Robust Small Area Estimation Under Unit Level Models

Sanjoy K. Sinha *   J.N.K. Rao †

Abstract

Small area estimation has received considerable attention in recent years because of an increasing demand for small area statistics. Basic area level and unit level models have been studied in the literature to obtain empirical best linear unbiased predictors for small area means. Although this classical method is useful for estimating the small area means efficiently under strict model assumptions, it can be highly influenced by the presence of outliers in the data. In this article, the authors investigate the robustness properties of the classical estimators and propose a resistant method for small area estimation, which is useful for downweighting any influential observations in the data when estimating the model parameters. Simulations are carried out to study the behavior of the robust estimators in the presence of outliers, and these estimators are also compared to the ordinary classical estimators. To estimate the mean squared errors of the robust estimators of small area means, a parametric bootstrap method is adopted here, which is applicable to the unit level models with block diagonal covariance structures. Performance of the bootstrap mean squared error estimator is also investigated in the simulation study. The proposed robust method is also applied to some real life data, referred to as the survey and satellite data, obtained from a study described in a statistical journal.

Key Words: Best linear unbiased prediction; Bootstrap method; Mean squared error; Mixed model; Robust estimation; Small area mean.

1. Introduction

Small area estimators have long been used in survey sampling. Demographers commonly use a variety of indirect methods for small area estimation of population and other characteristics of interest, based on either census or on administrative records. The demand for small area statistics has greatly increased worldwide in recent years, which is perhaps due to their growing use in formulating policies and programs, in the allocation of government funds and in regional planning.

Basic area level and unit level models are widely used to derive empirical best linear unbiased prediction (EBLUP), empirical Bayes (EB) and hierarchical Bayes (HB) estimators of small area means (or totals) and associated measures of variability. An extensive account of EBLUP, EB and HB methods for small area estimation can be found in Rao (2003). The classical EBLUP method is applicable for linear mixed models that cover the area level and unit level models. On the other hand, EB and HB methods are more generally applicable for generalized linear mixed models that are used to analyze binary and count data. MSE is used as a measure of variability under the EBLUP and EB approaches, whereas the HB approach uses the posterior variance as a measure of variability assuming a prior distribution on the model parameters. Recent review papers on small area estimation include Pfeffermann (2002), Rao (2003), and Jiang and Lahiri (2006).

Although the classical small area estimators are efficient under strict model assumptions, they can be very sensitive to outliers or departures from the underlying distributions. The deviations from the underlying distributions or assumptions refer to the fact that a small proportion of the data may come from an arbitrary distribution rather than the underlying "true" distribution, which may result in outliers or influential observations in the data. In this article, we address the non-robustness properties of the classical estimators and propose a robust technique for small area estimation that can be used to downweight any outliers in the data when fitting the underlying model.

The paper is organized as follows. Section 2 presents the basic small area models. Section 3 reviews some existing methods for small area estimation and introduces the proposed robust method, which is developed in the framework of maximum likelihood estimation. A small simulation study was carried out to investigate the performance of the robust method in the presence of outliers. The simulation results are presented in Section 4. In section 5, the proposed method of small area estimation is applied to some real life data obtained from a study described in Battese, Harter and Fuller (1988). Section 6 concludes the paper with some discussion.

*School of Mathematics and Statistics, Carleton University Ottawa, Ontario, K1S 5B6 Canada
†School of Mathematics and Statistics, Carleton University Ottawa, Ontario, K1S 5B6 Canada

2. Small Area Models

Models that are commonly used to derive small area estimators can be classified into two groups: (i) Area level models that relate the small area means to area-specific auxiliary variables. Such models are essential if unit level data are not available. (ii) Unit level models that relate the unit values of the study variable to unit-specific auxiliary variables with known area means.

2.1 Basic Area Level Model

To define a basic area level model (also referred to as Fay-Herriot model), we assume that $\theta_i = g(\bar{Y}_i)$ for some function g of the small area mean $\bar{Y}_i$ is related to area-specific auxiliary data $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})^t$ by a linear model

$$\theta_i = \mathbf{z}_i^t \boldsymbol{\beta} + v_i, \quad i = 1, \dots, k \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ is the $p \times 1$ vector of regression coefficients, and v_i's area-specific random effects assumed to be independent $N(0, \sigma_v^2)$. The variance component σ_v^2 is a measure of homogeneity of the areas after accounting for the covariates $\mathbf{z}_i$. To infer about the small area means $\bar{Y}_i$ in model (1), we assume that some direct estimators $\hat{\bar{Y}}_i$ of the means $\bar{Y}_i$ are available and that

$$\hat{\theta}_i = g(\hat{\bar{Y}}_i) = \theta_i + e_i, \quad i = 1, \dots, k \tag{2}$$

where $e_i|\theta_i$ are assumed to be independent $N(0, \tau_i)$ with known sampling variances τ_i. Combining (1) and (2), we obtain

$$\hat{\theta}_i = \mathbf{z}_i^t \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, k \tag{3}$$

where design-induced errors e_i and model errors v_i are assumed to be independent.

Basic area level model was studied by a number of authors for estimating small area means. Fay and Herriot (1979) used model (3) in the context of estimating per-capita income for small areas in the United States with population less than 1,000. They used $\theta_i = \log \bar{Y}_i$, where $\bar{Y}_i$ is the per-capita income in the ith area. Recently, National Research Council (2000) used model (3) to find model-based small area (county) estimates of poor school-age children in the United States. In this application, $\theta_i = \log Y_i$ was chosen, where Y_i represents the true poverty count of the ith small area. For further applications, the reader is referred to Rao (2003, Chapter 7). Various extensions of the basic area level model have been studied to deal with correlated sampling errors arising from spatial dependence of the model errors v_i, and time series and cross-sectional data. They are reviewed in Rao (2003, Chapter 8).

2.2 Basic Unit Level Model

In the basic unit level model, unit-specific auxiliary variables $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^t$ are related to the unit level variable of interest, y_{ij}, through a nested error linear regression model

$$y_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \dots, k, \ j = 1, \dots, n_i \tag{4}$$

where n_i is the number of sampled units observed in the ith area consisting of N_i units. The total number of sampled units observed in all k areas is $n_0 = \sum_{i=1}^k n_i$. The area-specific random effects v_i are assumed to be independent $N(0, \sigma_v^2)$, and independent of the random errors e_{ij}, which are assumed to be independent $N(0, \sigma_e^2)$.

When the population size N_i is large, the ith small area mean $\bar{Y}_i$ may be written as

$$\bar{Y}_i \approx \mu_i = \bar{\mathbf{X}}_i^t \boldsymbol{\beta} + v_i, \tag{5}$$

where $\bar{\mathbf{X}}_i$ is the vector of known means of the $\mathbf{x}_{ij}$ for the ith area. Here μ_i is a linear combination of fixed effects $\boldsymbol{\beta}$ and realized value of area-specific random effect v_i. In this paper, the focus is on μ_i.

