi\left(\left(\sigma_e^\psi\right)^{-1}\left\{y_j - \mathbf{x}_j^T \tilde{\boldsymbol{\beta}}^\psi - \mathbf{z}_j \tilde{\mathbf{u}}^\psi\right\}\right) \Big/ \left(\sigma_e^\psi\right)^{-1}\left\{y_j - \mathbf{x}_j^T \tilde{\boldsymbol{\beta}}^\psi - \mathbf{z}_j \tilde{\mathbf{u}}^\psi\right\}$, and $\mathbf{W}_{3s}$ is a $m \times m$ diagonal matrix of weights with i-th component $w_{3i} = \psi\left(\left(\sigma_u^\psi\right)^{-1} \tilde{u}_i^\psi\right) \Big/ \left(\sigma_u^\psi\right)^{-1} \tilde{u}_i^\psi$;
- $\tilde{\boldsymbol{\beta}}^\psi$ and $\tilde{u}_i^\psi$ are the solutions to (10) and (11) when variance components are known.

In addition, $\mathbf{1}_s$ is the n-vector with j-th component equal to one whenever the corresponding sample unit is in area i and is zero otherwise. The REBLUP (13) can be expressed in exactly the same way, except that all quantities in the weight vector $\mathbf{w}_{is}^{RBLUP}$ that depend on (unknown) variance components now need a 'hat', in which case we denote it by $\mathbf{w}_{is}^{REBLUP}$. Given this pseudo-linear representation for REBLUP, a simple first order approximation to its MSE is developed assuming the conditional version of the model (1), i.e. the random effects are considered to be fixed, but unknown, quantities. Let $I(j \in i)$ denote the indicator for whether unit j is in area i. The estimator of the conditional MSE of REBLUP is then

$$\widehat{MSE}\left(\hat{\bar{y}}_i^{REBLUP}\right) = \hat{V}\left(\hat{\bar{y}}_i^{REBLUP}\right) + \left\{\hat{B}\left(\hat{\bar{y}}_i^{REBLUP}\right)\right\}^2, \tag{21}$$

where

$$\hat{V}(\hat{\bar{y}}_i^{REBLUP}) = N_i^{-2} \sum_{j \in s} \left\{a_{ij}^2 + (N_i - n_i) n^{-1}\right\} \lambda_j^{-1} (y_j - \hat{\mu}_j)^2$$

is the estimate of the conditional prediction variance of (13), with $a_{ij} = N_i w_{ij}^{REBLUP} - I(j \in i)$ and

$$\hat{B}(\hat{\bar{y}}_i^{REBLUP}) = \sum_{j \in s} w_{ij}^{REBLUP} \hat{\mu}_j - N_i^{-1} \sum_{j \in (r_i \cup s_i)} \hat{\mu}_j$$

is the estimate of its conditional prediction bias. In order to calculate (21) we need to define $\hat{\mu}_j$ and $\hat{\lambda}_j$. Here $\hat{\mu}_j = \sum_{k \in s} \phi_{kj} y_k$ is an unbiased linear estimator of the conditional expected value $\mu_j = E\left(y_j \mid \mathbf{x}_j, \mathbf{u}^\psi\right)$ and $\lambda_j = \left\{1 - 2\phi_{jj} + \sum_{k \in s} \phi_{kj}^2\right\}$ is a scaling constant. Because of the well-known shrinkage effect associated with BLUPs, replacing $\hat{\mu}_j$ by the EBLUP of μ_j under (1) can lead to biased estimation of the conditional prediction variance. Chambers *et al*. (2011) therefore recommend that $\hat{\mu}_j$ be computed as the 'unshrunken' version of the EBLUP for μ_j. See also Salvati *et al*. (2012). Note that the MSE estimator (21) ignores the extra variability associated with estimation of the variance components, and hence is a first order approximation to the actual conditional MSE of REBLUP.

The MSE estimator for REBLUP-BC (19) is obtained using the same pseudo-linearization approach as outlined above. The only difference is that the weights w_{ij}^{REBLUP} used in (21) are now replaced by corresponding REBLUP-BC weights. Furthermore, since REBLUP-BC is an approximately unbiased estimator of the small area mean, the squared bias term in (21) is omitted.

It has been empirically demonstrated that this method of MSE has good repeated sampling properties for realistic small area applications - see Chandra and Chambers (2009), Chambers and Tzavidis (2006), Chandra *et al*. (2007), Tzavidis *et al*. (2010) and Salvati *et al*. (2010). Although empirical results (see Chambers *et al.*, 2011) show that (21) performs well in terms of bias, this improved bias performance comes at the cost of increased mean squared error, mainly due to variability of the squared bias term in this situation. In particular, when the area-specific sample sizes are very small, the use of (21) can lead to MSE estimates with high MSE.

4.2 Linearization based MSE estimation for small area predictors

In what follows we build on the linearization ideas set out in Booth and Hobert (1998) to propose a new estimator of the MSE of a small area estimator that is defined by the solution of a set of robust estimating equations. The MSE is shown to be a sum of a prediction variance, a squared bias term and a correction term that accounts for the sampling variability of parameter estimates. Our theoretical development is based on approximations that correspond to assuming that $\max(n_i) = O(1)$, so that, as $m \to \infty$, the prediction variance and the squared bias are $O(1)$ and the correction term is $O(m^{-1})$. We also make the standard assumption that a consistent estimator of the MSE of a linear approximation to the small area estimator of interest can be used as its MSE estimator. As noted by Harville and Jeske (1992), such an approach will not generally be consistent, and the resulting MSE estimator can be biased low. In small sample problems this is not generally an issue. However, it needs to be kept in mind in what follows.

We illustrate this approach by applying it to estimation of the conditional MSE of REBLUP (13) and REBLUP-BC (19). The corresponding MSE estimator for the EBLUP (2) can be obtained as a special case of the MSE estimator of REBLUP (13). In order to conserve space, the development omits some technical details, but these are available from the authors upon request. Note that when used with an estimator based on a mixed model, the proposed MSE estimator provides a second order approximation to the conditional MSE, since it includes a term for the contribution to the variability resulting from the estimation of the variance components. Throughout, we assume use of a Huber Proposal 2 influence function with tuning constant c. We also assume the regularity conditions (RC1) to (RC7) set out in the Appendix.

Under model (1) the conditional prediction variance of the Robust BLUP (RBLUP) of $\bar{y}_i$ can be expressed as

$$Var_{\mathbf{u}}\left(\hat{\bar{y}}_i^{RBLUP}-\bar{y}_i\right)=\left(1-n_iN_i^{-1}\right)^2\left(\bar{\mathbf{x}}_{ri}^T \quad \bar{\mathbf{z}}_{ri}^T\right)Var_{\mathbf{u}}\left(\tilde{\boldsymbol{\delta}}\right)\left(\bar{\mathbf{x}}_{ri}^T \quad \bar{\mathbf{z}}_{ri}^T\right)^T+\left(1-n_iN_i^{-1}\right)^2 Var_{\mathbf{u}}\left(\bar{e}_{ri}\right), \tag{22}$$

where $\tilde{\boldsymbol{\delta}}=\left(\tilde{\boldsymbol{\beta}}^{\psi T},\tilde{\mathbf{u}}^{\psi T}\right)^T$ with corresponding 'true' value $\boldsymbol{\delta}_0=\left(\boldsymbol{\beta}_0^{\psi T},\mathbf{u}_0^{\psi T}\right)^T$. Here $\bar{e}_{ri}=\left(N_i-n_i\right)^{-1}\sum_{j\in r_i}\left(y_j-\mathbf{x}_j^T\boldsymbol{\beta}_0^{\psi}-\mathbf{z}_j^T\mathbf{u}_0^{\psi}\right)$, $\tilde{\mathbf{u}}^{\psi}=\left(\tilde{\mathbf{u}}_1^{\psi T},\ldots,\tilde{\mathbf{u}}_m^{\psi T}\right)^T$ and a subscript of $\mathbf{u}$ is used to denote moments that are conditioned on the realised values of the area effects. In order to estimate (22) we need to estimate $Var_{\mathbf{u}}\left(\tilde{\boldsymbol{\delta}}\right)$. From (8) and (10) we see that $\mathbf{H}(\tilde{\boldsymbol{\delta}})=\mathbf{0}$ where

$$\mathbf{H}(\boldsymbol{\delta})=\begin{pmatrix}\mathbf{H}_{\beta^{\psi}}(\boldsymbol{\delta})\\ \mathbf{H}_{u^{\psi}}(\boldsymbol{\delta})\end{pmatrix}=\begin{pmatrix}\mathbf{X}_s^T\mathbf{V}_s^{-1}\mathbf{U}_s^{1/2}\psi\left(\mathbf{U}_s^{-1/2}\left\{\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}^{\psi}\right\}\right)\\ \mathbf{Z}_s^T\boldsymbol{\Sigma}_{es}^{-1/2}\psi\left(\boldsymbol{\Sigma}_{es}^{-1/2}\left\{\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}^{\psi}-\mathbf{Z}_s\mathbf{u}^{\psi}\right\}\right)-\boldsymbol{\Sigma}_u^{-1/2}\psi\left(\boldsymbol{\Sigma}_u^{-1/2}\mathbf{u}^{\psi}\right)\end{pmatrix}.$$