Battese, Harter, and Fuller (1988) used model (4) to estimate mean acreage under a crop for counties (small areas) in Iowa using sample survey data in conjunction with satellite information. In their application, N_i ranged from 394 to 965 and n_i from 1 to 6. Rao and Choudhry (1995) used model (4) in a slightly modified form to study the population of unincorporated tax filers from the province of Nova Scotia, Canada. Their proposed model was $y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + x_{ij}^{1/2} e_{ij}$, where y_{ij} and x_{ij} represent the total wages and salaries and gross business income for the jth firm in the ith small area. This model was used to estimate the small area totals Y_i or the means $\bar{Y}_i$.

3. Robust Estimators

Models (3) and (4) in Section 2 are special cases of the general linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \tag{6}$$

where $\mathbf{y}$ is the vector of observed data, $\mathbf{X}$ is the design matrix for the fixed effects and $\mathbf{Z}$ is the design matrix for the random effects. The vector of random effects $\mathbf{v}$ and the vector of random errors $\mathbf{e}$ are assumed to be independently normally distributed with means $\mathbf{0}$ and covariance matrices $\mathbf{G}$ and $\mathbf{R}$, respectively, depending on some vector $\boldsymbol{\theta}$ of variance components. The variance-covariance matrix of $\mathbf{y}$ is obtained as $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta}) = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^t$.

In small area problems, we are generally interested in estimating a linear combination, $\mu = \mathbf{l}^t\boldsymbol{\beta} + \mathbf{m}^t\mathbf{v}$, of the regression coefficients $\boldsymbol{\beta}$ and the realized values of the random effects $\mathbf{v}$, for some specified vectors $\mathbf{l}$ and $\mathbf{m}$ of constants. The function μ is estimated as $\hat{\mu} = \mathbf{l}^t\hat{\boldsymbol{\beta}} + \mathbf{m}^t\hat{\mathbf{v}}$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$ are some suitable estimators of $\boldsymbol{\beta}$ and $\mathbf{v}$, respectively.

3.1 BLUP Estimator

For known $\boldsymbol{\theta}$, Henderson (1975) obtained the best linear unbiased prediction (BLUP) estimator of $\mu = \mathbf{l}^t\boldsymbol{\beta} + \mathbf{m}^t\mathbf{v}$ as

$$\tilde{\mu}_H = t(\boldsymbol{\theta}, \mathbf{y}) = \mathbf{l}^t\tilde{\boldsymbol{\beta}} + \mathbf{m}^t\tilde{\mathbf{v}}, \tag{7}$$

where $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{y}$ is the generalized least squares estimator, also referred to as the best linear unbiased estimator (BLUE), of $\boldsymbol{\beta}$, and $\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\boldsymbol{\theta}) = \mathbf{G}\mathbf{Z}^t\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$ is the BLUP estimator of the vector $\mathbf{v}$ of random effects.

The BLUP estimator $t(\boldsymbol{\theta}, \mathbf{y})$ in (7) depends on the variance components $\boldsymbol{\theta}$, which are unknown in practice. Replacing $\boldsymbol{\theta}$ by an estimator $\hat{\boldsymbol{\theta}}$, we obtain $\hat{\mu}_H = t(\hat{\boldsymbol{\theta}}, \mathbf{y}) = \mathbf{l}^t\hat{\boldsymbol{\beta}} + \mathbf{m}^t\hat{\mathbf{v}}$, which is referred to as the empirical best linear unbiased predictor (EBLUP) of the small area mean μ, where $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{v}} = \tilde{\mathbf{v}}(\hat{\boldsymbol{\theta}})$. Standard procedures for estimating the variance components $\boldsymbol{\theta}$ include maximum likelihood (ML) and restricted maximum likelihood (REML) method of estimation. These methods have been discussed in detail by Searle, Casella and McCulloch (1992).

To calculate the mean squared error (MSE) of the EBLUP estimator $t(\hat{\boldsymbol{\theta}}, \mathbf{y})$, Kackar and Harville (1984) showed that

$$\text{MSE}[t(\hat{\boldsymbol{\theta}}, \mathbf{y})] = \text{MSE}[t(\boldsymbol{\theta}, \mathbf{y})] + E[t(\hat{\boldsymbol{\theta}}, \mathbf{y}) - t(\boldsymbol{\theta}, \mathbf{y})]^2, \tag{8}$$

under normality. An expression for $\text{MSE}[t(\boldsymbol{\theta}, \mathbf{y})]$ can be found in Henderson (1975). It can be shown (see Rao 2003) that

$$\text{MSE}[t(\boldsymbol{\theta}, \mathbf{y})] = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}), \tag{9}$$

where $g_1(\boldsymbol{\theta}) = \mathbf{m}^t(\mathbf{G} - \mathbf{G}\mathbf{Z}^t\mathbf{V}^{-1}\mathbf{Z}\mathbf{G})\mathbf{m}$ and $g_2(\boldsymbol{\theta}) = \mathbf{d}^t(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{d}$, with $\mathbf{d}^t = \mathbf{l}^t - \mathbf{b}^t\mathbf{X}$ and $\mathbf{b}^t = \mathbf{m}^t\mathbf{G}\mathbf{Z}^t\mathbf{V}^{-1}$.

The second term in (8) is generally intractable, and it is necessary to obtain an approximation to this term. Kackar and Harville (1984) proposed a Taylor series approximation. Prasad and Rao (1990) considered a further approximation to this term under some general conditions. It can be shown that

$$E[t(\hat{\boldsymbol{\theta}}, \mathbf{y}) - t(\boldsymbol{\theta}, \mathbf{y})]^2 \approx g_3(\boldsymbol{\theta}), \tag{10}$$

where $g_3(\boldsymbol{\theta}) = \text{tr}[(\partial\mathbf{b}^t/\partial\boldsymbol{\theta})\mathbf{V}(\partial\mathbf{b}^t/\partial\boldsymbol{\theta})^t\bar{\mathbf{V}}(\hat{\boldsymbol{\theta}})]$ with $\bar{\mathbf{V}}(\hat{\boldsymbol{\theta}})$ being the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$.

Combining (9) with (10), a second-order approximation to the MSE of $t(\hat{\boldsymbol{\theta}}, \mathbf{y})$ is obtained as

$$\text{MSE}[t(\hat{\boldsymbol{\theta}}, \mathbf{y})] \approx g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3(\boldsymbol{\theta}). \tag{11}$$

Here the terms $g_2(\boldsymbol{\theta})$ and $g_3(\boldsymbol{\theta})$, due to estimating $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ respectively, are of lower order than the leading term $g_1(\boldsymbol{\theta})$. Prasad and Rao (1990), and Datta and Lahiri (2000) obtained a nearly unbiased estimator of $\text{MSE}[t(\hat{\boldsymbol{\theta}}, \mathbf{y})]$, for linear mixed models with block diagonal covariance structure (see (21)), using a simple moment estimator of $\boldsymbol{\theta}$ and a REML estimator of $\boldsymbol{\theta}$, respectively. It is given by

$$\text{mse}[t(\hat{\boldsymbol{\theta}}, \mathbf{y})] = g_1(\hat{\boldsymbol{\theta}}) + g_2(\hat{\boldsymbol{\theta}}) + 2g_3(\hat{\boldsymbol{\theta}}), \tag{12}$$

and $E(\text{mse}[t(\hat{\boldsymbol{\theta}}, \mathbf{y})]) \approx \text{MSE}[t(\hat{\boldsymbol{\theta}}, \mathbf{y})]$. For example, the bias of (12) is of order $o(k^{-1})$ under the basic unit level model (4) of Section 2.2. Das, Jiang, and Rao (2004) established (12) under ANOVA linear mixed models and REML estimation of $\boldsymbol{\theta}$.