We compute the asymptotic variance of solutions to an estimating equation to obtain a first order approximation to $Var_{\mathbf{u}}(\tilde{\boldsymbol{\delta}})$ and by extension to the conditional prediction variance of RBLUP. Given (RC1) and (RC2), and following the same argument as in Booth and Hobert (1998), leads to the first order approximation

$$Var_{\mathbf{u}}(\tilde{\boldsymbol{\delta}})=\left\{E_{\mathbf{u}}\left(\partial_{\boldsymbol{\delta}}\mathbf{H}_0\right)\right\}^{-1}Var_{\mathbf{u}}\left\{\mathbf{H}(\boldsymbol{\delta}_0)\right\}\left[\left\{E_{\mathbf{u}}\left(\partial_{\boldsymbol{\delta}}\mathbf{H}_0\right)\right\}^{-1}\right]^T+o\left(m^{-1}\right).$$

After some simplification this approximation suggests the sandwich-type estimator of $Var_{\mathbf{u}}(\tilde{\boldsymbol{\delta}})$:

$$\hat{V}_{\mathbf{u}}\left(\tilde{\boldsymbol{\delta}}\right)=\left\{E_{\mathbf{u}}\left(\partial_{\boldsymbol{\delta}}\mathbf{H}_0\right)\right\}^{-1}\begin{bmatrix}\hat{V}_{\mathbf{u}}\left(\mathbf{H}_{0\beta^{\psi}}\right) & \widehat{Cov}_{\mathbf{u}}\left(\mathbf{H}_{0\beta^{\psi}},\mathbf{H}_{0u^{\psi}}\right)\\ \widehat{Cov}_{\mathbf{u}}\left(\mathbf{H}_{0u^{\psi}},\mathbf{H}_{0\beta^{\psi}}\right) & \hat{V}_{\mathbf{u}}\left(\mathbf{H}_{0u^{\psi}}\right)\end{bmatrix}\left[\widehat{\left\{E_{\mathbf{u}}\left(\partial_{\boldsymbol{\delta}}\mathbf{H}_0\right)\right\}^{-1}}\right]^T,$$

where

$$\widehat{\left\{E_{\mathbf{u}}\left(\partial_{\boldsymbol{\delta}}\mathbf{H}_0\right)\right\}^{-1}}=\begin{bmatrix}\left[\hat{E}_{\mathbf{u}}\left\{\partial_{\beta_0^{\psi}}\mathbf{H}_{0\beta^{\psi}}\right\}\right]^{-1} & -\left[\hat{E}_{\mathbf{u}}\left\{\partial_{\beta_0^{\psi}}\mathbf{H}_{0\beta^{\psi}}\right\}\right]^{-1}\hat{E}_{\mathbf{u}}\left\{\partial_{\beta_0^{\psi}}\mathbf{H}_{0u^{\psi}}\right\}\left[\hat{E}_{\mathbf{u}}\left\{\partial_{u_0^{\psi}}\mathbf{H}_{0u^{\psi}}\right\}\right]^{-1}\\ \mathbf{0} & \left[\hat{E}_{\mathbf{u}}\left\{\partial_{u_0^{\psi}}\mathbf{H}_{0u^{\psi}}\right\}\right]^{-1}\end{bmatrix},$$

with

- $\hat{E}_{\mathbf{u}}\left\{\partial_{\beta_0^{\psi}}\mathbf{H}_{0\beta^{\psi}}\right\}=-\mathbf{X}_s^T\mathbf{V}_s^{-1}\mathbf{U}_s^{1/2}\mathbf{R}\mathbf{U}_s^{-1/2}\mathbf{X}_s$;
- $\hat{E}_{\mathbf{u}}\left\{\partial_{u_0^{\psi}}\mathbf{H}_{0u^{\psi}}\right\}=-\mathbf{Z}_s^T\boldsymbol{\Sigma}_{es}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}_{es}^{-1/2}\mathbf{Z}_s-\boldsymbol{\Sigma}_u^{-1/2}\mathbf{D}\boldsymbol{\Sigma}_u^{-1/2}$;
- $\hat{E}_{\mathbf{u}}\left\{\partial_{\beta_0^{\psi}}\mathbf{H}_{0u^{\psi}}\right\}=-\mathbf{Z}_s^T\boldsymbol{\Sigma}_{es}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}_{es}^{-1/2}\mathbf{X}_s$
- $\hat{V}_{\mathbf{u}}\left\{\mathbf{H}_{0\beta^{\psi}}\right\}=\left(n-p\right)^{-1}\sum_{j=1}^{n}\psi^2(r_j)\mathbf{X}_s^T\mathbf{V}_s^{-1}\mathbf{U}_s\mathbf{V}_s^{-1}\mathbf{X}_s$;
- $\hat{V}_{\mathbf{u}}\left\{\mathbf{H}_{0u^{\psi}}\right\}=\left(n-p\right)^{-1}\sum_{j=1}^{n}\psi^2(t_j)\mathbf{Z}_s^T\boldsymbol{\Sigma}_{es}^{-1}\mathbf{Z}_s$; and
- $\widehat{Cov}_{\mathbf{u}}\left(\mathbf{H}_{0u^{\psi}},\mathbf{H}_{0\beta^{\psi}}\right)=\left(n-p\right)^{-1}\sum_{j=1}^{n}\left\{\psi\left(r_j\right)\psi\left(t_j\right)\right\}\mathbf{X}_s^T\mathbf{V}_s^{-1}\mathbf{U}_s^{1/2}\boldsymbol{\Sigma}_{es}^{-1/2}\mathbf{Z}_s$.

Here $\mathbf{R}$ is a $n\times n$ diagonal matrix with *j*-th diagonal element equal to 1 if $-c<r_j<c$, 0 otherwise, $\mathbf{T}$ is a diagonal matrix of dimension $n\times n$ with *j*-th diagonal element equal to 1 if $-c<t_j<c$, 0 otherwise, and $\mathbf{D}$ is a $m\times m$ diagonal matrix with *i*-th diagonal element equal to 1 if $-c<d_i<c$, 0 otherwise. From (22) an estimator of the conditional prediction variance of RBLUP can then be written as

$$\hat{V}_{\mathbf{u}}\left(\hat{\bar{y}}_i^{RBLUP}-\bar{y}_i\right)=h_{1i}(\tilde{\boldsymbol{\delta}})+h_{2i}(\tilde{\boldsymbol{\delta}}),\tag{23}$$

where

- $h_{1i}(\tilde{\boldsymbol{\delta}})=\left(1-n_iN_i^{-1}\right)^2\left(\bar{\mathbf{x}}_{ri}^T\;\;\bar{\mathbf{z}}_{ri}^T\right)\hat{V}_{\mathbf{u}}\left(\tilde{\boldsymbol{\delta}}\right)\left(\bar{\mathbf{x}}_{ri}^T\;\;\bar{\mathbf{z}}_{ri}^T\right)^T$ is due to the estimation of fixed and random effects in the model; and
- $h_{2i}(\tilde{\boldsymbol{\delta}})=\left(1-n_iN_i^{-1}\right)^2\hat{V}_{\mathbf{u}}\left(\bar{e}_{ri}\right)$ can be calculated just using the data from area *i*, i.e. $\hat{V}_{\mathbf{u}}(\bar{e}_{ri})=(N_i-n_i)^{-1}(n_i-1)^{-1}\sum_{j\in s_i}\left(y_j-\mathbf{x}_j^T\tilde{\boldsymbol{\beta}}^{\psi}-\mathbf{z}_j^T\tilde{\mathbf{u}}^{\psi}\right)^2$, or by pooling data from the entire sample, in

which case $\hat{V}_{\mathbf{u}}(\bar{e}_{ri}) = (N_i - n_i)^{-1}(n_i - 1)^{-1}\sum_h \sum_{j \in s_h}\left(y_j - \mathbf{x}_j^T \tilde{\boldsymbol{\beta}}^{\psi} - \mathbf{z}_j^T \tilde{\mathbf{u}}^{\psi}\right)^2$. Note that the pooled estimator leads to more stable MSE estimates when area sample sizes are very small.

Finally, we add an estimator of the squared conditional bias to (23), leading to an estimator of the MSE of RBLUP of the form

$$\widehat{MSE}_{\mathbf{u}}\left(\hat{\bar{y}}_i^{RBLUP}\right) = h_{1i}(\tilde{\boldsymbol{\delta}}) + h_{2i}(\tilde{\boldsymbol{\delta}}) + \left\{\hat{B}_{\mathbf{u}}\left(\hat{\bar{y}}_i^{RBLUP}\right)\right\}^2, \tag{24}$$

where $\hat{B}_{\mathbf{u}}\left(\hat{\bar{y}}_i^{RBLUP}\right)$ is the estimator of the conditional bias defined following (21).