3.2 Previously Studied Robust Methods

The classical least squares estimator of $\boldsymbol{\beta}$ and the ML or REML estimator of the variance components $\boldsymbol{\theta}$ are generally sensitive to outliers in the data. To handle these outliers, a number of resistant methods were studied in the literature. Fellner (1986) studied the basic unit level model (4) and, for known variance components, considered finding robust estimates of the fixed effects $\boldsymbol{\beta}$ and random effects $\mathbf{v}$ by solving the following set of equations simultaneously:

$$\mathbf{X}^t\mathbf{R}^{-1/2}\boldsymbol{\Psi}(\mathbf{R}^{-1/2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{v})) = \mathbf{0} \quad (13)$$
$$\mathbf{Z}^t\mathbf{R}^{-1/2}\boldsymbol{\Psi}(\mathbf{R}^{-1/2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{v})) - \mathbf{G}^{-1/2}\boldsymbol{\Psi}(\mathbf{G}^{-1/2}\mathbf{v}) = \mathbf{0}, \quad (14)$$

where $\boldsymbol{\Psi}(\mathbf{u}) = [\psi_b(u_1), \psi_b(u_2), \ldots]^t$ with ψ_b being the Huber's psi function $\psi_b(u) = u\min(1, b/|u|)$; $b > 0$ is a tuning constant. The common choice of b is 1.345. The choice $b = \infty$ leads to the classical best linear unbiased estimator of $\boldsymbol{\beta}$ and the BLUP estimator of $\mathbf{v}$.

For estimating the variance components $\boldsymbol{\theta} = (\sigma_e^2, \sigma_v^2)^t$ of the unit level model (4), Fellner (1986) suggested robust versions of the REML estimating equations of Harville (1977) in the form

$$\hat{\sigma}_v^2 = \hat{\sigma}_v^2\|\boldsymbol{\Psi}(\hat{\sigma}_v^{-1}\hat{\mathbf{v}})\|^2/[c(k-\lambda)]$$
$$\hat{\sigma}_e^2 = \hat{\sigma}_e^2\|\boldsymbol{\Psi}(\hat{\sigma}_e^{-1}\hat{\mathbf{e}})\|^2/[c\{(n_0-p)-(k-\lambda)\}],$$

where the tuning constant c is chosen as $c = E[\psi_b^2(u)]$ with u being a standard normal distribution and $\hat{\mathbf{e}} = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{Z}\hat{\mathbf{v}})$. Here p is the dimension of $\boldsymbol{\beta}$, and λ is obtained from $\lambda = \mathrm{tr}(\mathbf{D})/\hat{\sigma}_v^2$, where $\mathbf{D}$ is a matrix formed by the last k rows and columns of

$$\begin{bmatrix}\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^t\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}^t\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^t\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}.$$

The estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ depend on the EBLUP residuals $\hat{\mathbf{v}}$ and hence not efficient in the presence of outliers. To improve the efficiency of the estimators, we adopt an alternative robust approach where the fixed effects $\boldsymbol{\beta}$ and the variance components $\boldsymbol{\theta}$ are estimated first robustly using the marginal distribution of the response vector $\mathbf{y}$. Then we consider estimating the random effects $\mathbf{v}$ using the robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Robust estimators in the linear mixed model were also investigated by Richardson and Welsh (1995) and Richardson (1997). They, however, focused on estimating only the fixed effects and the variance components in the mixed model. In the context of small area estimation, it is also important to find suitable predictors of the area-specific random effects $\mathbf{v}$. Our robust method for small area estimation is introduced in the next section.

3.3 Proposed Robust Method

For the linear mixed model (6), it is easy to show that the marginal distribution of the response vector $\mathbf{y}$ is multivariate normal with mean $\mathbf{X}\boldsymbol{\beta}$ and variance-covariance matrix $\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^t$ depending on $\boldsymbol{\theta}$ of variance components. The ML estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are obtained by solving the normal equations

$$\mathbf{X}^t\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \quad (15)$$
$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^t\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) - \mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\right) = \mathbf{0}, \quad (16)$$

where θ_l is the lth component of $\boldsymbol{\theta}$. If certain components of the fitted vector $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ are unusually different from the corresponding observed values $\mathbf{y}$, then we have indication of apparent outliers in the data.

To handle such outliers in the $\mathbf{y}$ values, we consider estimating the fixed effects and variance components by solving the equations

$$\mathbf{X}^t\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}) = \mathbf{0} \quad (17)$$
$$\boldsymbol{\Psi}^t(\mathbf{r})\mathbf{U}^{1/2}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}) - \mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{K}\right) = \mathbf{0} \quad (18)$$

with $\mathbf{r} = \mathbf{U}^{-1/2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$. Here $\mathbf{U}$ is a diagonal matrix with its elements being equal to the diagonal elements of the covariance matrix $\mathbf{V}$, and $\mathbf{K}$ is a diagonal matrix chosen as $\mathbf{K} = c\mathbf{I}$ with $c = E[\psi_b^2(r)]$; the expectation is taken over a standard normal distribution. Note that for the choice $b = \infty$, the robust estimators reduce to the corresponding classical ML estimators.

Equations (17) and (18) can be solved iteratively to obtain robust estimators of both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Here we adopt a Newton-Raphson algorithm for solving these estimating equations. Using first-order Taylor series expansion around $\boldsymbol{\beta}_0$, the left side of (17) can be approximated as

$$\mathbf{X}^t\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}(\boldsymbol{\beta})) \approx \mathbf{X}^t\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}(\boldsymbol{\beta}_0)) - \mathbf{X}^t\mathbf{U}^{-1/2}\mathbf{D}(\boldsymbol{\beta}_0)\mathbf{U}^{1/2}\mathbf{V}^{-1}\mathbf{X}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$$

where $\mathbf{D}(\boldsymbol{\beta})$ is a diagonal matrix with its jth diagonal element being equal to $(\partial/\partial r_j)\psi_b(r_j)$. This approximation leads to an iterative equation of the form

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left(\mathbf{X}^t\mathbf{U}^{-1/2}\mathbf{D}(\boldsymbol{\beta}^{(m)})\mathbf{U}^{1/2}\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}(\boldsymbol{\beta}^{(m)})). \tag{19}$$