The corresponding estimator of the MSE of REBLUP (13) is obtained by adding an extra term to (24) to account for the increased variability due to the estimation of the variance components. Let $\boldsymbol{\theta} = \left(\sigma_u^{\psi 2}, \sigma_e^{\psi 2}\right)$ denote the vector of the variance components, with estimator $\hat{\boldsymbol{\theta}} = \left(\hat{\sigma}_u^{\psi 2}, \hat{\sigma}_e^{\psi 2}\right)$. Our development is similar to that of Prasad and Rao (1990) in that it is based on the decomposition

$$\begin{aligned} MSE_{\mathbf{u}}\left(\hat{\bar{y}}_i^{REBLUP}\right) &= MSE_{\mathbf{u}}\left(\hat{\bar{y}}_i^{RBLUP}\right) + E_{\mathbf{u}}\left[\left(\hat{\bar{y}}_i^{REBLUP} - \hat{\bar{y}}_i^{RBLUP}\right)^2\right] \\ &\quad + 2E_{\mathbf{u}}\left[\left(\hat{\bar{y}}_i^{RBLUP} - \bar{y}_i\right)\left(\hat{\bar{y}}_i^{REBLUP} - \hat{\bar{y}}_i^{RBLUP}\right)\right] \\ &= MSE_{\mathbf{u}}\left(\hat{\bar{y}}_i^{RBLUP}\right) + E_{\mathbf{u}}\left[\left(\hat{\bar{y}}_i^{REBLUP} - \hat{\bar{y}}_i^{RBLUP}\right)^2\right] + O(m^{-1}). \end{aligned} \tag{25}$$

Details of the proof that the cross-product term above is of lower order are available from the authors. An approximation to the second term on the right hand side of (25) can be obtained using Taylor series methods under conditions (RC1) - (RC7) set out in the Appendix. In order to develop this approximation, we first note that, using (RC7), we can write

$$\hat{\bar{y}}_i^{REBLUP} - \hat{\bar{y}}_i^{RBLUP} = \partial_{\boldsymbol{\theta}} \hat{\bar{y}}_i^{RBLUP}\left\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\} + O_p\left(m^{-1}\right).$$

Next, using the identity

$$\hat{\bar{y}}_i^{RBLUP} = \frac{1}{N_i}\sum_{j \in r_i} \mathbf{x}_i^T \boldsymbol{\beta}_0^{\psi} + \frac{1}{N_i}\sum_{j \in r_i} \mathbf{z}_j^T \mathbf{B}_s\left(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta}_0^{\psi}\right) + \left\{\frac{1}{N_i}\sum_{j \in r_i} \mathbf{x}_j^T - \frac{1}{N_i}\sum_{j \in r_i} \mathbf{z}_j^T \mathbf{B}_s \mathbf{X}_s\right\}\left(\tilde{\boldsymbol{\beta}}^{\psi} - \boldsymbol{\beta}_0^{\psi}\right),$$

and the fact that the derivatives of $\tilde{\boldsymbol{\beta}}^{\psi}$ with respect to $\boldsymbol{\theta}$ are of lower order, we can write

$$\hat{\bar{y}}_i^{REBLUP} - \hat{\bar{y}}_i^{RBLUP} = \left(N_i^{-1}\sum_{j \in r_i} \mathbf{z}_j^T\right)\sum_{k=1}^{2}\left(\partial_{\theta_k} \mathbf{B}_s\right)\left(\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta}_0^{\psi}\right)\left\{\hat{\theta}_k - \theta_k\right\} + O_p\left(m^{-1}\right).$$

Finally, using conditions (RC2)-(RC6) and noting that $\tilde{\boldsymbol{\beta}}^{\psi} - \boldsymbol{\beta}_0^{\psi} = O_p\left(m^{-1/2}\right)$ we get

$$Var_{\mathbf{u}}\left(\hat{\bar{y}}_i^{REBLUP}-\hat{\bar{y}}_i^{RBLUP}\right)=\left(N_i^{-1}\sum_{j\in r_i}\mathbf{z}_j^T\right)\left\{\sum_{k=1}^{2}\sum_{g=1}^{2}\left(\partial_{\theta_k}\mathbf{B}_s\right)Cov_{\mathbf{u}}\begin{bmatrix}\left(\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}_0^{\psi}\right)\left\{\hat{\theta}_k-\theta_k\right\},\\ \left(\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}_0^{\psi}\right)\left\{\hat{\theta}_g-\theta_g\right\}\end{bmatrix}\left(\partial_{\theta_g}\mathbf{B}_s\right)^T\right\}\left(N_i^{-1}\sum_{j\in r_i}\mathbf{z}_j^T\right)^T+o\left(m^{-1}\right)$$

where

$$Cov_{\mathbf{u}}\left\{\left(y_j-\mathbf{x}_j^T\boldsymbol{\beta}_0^{\psi}\right)\left(\hat{\theta}_k-\theta_k\right),\left(y_l-\mathbf{x}_l^T\boldsymbol{\beta}_0^{\psi}\right)\left(\hat{\theta}_g-\theta_g\right)\right\}=\left\{\begin{matrix}\left(\mathbf{z}_j^T\mathbf{u}_0^{\psi}\right)\left(\mathbf{z}_l^T\mathbf{u}_0^{\psi}\right)\\ +\sigma_e^{\psi 2}\mathbf{I}(j=l)\end{matrix}\right\}E_{\mathbf{u}}\left(\hat{\theta}_k-\theta_k\right)\left(\hat{\theta}_g-\theta_g\right)+O(m^{-1}).$$

Consequently, (25) can be approximated by:

$$\widehat{MSE}\left(\hat{\bar{y}}_i^{REBLUP}\right)=h_{1i}(\tilde{\boldsymbol{\delta}})+h_{2i}(\tilde{\boldsymbol{\delta}})+h_{3i}(\tilde{\boldsymbol{\delta}})+\left\{\hat{B}\left(\hat{\bar{y}}_i^{RBLUP}\right)\right\}^2. \tag{26}$$

where

$$h_{3i}(\tilde{\boldsymbol{\delta}})=\left(N_i^{-1}\sum_{j\in r_i}\mathbf{z}_j^T\right)\Upsilon Var_{\mathbf{u}}(\hat{\boldsymbol{\theta}})\left(N_i^{-1}\sum_{j\in r_i}\mathbf{z}_j^T\right)^T$$

and

$$\Upsilon=\sum_{k=1}^{2}\sum_{g=1}^{2}\left\{\left(\partial_{\theta_k}\mathbf{B}_s\right)\left[\sum_j\sum_l\left\{\left(\mathbf{z}_j^T\mathbf{u}_0^{\psi}\right)\left(\mathbf{z}_l^T\mathbf{u}_0^{\psi}\right)+\sigma_e^{\psi 2}\mathbf{I}(j=l)\right\}\right]\left(\partial_{\theta_g}\mathbf{B}_s\right)^T\right\}.$$

An estimate of the variance-covariance matrix of the variance components $Var_{\mathbf{u}}(\hat{\boldsymbol{\theta}})$ can be calculated using the results of Sinha and Rao (2009). An estimator of the conditional MSE of REBLUP is then obtained by replacing $\boldsymbol{\delta}_0=\left(\boldsymbol{\beta}_0^{\psi T},\mathbf{u}_0^{\psi T}\right)^T$ and $\tilde{\boldsymbol{\delta}}=\left(\tilde{\boldsymbol{\beta}}^{\psi T},\tilde{\mathbf{u}}^{\psi T}\right)^T$ by $\hat{\boldsymbol{\delta}}=\left(\hat{\boldsymbol{\beta}}^{\psi T},\hat{\mathbf{u}}^{\psi T}\right)^T$ in (26) and leads to:

$$mse\left(\hat{\bar{y}}_i^{REBLUP}\right)=h_{1i}(\hat{\boldsymbol{\delta}})+h_{2i}(\hat{\boldsymbol{\delta}})+h_{3i}(\hat{\boldsymbol{\delta}})+\left\{\hat{B}\left(\hat{\bar{y}}_i^{REBLUP}\right)\right\}^2. \tag{27}$$

Note that a corresponding estimate of the conditional MSE of the EBLUP is easily calculated by setting the tuning constant for the influence function in (27) so that no outlier modification occurs, e.g. setting $c > 100$.