Similarly, equation (18) can be solved for θ_l by using the Newton-Raphson algorithm that leads to the iterative equation

$$\theta_l^{(m+1)} = \theta_l^{(m)} - \left[\Phi'(\theta_l^{(m)})\right]^{-1}\Phi(\theta_l^{(m)}), \tag{20}$$

where

$$\Phi(\theta_l) = \boldsymbol{\Psi}^t(\mathbf{r})\mathbf{U}^{1/2}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}) - \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{K}\right)$$

and $\Phi'(\theta_l^{(m)}) = \partial\Phi(\theta_l)/\partial\theta_l$ evaluated at $\theta_l = \theta_l^{(m)}$. The derivative $\Phi'(\theta_l)$ in (20) can be obtained as

$$\begin{aligned}\frac{\partial\Phi(\theta_l)}{\partial\theta_l} &= 2\,\frac{\partial}{\partial\theta_l}\left[\Psi^t(\mathbf{r})\mathbf{U}^{1/2}\mathbf{V}^{-1}\right]\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\mathbf{U}^{1/2}\Psi(\mathbf{r}) \\ &\quad + \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\,\mathbf{K}\right),\end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial\theta_l}\left[\Psi^t(\mathbf{r})\mathbf{U}^{1/2}\mathbf{V}^{-1}\right] &= \left[\frac{\partial}{\partial\theta_l}\Psi^t(\mathbf{r})\right]\mathbf{U}^{1/2}\mathbf{V}^{-1} + \Psi^t(\mathbf{r})\left[\frac{\partial}{\partial\theta_l}\mathbf{U}^{1/2}\right]\mathbf{V}^{-1} \\ &\quad -\Psi^t(\mathbf{r})\mathbf{U}^{1/2}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\end{aligned}$$

with $\partial\mathbf{U}/\partial\theta_l = \mathbf{I}$ being an identity matrix, and $\partial\Psi(\mathbf{r})/\partial\theta_l = (\partial/\partial\theta_l)[\psi_b(r_1), \psi_b(r_2), \ldots]^t$ with $(\partial/\partial\theta_l)\psi_b(r_j) = -(1/2)U_j^{-1}r_j\psi_b'(r_j)$.

The complete algorithm for the robust estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be described as follows:

1. Choose initial values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$. Set $m = 0$.
2. (a) Calculate $\boldsymbol{\beta}^{(m+1)}$ from the iterative equation (19).
   (b) Calculate $\theta_l^{(m+1)}$ $(l = 1, 2, \ldots)$ from (20).
   (c) Set $m = m + 1$.
3. Continue step 2 until a convergence is achieved. Declare the estimates at convergence to be the robust maximum likelihood (RML) estimators $\hat{\boldsymbol{\beta}}_M$ and $\hat{\boldsymbol{\theta}}_M$ of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, where the subscript M stands for M (or maximum likelihood type) estimation introduced by Huber (1964).

Based on the robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, we then consider estimating the random effects $\mathbf{v}$ by solving (14), as suggested by Fellner (1986). We adopt a similar Newton-Raphson algorithm as described earlier for estimating the random effects. The robust estimators $\hat{\mathbf{v}}_M$ of the random effects $\mathbf{v}$ are then used to predict $\mu = \mathbf{l}^t\boldsymbol{\beta} + \mathbf{m}^t\mathbf{v}$ as $\hat{\mu}_M = t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \hat{\mathbf{v}}_M) = \mathbf{l}^t\hat{\boldsymbol{\beta}}_M + \mathbf{m}^t\hat{\mathbf{v}}_M$. The predictor $\hat{\mu}_M$ is referred to as the robust EBLUP (REBLUP) of μ, and it becomes the ordinary EBLUP for the choice $\psi_b(r) = r$ of Huber's psi function.

3.4 Asymptotics

A special case of the general linear mixed model (6) that covers many small area models with block diagonal covariance structure may be defined as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i + \mathbf{e}_i, \quad i = 1, \ldots, k \tag{21}$$

where the variance-covariance matrix of $\mathbf{y}_i$ is obtained as $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{R}_i + \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i^t$.

Robust EBLUP predictor of area-specific parameter $\mu_i = \mathbf{l}_i^t\boldsymbol{\beta} + \mathbf{m}_i^t\mathbf{v}_i$ is given by $\hat{\mu}_{iM} = \mathbf{l}_i^t\hat{\boldsymbol{\beta}}_M + \mathbf{m}_i^t\hat{\mathbf{v}}_{iM} = t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \hat{\mathbf{v}}_{iM})$. For example, under the basic unit level model (4), the small area mean $\bar{Y}_i \approx \mu_i$ with $\mathbf{l}_i = \bar{X}_i$ and $\mathbf{m}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^t$, where 1 is in the ith position of the vector $\mathbf{m}_i$.

Under suitable regularity conditions, $k^{1/2}(\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})$ is asymptotically distributed as Gaussian with mean zero and covariance matrix consistently estimated by $\hat{\mathbf{C}}_1^{-1}\hat{\mathbf{B}}_1\hat{\mathbf{C}}_1^{-1}$, where

$$\hat{\mathbf{B}}_1 = k^{-1}\sum_{i=1}^{k}\mathbf{X}_i^t\mathbf{V}_i^{-1}\mathbf{U}_i^{1/2}\mathbf{Q}_i\mathbf{U}_i^{1/2}\mathbf{V}_i^{-1}\mathbf{X}_i$$

with $\mathbf{Q}_i = \Psi(\mathbf{r}_i)\Psi^t(\mathbf{r}_i)$ and

$$\hat{\mathbf{C}}_1 = -k^{-1}\sum_{i=1}^{k}\mathbf{X}_i^t\mathbf{U}_i^{-1/2}\mathbf{D}_i\mathbf{U}_i^{1/2}\mathbf{V}_i^{-1}\mathbf{X}_i,$$

where $\mathbf{D}_i$ is a diagonal matrix with its jth diagonal element $D_{ij} = (\partial/\partial r_{ij})\psi_b(r_{ij})$.

Also, under regularity conditions, $k^{1/2}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta})$ is asymptotically distributed as Gaussian with mean zero and covariance matrix consistently estimated by $\hat{\mathbf{C}}_2^{-1}\hat{\mathbf{B}}_2\hat{\mathbf{C}}_2^{-1}$, where matrix $\hat{\mathbf{B}}_2$ has its (l,m)th element

$$\hat{\mathbf{B}}_{2,l,m} = k^{-1}\sum_{i=1}^{k}\eta_i(\theta_l)\eta_i(\theta_m)$$

and $\hat{\mathbf{C}}_2$ has its (l,m)th element

$$\hat{\mathbf{C}}_{2,l,m} = -k^{-1}\sum_{i=1}^{k}(\partial/\partial\theta_m)\eta_i(\theta_l)$$

where $\eta_i(\theta_l)$ has the form

$$\eta_i(\theta_l) = \mathbf{\Psi}^t(\mathbf{r}_i)\mathbf{U}_i^{1/2}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_l}\mathbf{V}_i^{-1}\mathbf{U}_i^{1/2}\mathbf{\Psi}(\mathbf{r}_i) - \mathrm{tr}\left(\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_l}\mathbf{K}_i\right).$$

The matrices $\hat{\mathbf{B}}_1$, $\hat{\mathbf{B}}_2$, $\hat{\mathbf{C}}_1$, and $\hat{\mathbf{C}}_2$ are all evaluated at $(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M)$.