We take a similar approach to defining an estimator of the conditional MSE of REBLUP-BC. To start, we develop an approximation to the conditional prediction variance of this predictor when the variance components are known, i.e. for RBLUP-BC. In this case the prediction error is

$$\hat{\bar{y}}_i^{RBLUP-BC}-\bar{y}_i=\left(1-n_iN_i^{-1}\right)\left\{\left(\bar{\mathbf{x}}_{ri}^T\tilde{\boldsymbol{\beta}}^{\psi}+\bar{\mathbf{z}}_{ri}^T\tilde{\mathbf{u}}^{\psi}\right)+n_i^{-1}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_j-\mathbf{x}_j^T\tilde{\boldsymbol{\beta}}^{\psi}-\mathbf{z}_j^T\tilde{\mathbf{u}}^{\psi}}{\omega_{ij}^{\psi}}\right)-\bar{y}_{ri}\right\}.$$

The second (BC) term inside the braces on the right hand side of this expression can be expanded using a Taylor series approximation. When the tuning constant used in ϕ is large, so $\phi'\approx 1$, this approximation becomes

$$n_i^{-1}\sum\nolimits_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_j-\mathbf{x}_j^T\tilde{\boldsymbol{\beta}}^{\psi}-\mathbf{z}_j^T\tilde{\mathbf{u}}^{\psi}}{\omega_{ij}^{\psi}}\right)=n_i^{-1}\sum\nolimits_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_j-\mathbf{x}_j^T\boldsymbol{\beta}_0^{\psi}-\mathbf{z}_j^T\mathbf{u}_0^{\psi}}{\omega_{ij}^{\psi}}\right)-\begin{pmatrix}\tilde{\boldsymbol{\beta}}^{\psi}-\boldsymbol{\beta}_0^{\psi}\\ \tilde{\mathbf{u}}^{\psi}-\mathbf{u}_0^{\psi}\end{pmatrix}^T\begin{pmatrix}\overline{\mathbf{x}}_{si}\\ \overline{\mathbf{z}}_{si}\end{pmatrix}+O_p\left(m^{-1}\right).$$

Substituting in the preceding expression for the prediction error of RBLUP-BC leads to

$$\hat{\bar{y}}_i^{RBLUP-BC}-\bar{y}_i=\left(1-n_iN_i^{-1}\right)\left(T_i-\bar{e}_{ri}+U_i\right)+O_p\left(m^{-1}\right) \tag{28}$$

where $T_i=\left(\overline{\mathbf{x}}_{ri}-\overline{\mathbf{x}}_{si}\right)^T\left(\tilde{\boldsymbol{\beta}}^{\psi}-\boldsymbol{\beta}_0^{\psi}\right)+\left(\overline{\mathbf{z}}_{ri}-\overline{\mathbf{z}}_{si}\right)^T\left(\tilde{\mathbf{u}}^{\psi}-\mathbf{u}_0^{\psi}\right)$ and $U_i=n_i^{-1}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left\{\left(y_j-\mathbf{x}_j^T\boldsymbol{\beta}_0^{\psi}-\mathbf{z}_j^T\mathbf{u}_0^{\psi}\right)/\omega_{ij}^{\psi}\right\}$.

Under the regularity conditions (RC1) – (RC5), the variance of T_i is $O\left(m^{-1}\right)$ and the covariance between T_i and U_i is of a lower order of magnitude than either of their variances, so from (28) we can write down an estimator of the conditional variance of RBLUP-BC of the form

$$\widehat{MSE}\left(\hat{\bar{y}}_i^{RBLUP-BC}\right)=h_{1i}^{BC}(\tilde{\boldsymbol{\delta}})+h_{2i}(\tilde{\boldsymbol{\delta}})+h_{3i}^{BC}(\tilde{\boldsymbol{\delta}}), \tag{29}$$

where

$$h_{1i}^{BC}(\tilde{\boldsymbol{\delta}})=\left(1-n_iN_i^{-1}\right)^2\left\{\begin{pmatrix}\overline{\mathbf{x}}_{ri}-\overline{\mathbf{x}}_{si}\\ \overline{\mathbf{z}}_{ri}-\overline{\mathbf{z}}_{si}\end{pmatrix}^T\hat{V}_{\mathbf{u}}\left(\tilde{\boldsymbol{\delta}}\right)\begin{pmatrix}\overline{\mathbf{x}}_{ri}-\overline{\mathbf{x}}_{si}\\ \overline{\mathbf{z}}_{ri}-\overline{\mathbf{z}}_{si}\end{pmatrix}\right\}$$

and

$$h_{3i}^{BC}(\tilde{\boldsymbol{\delta}})=\left(1-n_iN_i^{-1}\right)^2n_i^{-1}(n_i-p-q)^{-1}\sum_{j\in s_i}\left\{\omega_{ij}^{\psi}\phi\left(\frac{y_j-\mathbf{x}_j^T\tilde{\boldsymbol{\beta}}^{\psi}-\mathbf{z}_j^T\tilde{\mathbf{u}}^{\psi}}{\omega_{ij}^{\psi}}\right)\right\}^2.$$

The estimator of the conditional MSE of REBLUP-BC is then obtained by adding a term to (29) to account for the additional uncertainty due to estimation of the variance components. The same approach as already used for REBLUP can be applied, leading to the approximation

$$E_{\mathbf{u}}\left[\left(\hat{\bar{y}}_i^{REBLUP-BC}-\hat{\bar{y}}_i^{RBLUP-BC}\right)^2\right]=\left(1-n_iN_i^{-1}\right)^2\mathbf{D}_i^T\Upsilon Var_{\mathbf{u}}(\hat{\boldsymbol{\theta}})\mathbf{D}_i+O\left(m^{-1}\right)=h_{4i}^{BC}(\boldsymbol{\delta}_0)+O\left(m^{-1}\right) \tag{30}$$

where $\mathbf{D}_i=\overline{\mathbf{z}}_{ri}-n_i^{-1}\sum_{j\in s_i}\phi'\left[\left\{y_j-\mathbf{x}_j^T\boldsymbol{\beta}_0^{\psi}-\mathbf{z}_j^T\mathbf{B}_s\left(\mathbf{y}_s-\mathbf{X}_s\boldsymbol{\beta}_0^{\psi}\right)\right\}/\omega_{ij}^{\psi}\right]\mathbf{z}_j$. Note that $\mathbf{D}_i=\mathbf{0}$ when ϕ is the identity function, e.g. as in the version of BC described in Chambers *et al*. (1993), and the model only contains random intercepts. An estimator of the MSE of REBLUP-BC is then defined by replacing $\boldsymbol{\delta}_0=\left(\boldsymbol{\beta}_0^{\psi T},\mathbf{u}_0^{\psi T}\right)^T$ and $\tilde{\boldsymbol{\delta}}=\left(\tilde{\boldsymbol{\beta}}^{\psi T},\tilde{\mathbf{u}}^{\psi T}\right)^T$ by $\hat{\boldsymbol{\delta}}=\left(\hat{\boldsymbol{\beta}}^{\psi T},\hat{\mathbf{u}}^{\psi T}\right)^T$ in (29) and (30), to give

$$mse\left(\hat{\bar{y}}_i^{REBLUP-BC}\right) = h_{1i}^{BC}(\hat{\boldsymbol{\delta}}) + h_{2i}(\hat{\boldsymbol{\delta}}) + h_{3i}^{BC}(\hat{\boldsymbol{\delta}}) + h_{4i}^{BC}(\hat{\boldsymbol{\delta}}). \quad (31)$$

As with the estimator of the conditional MSE of REBLUP-BC based on the pseudo-linearization approach, no squared conditional bias estimator is used with (31) since the REBLUP-BC predictor is approximately unbiased for the small area mean. However, unlike (27), the MSE of REBLUP-BC has an extra term $h_{3i}^{BC}(\hat{\boldsymbol{\delta}})$ that arises due to the conditional bias correction in REBLUP-BC (19). Note that the term $h_{4i}^{BC}(\hat{\boldsymbol{\delta}})$ in (31) is equivalent to $h_{3i}(\hat{\boldsymbol{\delta}})$ in (27), i.e., they both estimate the increase in variability due to the estimation of the variance components. Estimators of the conditional MSEs of MQ (15) and MQ-BC (18) can be obtained similarly, with this development set out in the Appendix.

5. Model-Based Simulations

We provide model-based simulation results illustrating the performances of the different outlier robust small area predictors and of the corresponding MSE estimators described in Sections 3 and 4. Population data are generated for $m = 40$ small areas, with samples selected by simple random sampling without replacement within each area. Population and sample sizes are the same for all areas, and are fixed at either $N_i = 100$, $n_i = 5$ or $N_i = 300$, $n_i = 15$. Values for x are generated as independently and identically distributed from a lognormal distribution with a mean of 1.0 and a standard deviation of 0.5 on the log scale. Values for Y are generated as $y_{ij} = 100 + 5x_{ij} + u_i + \varepsilon_{ij}$, where the random area and individual effects are independently generated according to four scenarios:

- [0,0] – No outliers: $u \sim N(0,3)$ and $\varepsilon \sim N(0,6)$.
- [e,0] – Individual outliers only: $u \sim N(0,3)$ and $\varepsilon \sim \delta N(0,6) + (1-\delta)N(20,150)$, where δ is an independently generated Bernoulli random variable with $\Pr(\delta = 1) = 0.97$, i.e. the individual effects are independent draws from a mixture of two normal distributions, with 97% on average drawn from a 'well-behaved' $N(0,6)$ distribution and 3% on average drawn from an outlier $N(20,150)$ distribution.
- [0,u] – Area outliers only: $u \sim N(0,3)$ for areas 1-36, $u \sim N(9,20)$ for areas 37-40 and $\varepsilon \sim N(0,6)$,

i.e. random effects for areas 1–36 are drawn from a 'well behaved' $N(0,3)$ distribution, with those for areas 37–40 drawn from an outlier $N(9,20)$ distribution. Individual effects are not outlier-contaminated.