3.5 Estimation of MSE of Robust Small Area Estimators

Estimation of MSE of small area estimators is a challenging problem even in the case of classical EBLUP estimators. Several methods have been proposed by a number of authors for the case of EBLUP under the model (21) with block diagonal covariance structure. Methods studied include a linearization method of Prasad and Rao (1990), a jackknife method of Jiang, Lahiri and Wan (2002), and a bootstrap method of Hall and Maiti (2006). For a review of some of these methods see Rao (2003).

Given the complex form of robust estimators of small area means, the corresponding estimators of MSE cannot be obtained in any closed form. Here we adopt a bootstrap technique to estimate the MSEs of robust small area estimators under model (21) with block diagonal covariance structure. Initially, we studied the nonparametric bootstrap method by generating bootstrap samples of residuals from the original data. The results were not so encouraging, as the bootstrap samples were influenced by the outliers in the original data. Note that in sampling with replacement, a bootstrap sample may have outliers that can outnumber the "good" data points, and consequently, prediction for the small area means may not be valid at all.

Similarly to Hall and Maiti (2006), we then investigate a parametric bootstrap method for approximating $\mathrm{MSE}(\hat{\mu}_{iM}) = E\{t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \hat{\mathbf{v}}_{iM}) - \mu_i\}^2$. We adopt the following bootstrap algorithm:

1. For given $\hat{\boldsymbol{\beta}}_M$ and $\hat{\boldsymbol{\theta}}_M$, generate the area-specific random effects $\mathbf{v}_i^*$ and the random errors $\mathbf{e}_i^*$ from $N(\mathbf{0}, \mathbf{G}_i(\hat{\boldsymbol{\theta}}_M))$ and $N(\mathbf{0}, \mathbf{R}_i(\hat{\boldsymbol{\theta}}_M))$ respectively to create a bootstrap sample from the model

$$\mathbf{y}_i^* = \mathbf{X}_i\hat{\boldsymbol{\beta}}_M + \mathbf{Z}_i\mathbf{v}_i^* + \mathbf{e}_i^*, \quad i = 1, \ldots, k. \tag{22}$$

   Compute the corresponding bootstrap population parameter $\mu_i^* = t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \mathbf{v}_i^*) = \mathbf{l}_i^t\hat{\boldsymbol{\beta}}_M + \mathbf{m}_i^t\mathbf{v}_i^*$.

2. Generate B bootstrap samples $\{\mathbf{y}_i^{*(1)}, \ldots, \mathbf{y}_i^{*(b)}, \ldots, \mathbf{y}_i^{*(B)}\}$ from the bootstrap population model (22). Compute the corresponding bootstrap estimates of the population parameter $\mu_i^{*(b)} = t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \mathbf{v}_i^{*(b)})$ as $\hat{\mu}_{iM}^{*(b)} = t_M(\hat{\boldsymbol{\beta}}_M^{(b)}, \hat{\boldsymbol{\theta}}_M^{(b)}, \hat{\mathbf{v}}_{iM}^{(b)}) = \mathbf{l}_i^t\hat{\boldsymbol{\beta}}_M^{(b)} + \mathbf{m}_i^t\hat{\mathbf{v}}_{iM}^{(b)}$, for $b = 1, \ldots, B$, where $\hat{\boldsymbol{\beta}}_M^{(b)}$, $\hat{\boldsymbol{\theta}}_M^{(b)}$ and $\hat{\mathbf{v}}_{iM}^{(b)}$ are the robust bootstrap estimates of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\mathbf{v}_i$ for the bth bootstrap sample $\mathbf{y}_i^{*(b)}$.

3. Obtain a bootstrap estimate of MSE$[t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \hat{\mathbf{v}}_{iM})]$ as

$$\text{MSE}^B[t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \hat{\mathbf{v}}_{iM})] = \frac{1}{B}\sum_{b=1}^{B}\left\{t_M(\hat{\boldsymbol{\beta}}_M^{(b)}, \hat{\boldsymbol{\theta}}_M^{(b)}, \hat{\mathbf{v}}_{iM}^{(b)}) - t_M(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M, \mathbf{v}_i^{*(b)})\right\}^2.$$

The proposed bootstrap may be motivated by noting that the focus is on deviations from the working assumption of normality of $\mathbf{v}_i$ and $\mathbf{e}_i$ and that it is natural to use robust parameter estimates for drawing bootstrap samples since MSE$(\hat{\mu}_{iM})$ is not sensitive to outliers. Small simulations under the basic unit level model (4) were carried out to investigate the performance of this bootstrap method in Section 4, and the results were found to be very encouraging.

4. Simulation Study

4.1 Contamination Distributions

We ran a series of simulations using the basic unit level model (4) with a single auxiliary variable x:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij} \tag{23}$$

for $i = 1, \ldots, k$, $j = 1, \ldots, n$ with $k = 40$ and $n = 4$. The values x_{ij} of the auxiliary variable were generated from the normal distribution $N(\mu_x, \sigma_x^2)$ with $\mu_x = 1$ and $\sigma_x^2 = 1$. The area-specific random effects v_i were generated from the contamination distribution $(1-\gamma_1)N(0,\sigma_v^2) + \gamma_1 N(0, \sigma_v^{*2})$. This means that a $(1-\gamma_1)$ proportion of the v_i's were generated from the underlying "true" distribution $N(0,\sigma_v^2)$ and the remaining γ_1 proportion of the random effects were generated from the "arbitrary" contaminated distribution $N(0,\sigma_v^{*2})$. The choice $\gamma_1 = 0$ indicates no contamination of the distribution. Similarly, the random errors e_{ij} were generated from the contamination distribution $(1-\gamma_2)N(0,\sigma_e^2) + \gamma_2 N(0,\sigma_e^{*2})$. For the underlying distributions, we set $\sigma_e^2 = \sigma_v^2 = 1$ and for the contaminated distributions, we set $\sigma_e^{*2} = \sigma_v^{*2} = 25$ and the proportion of contamination $\gamma_1 = \gamma_2 = .10$. We considered four possible combinations $\{(0,0), (0,v), (0,e), (v,e)\}$ of contamination, where $(0,0)$ indicates no contamination of the distributions, $(0,v)$ indicates the contamination only in the distribution of the area-specific random effects v_i, and so on. The regression coefficients were fixed at $(\beta_0, \beta_1) = (1,1)$. We ran four sets of simulations each of size 500. Also when computing the mean squared error of a small area estimator by the bootstrap method, we used bootstrap samples of sizes 100, 200 and 500. We chose different sizes of bootstrap samples to study the variability among the bootstrap estimates of mean squared errors.

Table 1 presents the simulated biases and mean squared errors of the estimators of the regression coefficients (β_0, β_1) and the underlying variance components (σ_e^2, σ_v^2) obtained from the two robust methods (proposed and Fellner's) and the classical generalized least squares method discussed earlier. Here the classical generalized least squares and the proposed RML method perform almost equally well in the case of no contamination of the data. This indicates that when there are no outliers in the data, the proposed RML is almost as efficient as the generalized least squares. On the other hand, Fellner's approach performed poorly by exhibiting larger biases and mean squared errors for the estimators of the variance components as compared to the other two methods.

In the presence of outliers in the random effects v_i, the classical estimators of the variance component σ_v^2 are heavily influenced by these outliers and produced much larger biases and mean squared errors as compared to the robust methods. Between the two robust methods, our proposed RML method provides better results for the estimation of variance components.

In the case of outliers in the random errors e_{ij}, the RML method appears to provide uniformly better results than the classical least squares and Fellner's robust approaches. Also when we consider outliers in both the area-specific random effects v_i and the random errors e_{ij}, the RML method still provides almost uniformly better results than the other two methods. As the RML method provides efficient estimates in almost all situations considered, we naturally recommend this approach for the estimation of model parameters as well as for the prediction of small area means.

In the next step, we consider the estimation of small area means $\mu_i = \beta_0 + \beta_1 \bar{X}_i + v_i$, where $\bar{X}_i$ is the known population mean of the auxiliary variable in the ith area. We chose $\bar{X}_i = 1$ for each $i = 1, \ldots, k$, for simplicity. We considered the same set of simulated samples as used for Table 1. Table 2 presents average simulated absolute biases and average simulated mean squared errors (averages are taken over the areas) of the estimators of small area means for both robust and classical methods. In the case of uncontaminated data, classical EBLUP appears to be the most efficient, as expected. The robust estimators are also seen to be almost as efficient as the EBLUP. Under contamination of the distribution of random errors e_{ij} only, the two robust methods produced much more efficient estimators than those obtained from the classical EBLUP. But under contamination of the distribution of random

Table 1: Simulated biases and mean squared errors (MSEs) of robust and classical estimators.

	ML		Proposed RML		Fellner's method	
Parameter	Bias	MSE	Bias	MSE	Bias	MSE
No contamination						
$\beta_0 = 1$	0.000	0.041	0.001	0.043	-0.001	0.043
$\beta_1 = 1$	-0.002	0.008	-0.004	0.009	-0.003	0.009
$\sigma_e^2 = 1$	-0.019	0.016	0.016	0.022	0.095	0.035
$\sigma_v^2 = 1$	-0.011	0.071	-0.042	0.077	0.152	0.121
Contamination in $\mathbf{v}$						
$\beta_0 = 1$	-0.003	0.099	-0.005	0.055	-0.006	0.054
$\beta_1 = 1$	0.002	0.008	0.001	0.009	0.001	0.008
$\sigma_e^2 = 1$	-0.010	0.017	0.028	0.024	0.109	0.041
$\sigma_v^2 = 1$	2.212	7.823	0.379	0.303	0.512	0.453
Contamination in $\mathbf{e}$						
$\beta_0 = 1$	-0.010	0.068	-0.010	0.049	-0.010	0.048
$\beta_1 = 1$	-0.003	0.025	-0.002	0.013	-0.003	0.012
$\sigma_e^2 = 1$	2.485	7.009	0.522	0.329	0.542	0.345
$\sigma_v^2 = 1$	-0.037	0.179	-0.092	0.116	0.166	0.173
Contamination in both $\mathbf{v}$ and $\mathbf{e}$						
$\beta_0 = 1$	0.012	0.128	0.009	0.067	0.011	0.064
$\beta_1 = 1$	0.000	0.029	0.004	0.014	0.003	0.013
$\sigma_e^2 = 1$	2.367	6.408	0.600	0.435	0.566	0.380
$\sigma_v^2 = 1$	2.265	8.506	0.420	0.376	0.616	0.623

Table 2: Simulated absolute biases and MSEs of estimators of small area means (averaged over areas).

	EBLUP		REBLUP		Fellner's method	
Contamination	Bias	MSE	Bias	MSE	Bias	MSE
$(0, 0)$	.0174	.2066	.0165	.2140	.0168	.2122
$(v, 0)$	.0152	.2290	.0151	.2235	.0150	.2229
$(0, e)$	.0261	.5055	.0216	.2993	.0220	.2950
(v, e)	.0310	.6578	.0201	.3226	.0189	.3425

effects v_i only, the robust and classical methods gave similar results, which is an interesting finding here. When the distributions of both v_i and e_{ij} are contaminated, the proposed robust method appears to provide the most efficient estimators of the small area means in terms of both biases and mean squared errors.

In the last step of the simulations, we studied the proposed bootstrap method for estimating the mean squared errors of the proposed REBLUP estimators of the small area means μ_i. Table 3 presents the simulated mean squared errors (actual) and the simulated expected values of the estimated mean squared errors of the small area estimators $\hat{\mu}_{iM} = \hat{\beta}_{0M} + \hat{\beta}_{1M}\bar{X}_i + \hat{v}_{iM}$. We ran the simulations for three different sizes $(100, 200, 500)$ of bootstrap samples. The simulation size is 500 in each case, as used earlier. For each sample generated in the simulation study, we obtained a set of estimates of model parameters. Using these estimates, we then generated bootstrap samples of size B (B=200, for example) to approximate the mean squared errors of small area estimates for that particular sampled data. It appears that the simulated expectation of the estimated mean squared errors are close to the corresponding simulated mean squared errors under all contamination schemes. It is clear from Table 3 that even the size 100 of bootstrap samples is large enough to provide a good approximation to the simulated expectation. However, the variability in the bootstrap estimates was found to be less for larger sizes of bootstrap samples, as expected.

Table 3: Actual (simulated) and estimated MSEs of proposed REBLUP estimators (averaged over areas).

	Boot size = 100		Boot size = 200		Boot size = 500	
Contamination	Actual	Estimated	Actual	Estimated	Actual	Estimated
$(0,0)$	.2134	.2156	.2140	.2130	.2147	.2146
$(v,0)$	.2259	.2321	.2235	.2290	.2245	.2292
$(0,e)$	.2925	.2824	.2993	.2854	.2946	.2832
(v,e)	.3333	.3315	.3226	.3315	.3345	.3294

Table 4: Simulated biases and MSEs of estimators of fixed effects under misspecified t distributions.