- [e,u] – Outliers in both area and individual effects: $u \sim N(0,3)$ for areas 1-36, $u \sim N(9,20)$ for areas 37-40 and $\varepsilon \sim \delta N(0,6)+(1-\delta)N(20,150)$.

Each scenario is independently simulated 500 times. For each simulation the population values are generated according to the underlying scenario, a sample is selected in each area and the sample data are then used to compute estimates of each of the actual area means for *y*.

Five different estimators are used for this purpose - the standard EBLUP, see (2), which serves as a reference; the projective M-quantile estimator MQ, see (15); the robust bias-corrected predictive MQ estimator MQ-BC, see (18); the robust projective REBLUP estimator of Sinha and Rao (2009), see (13); and its robust bias-corrected version REBLUP-BC, see (19). In all cases the 'projective' influence function ψ is a Huber Proposal 2 type with tuning constant $c = 1.345$. In contrast, the 'predictive', less restrictive, influence function ϕ used in MQ-BC and REBLUP-BC is also a Huber Proposal 2 type, but with a larger tuning constant, $c = 3$.

The performance of these estimators across the different simulations is assessed by computing the median values of their area specific relative bias and relative root mean squared error, where the relative bias of an estimator $\hat{\bar{y}}_i$ for the actual mean $\bar{y}_i$ of area i is the average across simulations of the errors $\hat{\bar{y}}_i - \bar{y}_i$ divided by the corresponding average value of $\bar{y}_i$, and its relative root mean squared error is the square root of the average across simulations of the squares of these errors, again divided by the average value of $\bar{y}_i$. Table 1 presents these median values for the different simulation scenarios and different estimators.

The relative bias results set out in Table 1 confirm our expectations regarding the behaviour of the projective estimators (EBLUP, REBLUP and MQ) and the predictive estimators (REBLUP-BC and MQ-BC). The former are more biased than the latter (see scenarios with area and individual outliers) as a consequence of their implicit assumption that although outlier variances may be inflated relative to non-outliers, outlier effects still have zero expectation. This increase in bias is most pronounced when there are outliers in the area effects, which is not unexpected since that is when area means are most affected by the presence of outliers in the population data. Turning to the median RRMSE results, we see that claims in the literature (e.g. Chambers and Tzavidis, 2006) about the superior outlier robustness of MQ compared with the EBLUP certainly hold true – provided the outliers are in individual effects. If there are outliers in area effects, then MQ appears to offer no extra protection compared to the EBLUP, and in fact performs worse, mainly due to its sharply increasing bias in this situation. Similarly, when we compare the EBLUP and the REBLUP we see that if outliers are associated with individual effects, then the REBLUP offers better RRMSE performance than the EBLUP. However, the gap between these two estimators narrows considerably when outliers are associated with area effects. In contrast, the two predictive estimators seem relatively robust in terms of RRMSE performance. Nevertheless, due to the increased variability as a consequence of their bias corrections, both predictive estimators are not as efficient as the projective estimators when outliers are associated with individual effects, but both also do not fail when there are outliers in the area effects. Finally, the REBLUP-BC estimator appears to perform better than the MQ-BC estimator for those scenarios where the use of predictive estimators offers gains.

We now examine the performance of the different MSE estimators. We are mainly interested in the performance of MSE estimators for the predictive estimators REBLUP-BC and MQ-BC. However, we also comment on the performance of the different MSE estimators when used for estimating the MSE of projective estimators under a range of scenarios. MSE estimation for REBLUP and REBLUP-BC is implemented via the pseudo-linearization MSE estimator (21) (hereafter CCT) and via the linearization-based MSE estimators (27) and (31) (hereafter CCST). For MQ and MQ-BC the MSE estimators (A6) and (A8) - see the Appendix for details - which correspond to CCST, as well as those which correspond to CCT (see Chambers *et al*. 2011 for details) are used. For REBLUP and REBLUP-BC we investigated the parametric bootstrap procedure of Sinha and Rao (2009), hereafter BOOT, which we implemented by generating 100 bootstrap samples in each Monte Carlo run (using more bootstrap samples did not change our results to any significant extent). Finally, the MSE of the EBLUP was estimated via the Prasad-Rao (1990) estimator, hereafter PR, as well as via CCT and CCST. The results of the MSE estimators for each scenario and for each estimator are shown in Table 2 where we report the median values of their area specific relative biases and their relative root mean squared errors.

As a first general comment we note that for all estimators and scenarios we considered, CCST offers better stability than CCT. We now focus on the performance of these two MSE estimators for estimating the MSE of the robust predictive estimators REBLUP-BC and MQ-BC. As already pointed out, CCST exhibits overall better stability than the CCT. In terms of bias the picture is not as clear cut, however, with the biases of the two MSE estimators of same order of magnitude. In particular, while CCST has a somewhat better bias performance for REBLUP-BC, CCT performs better in terms of bias for MQ-BC. We also see that BOOT is generally a more stable MSE estimator for REBLUP-BC compared to CCT and CCST. However, this is a computationally intensive method of MSE estimation, and so may not always be appropriate in a practical survey setting. In this context we note that semiparametric bootstrap methods have recently been proposed for MQ and the non-robust version of MQ-BC (Tzavidis *et al*., 2010). We have not fully evaluated these different bootstrap methods here because the focus of this paper is on analytic methods of MSE estimation, and also because to do so would have substantially increased the length of this paper.

Turning now to MSE estimation for the projective estimators we observe that CCST for both REBLUP and MQ has lower bias than CCT and is also more stable. In contrast, BOOT is sometimes more biased than CCST but is more stable. However, both CCT and CCST appear to substantially overestimate the MSE of MQ when the population contains both area and unit outliers. Reasons for why this happens are not clear at present. Since this scenario is one where robust predictive estimation is advised, it is perhaps of more interest to note that both CCT and CCST work well in terms of estimating the MSE of MQ-BC, with CCST being the more stable of the two MSE estimators in this situation.

As one would expect, the PR estimator of the MSE of the EBLUP performs well in the [0,0] scenario and also records small relative bias for the [e,0] scenario, i.e. only individual outliers (when both CCT and CCST record large positive biases). However, it records large negative biases for situations where there are area level outliers (when both CCT and CCST record negligible biases). The main strength of PR is its stability - its median RRMSE is consistently lower than that of CCT and it is also more stable than CCST when there are no area level outliers. In large part, this is due to the small sample instability caused by the squared conditional bias term used in both CCT and CCST. This issue was noted by Chambers *et al*. (2011), who pointed out that the bias robustness of CCT for the EBLUP that is evident in Table 2 comes at the price of higher variability, especially in the case of very small area sample sizes. In comparison, CCST for the EBLUP demonstrates very similar bias robustness to CCT and is more stable.

6. Design-Based Simulation

Design-based simulations complement model-based simulations for SAE since they allow us to evaluate the performance of SAE methods in the context of a real population and realistic sampling methods where we do not know the precise source of contamination. From a finite population perspective we believe that this type of simulation constitutes a more practical and appropriate representation of the SAE problem. Furthermore, it provides a good illustration of why a focus on conditional MSE is likely to be closer to the MSE of interest for analysts using small area methods.

The population underpinning the design-based simulation is based on a data set obtained under the Environmental Monitoring and Assessment Program (EMAP) of the U.S. Environmental Protection Agency. The background to this data set is that between 1991 and 1995 EMAP conducted a survey of lakes in the North-Eastern states of the U.S. The data collected in this survey consists of 551 measurements from a sample of 334 of the 21,026 lakes located in this area. The lakes making up this population are grouped into 113 8-digit Hydrologic Unit Codes (HUCs), of which 64 contained less than 5 observations and 27 did not have any observations. In our simulation, we defined HUCs as the small areas of interest, with lakes grouped within HUCs. The variable of interest is Acid Neutralizing Capacity (ANC), an indicator of the acidification risk of water bodies. In addition to ANC values for the sampled locations, the EMAP data set also contained the elevation of each lake in the target area. In this simulation, elevation is used as the only model covariate.

A synthetic population of 21,026 ANC lake-specific individual values was constructed following the same procedure as in Salvati *et al.* (2012). This corresponds to non-parametrically simulating a population of ANC values for all 21,026 lakes using a nearest-neighbour imputation algorithm that retains the spatial structure of the observed 334 lake-specific ANC values in the EMAP sample data. This synthetic population of ANC values is then kept fixed over the Monte-Carlo simulations. Details on the exact data generation mechanism and the characteristics of the population can be found in Salvati *et al.* (2012). A total of 1000 independent random samples of lakes were taken from the population of 21,026 lakes by randomly selecting lakes in the

86 HUCs that contained EMAP sampled lakes, with sample sizes in these HUCs set to the greater of five and the original EMAP sample size. A two-level (level 1 is the lake and level 2 is the HUC) mixed model was fitted to the synthetic population data. The Shapiro-Wilk normality test rejects the null hypothesis that the residuals follow a normal distribution, with p-values of 0.0356 (level 1), and <0.0001 (level 2), indicating that the Gaussian assumptions of the mixed model are not met. Using a model that relaxes these assumptions, e.g. an M-quantile model with a bounded influence function, therefore seems reasonable for these data.

Table 3 shows the median relative biases and the median RRMSEs of the different predictors (EBLUP, REBLUP, MQ, REBLUP-BC, MQ-BC) and Table 4 reports the median relative biases and the median RRMSEs of the corresponding estimators of the MSEs of these predictors. The robust predictive estimators MQ-BC and REBLUP-BC work well both in terms of bias and RRMSE, while the EBLUP and MQ have the highest RRMSE, with MQ also recording the largest negative bias. The REBLUP shows a good performance in terms of RRMSE but records a large negative bias. These results suggest that predictive estimators offer the most balanced performance both in terms of bias and MSE for this population.

We now examine the performance of the different methods of MSE estimation. To start, Table 4 indicates that on average across areas CCST performs better or comparably to CCT for all predictors with these data. It also shows that the performance of the parametric bootstrap BOOT of Sinha and Rao (2009) depends on the predictor. We see that BOOT performs similarly to CCST for REBLUP-BC but exhibits substantial bias for REBLUP. Finally, we observe that for the EBLUP, MSE estimation via PR is essentially no different to that via CCT and CCST, with the CCST being more stable.

The analysis in Table 4 focuses on median RMSE estimation performance across areas. This hides the large variability in MSE estimation performance between areas. The relationship between the 'true' (empirical) RMSE of each predictor and its estimators for each area is shown in Figure 1, where box plots illustrating the variability in the RMSE Ratio, defined as the ratio of the average estimated RMSE for each area to the true RMSE, are shown for each predictor and each MSE estimation method. Here we see that both PR and CCST behave rather similarly for the EBLUP and do not adequately capture the between area differences in the area-specific MSE of this predictor. In contrast, CCT tracks this area-specific empirical MSE very well. These results are consistent with the comments in both Longford (2007) and Chambers *et al.*

(2011) that PR should not be used if area-specific estimation of MSE is a requirement. The main difference between CCT and CCST is that CCT essentially only assumes a linear mean structure for the data, while CCST, like PR, assumes that the sample data follow a linear mixed model. The fact that CCST does not perform as well as CCT in terms of tracking the area-specific MSE of the EBLUP in our simulations is therefore some evidence for recommending that CCST not be used with a non-robust predictor like the EBLUP. Turning to MSE estimation for REBLUP and REBLUP-BC, we see that CCST improves somewhat for REBLUP, which is consistent with this estimator being somewhat more robust than the EBLUP, and performs very well for the robust bias corrected REBLUP-BC, while CCT, though still an efficient 'tracker' of the area-specific MSEs of these two predictors, also exhibits a small downward bias. In contrast, BOOT does not track the area-specific MSEs of REBLUP and REBLUP-BC. Finally, we note that CCT and CCST behave very similarly for MQ and MQ-BC. Both track the area-specific MSEs of these predictors quite well. Overall, it appears that for the EMAP population data that were used in our simulations, CCST is the method of choice for area-specific MSE estimation for the robust predictive estimators REBLUP-BC and MQ-BC.

7. Final Remarks

In this paper we explore the extension of the Robust Predictive approach to SAE and we propose two analytic linearization-based mean squared error (MSE) estimators for outlier robust predictors of small area means. The first is a bias-robust MSE estimator that is based on the 'pseudo-linearization' approach of Chambers *et al*. (2011). The second method is based on first order approximations to the variances of solutions of robust estimating equations.

The empirical results reported in Sections 5 and 6 show that the robust predictive estimators (REBLUP-BC and MQ-BC) are less biased and can be more efficient than the robust projective estimators (REBLUP and MQ) in the presence of area and individual outliers. What is also evident from these results is that the bias correction of the robust predictive estimators comes at the cost of higher variability. As a result we expect that the use of these estimators will pay dividends when model diagnostics suggest that there are significant departures from the assumed working small area model. One approach for controlling the bias-variance trade off when using the robust predictive approach is by selecting optimal tuning constants c and

ϕ to be used in these estimators. In general, c is set equal to 1.345 for the ψ influence function and is set to a larger value for ϕ influence function. Applications with real data show that setting c=2 or c=3 for ϕ provides a good balance between bias and variance. An 'optimal' c value can be potentially achieved by cross-validation, and is an avenue for future research.

The pseudo-linearization MSE estimator CCT and the linearization-based MSE estimator CCST offer a promising approach to analytic estimation of the MSE of robust predictive estimators. As has already been noted by Chambers *et al.* (2011), CCT represents a method of MSE estimation that is well suited to tracking area-specific variability in MSE. However, this comes at the price of increased instability. Although we do not explore this issue fully in this paper, it seems clear that CCST, when used in conjunction with robust predictive estimation methods, also tracks area-specific variability in MSE and is more stable than CCT. This opens up the possibility that CCST may be competitive with more numerically intensive bootstrap methods for MSE estimation. A more complete comparison of this MSE estimation method with alternative parametric and semiparametric bootstrap methods of MSE estimation is beyond the scope of this paper, however, and is left for further research. Finally, we note that although CCST was developed under a conditional version of the linear mixed model, it should be possible to develop an unconditional version of CCST that averages over the distribution of the random area effects under a linear mixed model, and so reduces to the widely used MSE estimator PR in the case of the EBLUP. This presents an additional avenue for further research.

Acknowledgements

Work partially supported by the SAMPLE project (SSH-CT-2007-217565; FP7-SSH-2007-1) and by ARC Linkage Grant LP0776810. The authors are grateful to Space-Time Aquatic Resources Modeling and Analysis Program (STARMAP) for data availability. They are also indebted to the Associate Editor and two anonymous referees for positive comments that significantly improved the development in the paper.

Appendix

1. Regularity Conditions

The following regularity conditions are required for the development of the linearization-based MSE estimator set out in Section 4.2, and uses the same notation as employed there.

(RC1) The influence function ψ is a bounded continuous function with a derivative which, except for a finite number of points, is defined everywhere and is also bounded;

(RC2) The elements of $\mathbf{X}_s$ and $\mathbf{Z}_s$ are uniformly bounded as $m \to \infty$, so that $\mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{U}_s \mathbf{V}_s^{-1} \mathbf{X}_s = \left[O(m)\right]_{p \times p}$, $\mathbf{Z}_s^T \mathbf{\Sigma}_{es}^{-1} \mathbf{Z}_s = \left[O(m)\right]_{q \times q}$ and $\mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{U}_s^{1/2} \mathbf{\Sigma}_{es}^{-1/2} \mathbf{Z}_s = \left[O(m)\right]_{p \times q}$ are uniformly bounded, where $\mathbf{V}_s = \mathbf{\Sigma}_{es} + \mathbf{Z}_s \mathbf{\Sigma}_u \mathbf{Z}_s^T$ with $\mathbf{U}_s = diag\left(\mathbf{V}_s\right)$;

(RC3) The covariance matrices $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_{es}$ have linear structure (Prasad and Rao, 1990) and are known positive definite matrices of order $mq \times mq$ and $n \times n$ respectively, with elements that are also uniformly bounded as $m \to \infty$;

(RC4) The dimension q of the area random effect is a fixed finite number with $\sup_{i \ge 1} n_i = \lambda_1 < \infty$;

(RC5) There exist constants $\varsigma > 0$ and $L < \infty$ such that if $r_j = U_j^{-1/2}\left(y_j - \mathbf{x}_j^T \boldsymbol{\beta}_0^{\psi}\right)$, $t_j = \left(\sigma_e^{\psi}\right)^{-1/2}\left(y_j - \mathbf{x}_j^T \boldsymbol{\beta}_0^{\psi} - \mathbf{z}_j^T \mathbf{u}_0^{\psi}\right)$ and $d_i = \left(\sigma_u^{2\psi}\right)^{-1/2} u_{0i}^{\psi}$, then $E_{\mathbf{u}}\left|\psi\left(r_j\right)\right|^{4+\varsigma}$, $E_{\mathbf{u}}\left\|\psi'\left(r_j\right)\right\|$, $E_{\mathbf{u}}\left|\psi\left(t_j\right)\right|^{4+\varsigma}$, $E_{\mathbf{u}}\left\|\psi'\left(t_j\right)\right\|$, $E_{\mathbf{u}}\left|\psi\left(d_i\right)\right|^{4+\varsigma}$ and $E_{\mathbf{u}}\left\|\psi'\left(d_i\right)\right\|$ are all bounded by L.

(RC6) $\partial_{\theta_k} \mathbf{X}_s^T \mathbf{B}_s = \left[O(1)\right]_{p \times m}$ for $k = 1, \ldots, K$, where $\mathbf{B}_s$ was defined following (20).

(RC7) The elements of $\mathbf{V}$ and $\mathbf{U}$ are differentiable with respect to the variance components, with $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p\left(m^{-1/2}\right)$ and $\partial_{\boldsymbol{\theta}}\left(\partial_{\boldsymbol{\theta}} \hat{\bar{y}}_i^{REBLUP}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = O_p(1)$ when $\left\|\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right\| \le \left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|$.