	ML		RML	
Parameter	Bias	MSE	Bias	MSE
$\beta_0 = 1$	-0.041	0.394	-0.008	0.123
$\beta_1 = 1$	0.003	0.090	-0.004	0.024

4.2 t Distributions

Note that in the simulation study presented in Section 4.1, we considered contaminated normal distributions for both random effects and random error terms. Here we investigate the effects of misspecified errors on the small area estimates when the area-specific random effects v_i and the random errors e_{ij} in (23) both follow a t distribution with two degrees of freedom. The t distribution is also symmetric and exhibits long tails that can lead to outliers. The same design as considered earlier is used to generate the data and to compute the small area estimators. The results are presented in Tables 4 and 5. Table 4 shows that our proposed robust method is substantially better than the classical method in estimating β_0 and β_1. Also note that the bootstrap method based on normal residuals produced expected estimated MSE .518, which is close to the actual (simulated) MSE value of 0.551, as shown in Table 5.

5. Example: Survey and Satellite Data

Battese, Harter and Fuller (1988) presented and analyzed survey and satellite data for corn and soybeans in Iowa counties (small areas). The authors considered the estimation of areas under corn and soybeans for 12 counties in north-central Iowa, based on the 1978 June Enumerative Survey and satellite pixel data. Areas of corn and soybeans were obtained in 37 sample segments from the 12 counties by interviewing farm operators. The data set contains the number segments in each county, the number of hectares of corn and soybeans for each sample segment, the number of pixels classified as corn and soybeans for each sample segment, and the mean number of pixels per segment in each county classified as corn and soybeans (a segment is about 250 hectares and a pixel is about 0.45 hectares).

Battese, Harter and Fuller (1988) identified one observation in Hardin county to be an influential outlier, and they simply deleted this observation when predicting the corn and soybeans areas based on the unit level model. We have particularly chosen this data set to investigate the influence of the outlier on the classical estimates and also to explore the ability of the robust method to identify and downweight such influential points when estimating the parameters of interest.

Table 5: Simulated absolute biases and mean squared errors of EBLUP and REBLUP estimators of small area means (averaged over areas) under misspecified t distributions.

EBLUP		REBLUP	
Bias	MSE	Bias	MSE
0.0674	1.719	0.0334	0.551

Table 6: Robust and classical analysis of survey and satellite data (standard errors in parentheses).

Coefficients	Classical ML	Proposed RML	Fellner's method
Intercept (β_0)	$18.09_{(31.11)}$	$29.14_{(40.07)}$	35.73
Corn pixels (β_1)	$.3657_{(.0652)}$	$.3576_{(.0808)}$	.3480
Soybeans pixels (β_2)	$-.0302_{(.0679)}$	$-.0694_{(.0978)}$	-.0871
σ_e^2	$280.2_{(71.55)}$	$225.6_{(63.94)}$	229.7
σ_v^2	$47.80_{(56.51)}$	$102.7_{(64.57)}$	154.4

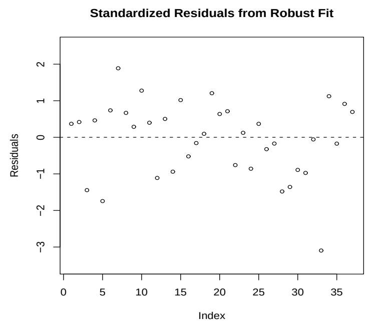

Figure 1: Residual analysis of survey and satellite data from robust method.

The crop hectares for corn in sample segments within counties are expressed as a function of the satellite data for those sample segments in the form

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + v_i + e_{ij}, \tag{24}$$

where y_{ij} is the number of hectares of corn in the jth segment of the ith county ($i = 1, \ldots, k$, where $k = 12$; $j = 1, \ldots, n_i$) with n_i being the number of sample segments in the ith county, and with x_{1ij} and x_{2ij} being the corresponding number of pixels classified as corn and soybeans, respectively. The random effects, v_i ($i = 1, \ldots, k$), are assumed to be iid $N(0, \sigma_v^2)$ independent of the random errors, e_{ij} ($i = 1, \ldots, k; j = 1, \ldots, n_i$), which are assumed to be iid $N(0, \sigma_e^2)$.

The small area mean hectares of corn per segment in the ith county under model (24) is taken as

$$\mu_i = \beta_0 + \beta_1 \bar{X}_{1i} + \beta_2 \bar{X}_{2i} + v_i, \tag{25}$$

where $\bar{X}_{1i}$ and $\bar{X}_{2i}$ are the population mean number of pixels classified as corn and soybeans per segment, respectively, in the ith county. These population means are known from the satellite classifications for all segments in the ith county.

Table 6 presents the robust and classical estimates of the fixed effects parameters ($\beta_0, \beta_1, \beta_2$) and the variance components (σ_e^2, σ_v^2). For given (σ_e^2, σ_v^2), the classical estimates of ($\beta_0, \beta_1, \beta_2$) are obtained as the generalized least squares estimates. Also we used the ML estimates of (σ_e^2, σ_v^2) as the classical estimates. As robust methods, we used the proposed M estimation and Fellner's approach for estimating the regression coefficients and the variance components. It is clear from Table 6 that the two robust methods produce similar estimates for all the model parameters except for the variance component σ_v^2, which is overestimated by the Fellner's method. On the other hand, the intercept parameter β_0 and the variance component σ_v^2 are heavily underestimated by the classical method. The ML estimate of σ_e^2 is overestimated here, which is perhaps due to the large outlier in the data set.

We also studied the standardized residuals $r_{ij} = (y_{ij} - \hat{\beta}_{0M} - \hat{\beta}_{1M} x_{1ij} - \hat{\beta}_{2M} x_{2ij})/(\hat{\sigma}_v^2 + \hat{\sigma}_e^2)^{1/2}$ from the robust fit to investigate if the robust method can detect the outlier and hence downweight this when fitting the model. The residual plot is shown in Figure 1. As expected, the largest standardized residual (which is less than -3) corresponds to the outlying observation in the Hardin county.

We also obtained the estimates of the mean hectares of corn per segment using model (25). Results from the robust and classical methods are given in Table 7. Predictions from these methods are somewhat similar for most of the counties with a few exceptions. As expected, the largest difference in estimates is observed in the Hardin county, where the two robust methods give estimates of about 137 hectares corresponding to the estimate of about

Table 7: Predicted mean hectares of corn per segment (bootstrap MSE estimates in parentheses).

		Estimated hectares	
County	Sample segments	EBLUP	REBLUP
Cerro Gordo	1	$122.2_{(64.0)}$	$123.7_{(97.0)}$
Hamilton	1	$123.2_{(61.9)}$	$125.3_{(93.4)}$
Worth	1	$113.9_{(87.6)}$	$110.3_{(90.4)}$
Humboldt	2	$115.4_{(61.8)}$	$114.1_{(73.6)}$
Franklin	3	$136.1_{(117.6)}$	$140.8_{(53.9)}$
Pocanhontas	3	$108.4_{(59.3)}$	$110.8_{(52.0)}$
Winnibago	3	$116.8_{(57.2)}$	$115.2_{(48.4)}$
Wright	3	$122.6_{(49.8)}$	$122.7_{(58.7)}$
Webster	4	$110.9_{(71.4)}$	$113.5_{(39.9)}$
Hancock	5	$124.4_{(53.2)}$	$124.1_{(40.0)}$
Kossuth	5	$113.4_{(106.5)}$	$109.5_{(37.5)}$
Hardin	6	$131.3_{(51.0)}$	$136.9_{(39.3)}$

131 hectares obtained from the classical method. We also calculated the bootstrap mean squared errors of the classical EBLUP and the REBLUP estimates of the area means based on 500 bootstrap samples. It is interesting to note that the REBLUP method gives a nice pattern in the values of the estimates of the mean squared errors in which larger sample sizes correspond to smaller mean squared errors. But the classical bootstrap method does not reflect this pattern in general. From a closer look on the generated bootstrap samples, we observed that because of the small sample size and large estimates of the variance components as used here, the bootstrap samples tend to have a few outlying points in many situations. The classical bootstrap estimates of the variance components are generally affected by those outlying points. In some cases, the classical estimates of σ_v^2 were found to be close to 0, which in turn underestimate the mean squared errors. One should, therefore, take extra precaution when using the classical bootstrap method for approximating the mean squared errors, especially when the sample size is small and the variance components are large.