2. Linearization-based MSE estimation for MQ and MQ-BC

In what follows we assume that for every area *i* there exists a value $q_0(i)$ such that the 'true' value of the regression vector for area *i* is $\boldsymbol{\beta}_{q_0(i)}$. We also use a subscript of 0 to 'true' values under this area-specific model. Then, denoting the M-quantile coefficient for area *i* by $q(i)$, the prediction variance of the MQ (15) is

$$Var\left(\hat{\bar{y}}_i^{MQ} - \bar{y}_i \,|q(i)\right) = \left(1 - n_i N_i^{-1}\right)^2 \left\{\bar{\mathbf{x}}_{ri}^T Var_0(\hat{\boldsymbol{\beta}}_{q(i)})\bar{\mathbf{x}}_{ri}\right\} + \left(1 - n_i N_i^{-1}\right)^2 Var_0(\bar{e}_{ri}). \tag{A1}$$

A first order approximation to $Var_0(\hat{\boldsymbol{\beta}}_{q(i)})$ is

$$Var_0\left(\hat{\boldsymbol{\beta}}_{q(i)}\right) = \left\{E_0\left(\partial_{\beta_{q(i)}}\mathbf{H}_0\right)\right\}^{-1} Var_0\left\{\mathbf{H}\left(\boldsymbol{\beta}_{q_0(i)}\right)\right\}\left[\left\{E_0\left(\partial_{\beta_{q(i)}}\mathbf{H}_0\right)\right\}^{-1}\right]^T + o(n^{-1}) \tag{A2}$$

with $\mathbf{H}(\boldsymbol{\beta}_{q_0(i)}) = \sum_{j=1}^{n} \mathbf{x}_j \psi_q(r_{j0i}) = \mathbf{X}_s^T \psi_q(r_{0i})$, where ψ_q is a bounded M-quantile influence function of order *q*, $\psi_q(r_{0i})$ is the *n*-vector with elements $\psi_q(r_{j0i}) = \psi_q\left\{\omega_{j0i}^{-1}\left(y_j - \mathbf{x}_j^T\boldsymbol{\beta}_{q_o(i)}\right)\right\}$ and ω_{j0i} is a robust estimator of the scale of the residual $r_{j0i} = y_j - \mathbf{x}_j^T\boldsymbol{\beta}_{q_0(i)}$. The $Var_0\left\{\mathbf{H}(\beta_{0q})\right\}$ component of (A2) can be written as

$$Var_0\left\{\mathbf{H}(\beta_{q_0(i)})\right\} = \mathbf{X}_s^T\left\{E_0\left\{\psi_q(r_{0i})\psi_q^T(r_{0i})\right\}\right\}\mathbf{X}_s,$$

since $E_0\left\{\psi_q(r_{j0i})\right\} = 0$. Assuming a Huber-type influence function, we obtain

$$E_0\left(\partial_{\beta_{q(i)}}\mathbf{H}_0\right) = \mathbf{X}_s^T E_0\left[2\frac{d}{d\boldsymbol{\beta}_{q(i)}}\psi_q(r_{0i})\Big|_{\boldsymbol{\beta}_{q(i)}=\boldsymbol{\beta}_{q_0(i)}}\right] = -2\mathbf{X}_s^T\mathbf{C}\mathbf{X}_s$$

where $\mathbf{C}$ is a $n \times n$ diagonal matrix with *j*-th diagonal component

$$\omega_{j0i}^{-1}E_0\left\{qI\left(0 < r_{j0i} \le c\right) + (1-q)I\left(-c < r_{j0i} \le 0\right)\right\}.$$

These expressions lead to the following estimator of (A2):

$$\hat{V}(\hat{\boldsymbol{\beta}}_{q(i)}) = (n-p)^{-1}\left\{\sum_{j\in s}\psi_q^2\left(\hat{r}_{jq(i)}\right)\right\}\left[n^{-1}\sum_{j\in s}\psi_q'\left(\hat{r}_{jq(i)}\right)\right]^{-2}\left(\mathbf{X}_s^T\mathbf{X}_s\right)^{-1} \tag{A3}$$

where $\hat{r}_{jq(i)} = \hat{\omega}_{jq(i)}^{-1}\left(y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{q(i)}\right)$. When $q(i) = 0.5$, (A3) is the estimator proposed by Street *et al*. (1988).

An estimator of the first order approximation (A1) is then

$$\hat{V}\left(\hat{\bar{y}}_i^{MQ}\right) = \left(1 - n_i N_i^{-1}\right)^2 \left\{ \bar{\mathbf{x}}_{ri}^T \hat{V}\left(\hat{\boldsymbol{\beta}}_{\bar{q}_i}\right) \bar{\mathbf{x}}_{ri} \right\} + \left(1 - n_i N_i^{-1}\right)^2 \hat{V}\left(\bar{e}_{ri}\right) \tag{A4}$$

where $\hat{V}(\bar{e}_{ri}) = (N_i - n_i)^{-1}(n-1)^{-1}\sum_h \sum_{j\in s_h}\left(y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_h}\right)^2$. A corresponding estimator of the area-specific bias of MQ is

$$\hat{B}\left(\hat{\bar{y}}_i^{MQ}\right) = N_i^{-1}\left\{ \sum_k \sum_{j\in s_k} w_{ij}\mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_k} - \sum_{j\in i}\mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i} \right\} \tag{A5}$$

where $w_{ij} = b_{ij} + N_i n_i^{-1} I(j \in i)$ and $\mathbf{b}_i = \left(b_{ij}\right) = \mathbf{W}(\bar{q}_i)\mathbf{X}_s\left(\mathbf{X}_s^T\mathbf{W}(\bar{q}_i)\mathbf{X}_s\right)^{-1}\left(N_i - n_i\right)\left(\bar{\mathbf{x}}_{ri} - \bar{\mathbf{x}}_{si}\right)$. The final expression for the estimator of the MSE of MQ is just the sum of (A4) and the square of (A5):

$$mse\left(\hat{\bar{y}}_i^{MQ}\right) = \hat{V}\left(\hat{\bar{y}}_i^{MQ}\right) + \left\{\hat{B}\left(\hat{\bar{y}}_i^{MQ}\right)\right\}^2. \tag{A6}$$

In order to develop a corresponding MSE estimator for MQ-BC, we first note that its prediction error is

$$\hat{\bar{y}}_i^{MQ-BC} - \bar{y}_i = N_i^{-1}\sum_{j\in r_i}\mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i} - N_i^{-1}\sum_{j\in r_i} y_j + \frac{N_i - n_i}{N_i n_i}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i}}{\omega_{ij}^{\psi}}\right)$$

where the right-most (BC) term in this error can be approximated by

$$\frac{1}{n_i}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i}}{\omega_{ij}^{\psi}}\right) \approx \frac{1}{n_i}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_j - \mathbf{x}_j^T\boldsymbol{\beta}_{q_0(i)}}{\omega_{ij}^{\psi}}\right) + \left(\hat{\boldsymbol{\beta}}_{\bar{q}_i} - \boldsymbol{\beta}_{q_0(i)}\right)^T \partial_{\beta_{q_0(i)}}\left\{\frac{1}{n_i}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}_{q_0(i)}}{\omega_{ij}^{\psi}}\right)\right\}.$$

Under the condition that the tuning constant used in ϕ is large, so $\phi' \approx 1$, we then have the corresponding approximation

$$\hat{\bar{y}}_i^{MQ-BC} - \bar{y}_i \approx N_i^{-1}\sum_{j\in r_i}\mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i} - N_i^{-1}\sum_{j\in r_i} y_j + \frac{N_i - n_i}{N_i n_i}\sum_{j\in s_i}\omega_{ij}^{\psi}\phi\left(\frac{y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{q_0(i)}}{\omega_{ij}^{\psi}}\right). \tag{A7}$$

The covariance between the first and third terms on the right hand side of (A7) will be of a lower order than either of their variances, so a first order approximation to the prediction variance of MQ-BC is

$$Var_0\left(\hat{\bar{y}}_i^{MQ-BC} - \bar{y}_i\right) = \left(1 - n_i N_i^{-1}\right)^2\left[\left\{\bar{\mathbf{x}}_{ri} - \bar{\mathbf{x}}_{si}\right\}^T Var\left(\hat{\boldsymbol{\beta}}_{\bar{q}_i}\right)\left\{\bar{\mathbf{x}}_{ri} - \bar{\mathbf{x}}_{si}\right\} + Var\left(\bar{e}_{ri}\right) + \frac{1}{n_i^2}\sum_{j\in s_i}E\left\{\omega_{ij}^{\psi}\phi\left(\frac{y_j - \mathbf{x}_j^T\boldsymbol{\beta}_{q_0(i)}}{\omega_{ij}^{\psi}}\right)\right\}^2\right].$$

The corresponding estimator of this first order approximation to the MSE of MQ-BC is therefore

$$mse(\hat{\bar{y}}_i^{MQ-BC}) = \left(1 - n_i N_i^{-1}\right)^2\left[\left\{\bar{\mathbf{x}}_{ri} - \bar{\mathbf{x}}_{si}\right\}^T \hat{V}\left(\hat{\boldsymbol{\beta}}_{\bar{q}_i}\right)\left\{\bar{\mathbf{x}}_{ri} - \bar{\mathbf{x}}_{si}\right\} + \hat{V}\left(\bar{e}_{ri}\right) + \frac{1}{n_i^2}\sum_{j\in s_i}\left\{\omega_{ij}^{\psi}\phi\left(\frac{y_j - \mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{\bar{q}_i}}{\omega_{ij}^{\psi}}\right)\right\}^2\right]. \tag{A8}$$

Note that, unlike (A6), there is no squared bias term in (A8), since this bias is (approximately) corrected by the BC term of MQ-BC. Also note that both (A6) and (A8) do not allow for variability associated with the 'parameter error' $\bar{q}_i - q_0(i)$, and so could underestimate the MSEs of MQ and MQ-BC under the area-specific model.

References

Battese, G., Harter, R. and Fuller, W. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *J. Am. Statist. Ass.*, **83**, 28-36.

Booth, J.G. and Hobert, J.P. (1998). Standard errors of prediction in Generalized Linear Mixed Models. *J. Am. Statist. Ass.*, **93**, 262-272.

Chambers, R.L. (1986). Outlier robust finite population estimation. *J. Am. Statist. Ass.*, **81**, 1063-1069.

Chambers, R., Chandra, H. and Tzavidis, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Method.*, **37**, 153-170.

Chambers, R.L., Dorfman, A.H. and Wehrly, T.E. (1993). Bias robust estimation in finite populations using nonparametric calibration. *J. Am. Statist. Ass.*, **88**, 268-277.

Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika,* **93**, 255-268.

Chambers, R. and Clark, R. (2012). *An Introduction to Model-Based Survey Sampling with Applications*, Oxford Statistical Science Series, Oxford University Press.

Chandra H., Salvati N. and Chambers R. (2007). Small area estimation for spatially correlated populations - a comparison of direct and indirect model-based methods. *Statist. in Trans.*, **8**, 887-906.

Chandra, H. and Chambers, R. (2009). Multipurpose weighting for small area estimation. *J. Off. Stat.*, **25**, 379-395.

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

Harville, D.A. and Jeske, D.R. (1992). Mean squared error of estimation or prediction under a general linear model. *J. Am. Statist. Ass.*, **87**, 724-731.

Huber, P. (1981). *Robust Statistics*. Wiley, New York.

Longford, N.T. (2007). On standard errors of model-based small-area estimators. *Survey Method.*, **33**, 69-79.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small area estimators. *J. Am. Statist. Ass.*, **85**, 163-171.

Royall, R.M. and Cumberland, W.G. (1978). Variance estimation in finite population sampling. *J. Am. Statist. Ass.*, **73**, 351 - 358.

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

Richardson, A.M. and Welsh, A.H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics,* **51**, 1429-1439.

Salvati, N., Chandra, H., Ranalli, M.G. and Chambers, R. (2010). Small area estimation using a nonparametric model-based direct estimator. *Comput. Statist. Data Anal.*, **54**, 2159-2171.

Salvati, N., Tzavidis, N., Pratesi, M. and Chambers, R. (2012). Small area estimation via M-quantile geographically weighted regression. *Test,* **21**, 1-28.

Sinha, S.K. and Rao, J.N.K. (2009). Robust small area estimation. *Canad. J. Stat.*, **37**, 381-399.

Street, J. O., Carroll, R. J., and Ruppert, D. (1988). A note on computing robust regression estimates via iterative reweighted least squares. *Am. Statist*, **42**, 152–154.

Tzavidis, N., Marchetti, S. and Chambers, R. (2010). Robust prediction of small area means and distributions. *Aust. & NZ J. Stat.*, **52**, 167-186.

Welsh, A.H. and Richardson, A.M. (1997). Approaches to the robust estimation of mixed models. *Handbook of Statistics* (G. S. Maddala and C. R. Rao, eds.), **15**, Chapter 13.

Welsh, A.H. and Ronchetti, E. (1998). Bias-calibrated estimation from sample surveys containing outliers. *J. R. Statist. Soc.* B, **60**, 413-428.

Table 1. Model-based simulation results: performances of predictors of small area means.

Scenario	[0,0]	[e,0]	[0,u]	[0,u]	[e,u]	[e,u]
Areas	1-40	1-40	1-36	37-40	1-36	37-40
Median values of Relative Bias (expressed as a percentage)						
EBLUP	0.02	-0.02	0.10	-0.54	0.17	-1.59
REBLUP	0.03	-0.39	0.11	-0.47	-0.30	-1.00
MQ	0.02	-0.43	0.09	-0.94	-0.32	-0.99
REBLUP-BC	0.02	-0.29	0.03	0.02	-0.28	-0.32
MQ-BC	0.02	-0.28	0.03	-0.07	-0.26	-0.30
Median values of Relative RMSE (expressed as a percentage)						
EBLUP	0.81	1.22	0.85	0.97	1.37	2.39
REBLUP	0.82	1.01	0.84	1.02	0.99	1.44
MQ	0.82	1.03	0.83	1.46	1.01	1.57
REBLUP-BC	0.91	1.23	0.92	0.86	1.24	1.27
MQ-BC	0.91	1.24	0.92	0.93	1.26	1.49

Table 2. Performance of RMSE estimators in model-based simulation experiments.

Scenario		[0,0]	[e,0]	[0,u]	[0,u]	[e,u]	[e,u]
Areas		1-40	1-40	1-36	37-40	1-36	37-40
Predictor	MSE Estimator	*Median values of Relative Bias (expressed as a percentage)*					
EBLUP	PR	-0.34	1.74	3.82	-17.31	11.32	-40.86
	CCT	3.61	31.24	1.55	2.15	5.95	-3.05
	CCST	0.55	31.22	-3.91	-0.30	2.96	-4.17
REBLUP	CCT	-17.71	-15.76	-20.24	-34.79	-19.51	-36.63
	CCST	-2.01	-8.46	-5.31	-3.58	-7.91	-22.51
	BOOT	-1.19	-4.42	7.38	-19.42	11.37	-31.44
MQ	CCT	-2.98	-16.29	-12.56	6.69	-24.02	177.42
	CCST	0.11	-8.21	-7.77	8.95	-14.10	163.38
REBLUP-BC	CCT	-10.56	-12.46	-11.88	-10.54	-12.57	-18.37
	CCST	-2.95	-2.83	-4.21	-11.27	-5.81	-8.48
	BOOT	-0.21	-6.76	-0.52	-1.25	-4.90	-12.96
MQ-BC	CCT	-6.35	3.48	-7.19	3.92	1.87	5.96
	CCST	-7.18	-11.38	-7.42	3.21	-11.42	-9.20
		Median values of Relative RMSE (expressed as a percentage)					
EBLUP	PR	6.24	18.57	7.20	17.90	22.28	43.19
	CCT	31.51	76.20	31.25	28.37	61.57	51.30
	CCST	22.92	66.27	7.68	18.98	27.15	39.13
REBLUP	CCT	29.52	30.82	28.67	28.58	29.00	38.70
	CCST	27.86	28.47	20.89	22.87	20.25	29.24
	BOOT	10.27	34.92	10.67	14.62	16.61	33.04
MQ	CCT	61.94	61.50	59.88	43.76	59.67	205.30
	CCST	54.77	49.14	50.63	40.58	45.34	189.92
REBLUP-BC	CCT	33.64	45.20	33.21	33.56	45.48	47.18
	CCST	33.30	45.17	33.11	32.99	45.13	47.10
	BOOT	10.12	15.27	10.20	10.60	14.53	18.35
MQ-BC	CCT	36.68	65.37	36.19	38.33	65.70	64.26
	CCST	33.93	44.81	33.55	35.30	44.65	50.55

Table 3. Median values of the Relative Bias (RB) and Relative RMSE (RRMSE) of point estimators in the design-based simulation. All values are expressed as percentages and medians are over the regions of interest.

Estimator	RB(%)	RRMSE(%)
EBLUP	10.79	35.18
REBLUP	-13.08	30.59
MQ	-22.98	35.07
REBLUP-BC	-4.13	31.94
MQ-BC	-6.17	31.57

Table 4. Performance of RMSE estimators in design-based simulation: median values of the percentage Relative Bias and Relative RMSE.

MSE Estimator	PR	CCT	CCST	BOOT
	Median values of Relative Bias (expressed as a percentage)			
EBLUP	6.37	1.79	3.23	
REBLUP		-23.06	3.59	32.12
MQ		-31.59	-24.48	
REBLUP-BC		-14.58	3.00	0.48
MQ-BC		-6.40	-11.01	
	Median values of Relative RMSE (expressed as a percentage)			
EBLUP	30.61	30.67	28.86	
REBLUP		45.79	43.72	61.95
MQ		62.19	55.88	
REBLUP-BC		39.78	39.47	39.81
MQ-BC		45.53	38.38	

Figure 1. Box plots showing area-specific values of the RMSE Ratios for the MSE estimators evaluated in the design-based simulation. The RMSE Ratio is defined as the ratio of the average over repeated sampling of the RMSE estimator for a predictor to the actual RMSE of this predictor under repeated sampling.