6. Discussion

Outliers are common in many real life data, and it is important to adopt suitable robust methods to deal with such outliers. The proposed REBLUP method of estimation of small area means is a significant development toward this direction.

In the absence of a suitable analytical approach for estimating the mean squared error of a small area estimator, in this article we have relied on the bootstrap approach to obtain an estimator of the mean squared error. For moderate sized bootstrap samples as used in the simulation study, these bootstrap estimates appear to provide good approximations to the actual mean squared errors. The proposed model based bootstrap technique, however, has a limitation in the sense that it depends on the working assumption on the distributions of the area-specific effects v_i and the random errors e_{ij}. Although fully nonparametric bootstrap can be used under no contamination of the data, this technique can be very sensitive to potential outliers in the data, and one should not use this nonparametric approach without adopting any robust adjustment. As a trade-off between the two approaches, one may consider a nonparametric weighted bootstrap approach, where the weights are determined by the fitted mixed effects model using the robust method. The bootstrap samples are then generated from the original sample and the corresponding weights associated with each data points. Work remains to be done in this direction.

In this paper, we studied outliers in the distributions of the random effects and the random errors, which are the most common type of outliers in the context of small area estimation. However, if there are outliers in the auxiliary variables, we can extend the proposed RML method to accommodate such outliers. In this case, the RML estimating equation (16) for $\boldsymbol{\beta}$ can be modified as

$$\mathbf{X}^t\mathbf{W}\mathbf{V}^{-1}\mathbf{U}^{1/2}\boldsymbol{\Psi}(\mathbf{r}) = \mathbf{0}$$

where $\mathbf{W}$ is a diagonal matrix with its diagonal elements being equal to some weight function $w(x)$ obtained such that the outliers in the auxiliary variables x's are downweighted when estimating the parameters. When x's are continuous, a common choice of the weight function w is a function of the Mahalanobis distance $(\mathbf{x}-\mathbf{m}_x)^t\mathbf{S}_x^{-1}(\mathbf{x}-\mathbf{m}_x)$, for some robust estimates $\mathbf{m}_x$ and $\mathbf{S}_x$ of the location and scale of $\mathbf{x}$, respectively (Sinha 2004). One limitation of such weight function is its inability to handle outliers in other-than-continuous variables. When the auxiliary variables are discrete, alternatively, one can use Mallows weights described in de Jongh, de Wet and Welsh (1988). These weights are nonparametric in the sense that no assumption is needed for the distribution of the auxiliary variables.

Chambers and Tzavidis (2006) proposed a robust M-quantile approach to small area estimation that is based on quantile regression without specifying the random effects to explain between small area variation unlike the linear mixed model (6) studied in our paper. However, they observed that the M-quantile estimators can be significantly biased. Tzavidis and Chambers (2007) proposed a bias correction to the M-quantile estimator. It would be interesting to compare our robust method of small area estimation with the bias-adjusted M-quantile method in terms of efficiency.

REFERENCES

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988), "An Error Components Model for Prediction of County Crop Areas Using Survey And Satellite Data," *Journal of the American Statistical Association*, 83, 28-36.

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

Das, K., Jiang, J., and Rao, J. N. K. (2004), "Mean Squared Error of Empirical Predictor," *The Annals of Statistics*, 32, 818-840

Datta, G. S. and Lahiri, P. (2000), "A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems," *Statistica Sinica*, 10, 613-627

de Jongh, P. J., de Wet, T. and Welsh, A. H. (1988), "Mallows-Type Bounded- Influence-Regression Trimmed Means," *Journal of the American Statistical Association*, 83, 805-810.

Fay, R. E. and Herriot, R. A. (1979), "Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, 74, 269-277.

Fellner, W. H. (1986), "Robust Estimation of Variance Components," *Technometrics*, 28, 51-60.

Hall, P. and Maiti, T. (2006),"On parametric bootstrap methods for small area prediction," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 68, 221-238.

Harville, D. A. (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to related Problems," *Journal of the American Statistical Association*, 72, 320-340.

Henderson, C. R. (1975), "Best Linear Unbiased Estimation and Prediction Under a Selection Model," *Biometrics*, 31, 423-447.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73-101.

Jiang, J., and Lahiri, P. (2006), "Estimation of Finite Population Domain Means: A Model-assisted Empirical Best Prediction Approach," *Journal of the American Statistical Association*, 101, 301-311.

Jiang, J., Lahiri, P., and Wan, S.-M. (2002), "A Unified Jackknife Theory for Empirical Best Prediction With M-Estimation," *Annals of Statistics*, 30, 1782-1810.

Kackar, R. N., and Harville, D. A. (1984), "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models," *Journal of the American Statistical Association*, 79, 853-862.

National Research Council (2000), *Small Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*, C. F. Citro and G. Kalton (Eds.), Committee on National Statistics, Washington, DC: National Academy Press.

Pfeffermann, D. (2002), "Small Area Estimation - New Developments and Directions," *International Statistical Review*, 70, 125-143.

Prasad, N. G. N. and Rao, J. N. K. (1990), "The Estimation of the Mean Squared Error of Small Area Estimators," *Journal of the American Statistical Association*, 85, 163-171.

Qaqish, B. F., and Preisser, J. S. (1999), "Resistant Fits for Regression with Correlated Outcomes: An Estimating Equations Approach," *Journal of Statistical Planning and Inference*, 75, 415-431.

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

Richardson, A. M. (1997), "Bounded Influence Estimation in the Mixed Linear Model", *Journal of the American Statistical Association*, 92, 154-161.

Richardson, A. M. and Welsh, A. H. (1995), "Robust Estimation in the Mixed Linear Model", *Biometrics*, 51, 1429-1439.

Searl, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.

Sinha, S. K. (2004), "Robust analysis of generalized linear mixed models," *Journal of the American Statistical Association*, 99, 451-460.

Tzavidis, N., and Chambers, R. (2007), "Robust prediction of small area means and distributions," *CCSR Working Paper 2007-08*, University of Manchester [available from http://www.ccsr.ac.uk/publications/working/#2007-08]