THE MEAN-SQUARED ERROR OF A RANK-BASED PREDICTOR OF THE MEAN OF A SMALL AREA

M. Mushfiqur Rashid and Balgobin Nandram

M. M. Rashid, Department of Mathematical Sciences, WPI, 100 Institute Road, Worcester, MA 01609

Key words: R-Predictor, Dispersion function, Mean-Squared Error, Nested Error Model

Abstract

Given data from a number of similar areas, we use a rank-based method to estimate the finite population mean of one of these areas. As the areas are similar, we use the nested error regression model which permits "borrowing of strength" from other areas. One important feature of our method is that there is no need to assess assumption of normality which is necessary to implement most methods for small areas. We use R-estimates of the model parameters to construct a predictor of the population mean of a small area, and construct an estimate of the mean squared error of the predictor. Finally, we illustrate the methodology using data obtained from a survey of agricultural areas.

1. Introduction

Small area estimation is extremely useful to government agencies that address the issues of distribution, equity, and disparity. As the field suggests, it is particularly appropriate in a setting that involves several areas with only a small sample available from each area. One parameter of interest is the finite population mean of one area and a procedure which borrows strength from other neighboring areas is usually used. Further, it requires normality of the responses, and therefore transformation might be needed to obtain normality; see for example Calvin and Sedransk (1991), Nandram and Sedransk (1993) and Nandram (1994). Rashid and Nandram (1994) developed a rank-based predictor of the finite population of the mean of a small area assuming a nested regression model. The objective of this article is to develop the mean-squared error of the rank-based predictor. Ghosh and Rao (1994) gave an excellent review of the recent literature on small area estimation.

In recent years model based estimation procedures for small area estimation have been widely addressed. The methods usually involve either an empirical Bayes (EB) approach or a full classical approach in which variance components are estimated. Both these procedures assume certain mixed models to perform prediction. Certain best linear unbiased predictors (BLUP) or EB predictors are obtained for the unknown parameters of interest assuming the variance components are known. Then the unknown variance components are estimated typically by Henderson's method of fitting constants or REML method, and the resulting estimated BLUP's are used for final prediction. Details are presented by Prasad and Rao (1990) and Battese, Harter and Fuller (1988). Although, the approaches are usually quite satisfactory for point prediction, it is very difficult to estimate the standard errors. Ghosh and Lahiri (1989) proposed a hierarchical Bayes (HB) procedure as an alternative to the estimated BLUP or the EB procedure; see also Datta and Ghosh (1991). The standard errors of the estimates, though complicated, can be obtained via numerical integration without any further approximation.

In this article we neither require any distributional assumptions nor estimation of variance components in predicting the mean of the small areas. However, estimate of variance components are required to obtain mean-squared error of the point predictor, called R-predictor. In Section 2 we present the model assumptions and a predictor of the finite population mean. In Section 3 we use R-estimators to construct the predictor and its mean-squared error. A numerical example is given in Section 4 and Section 5 has conclusions.

2. Model and Assumptions

We assume that data are available from c small areas, and the i^{th} area has M_i individuals, $i = 1, 2, \ldots, c$. Let Y_{ij} be the value of the j^{th} unit in the i^{th} area and let $\underline{Y}_i = [Y_{i1}, Y_{i2}, \ldots, Y_{iM_i}]'$ denote the vector of all values of all individuals from the i^{th} area. Let $\bar{Y}_i = M_i^{-1} \sum_{j=1}^{M_i} Y_{ij}$ be the finite population mean of the ith small area, $i = 1, 2, \ldots, c$.

Suppose we take a random sample of m_i individuals from the ith small area. Letting $f_i = m_i/M_i$ be the sampling fraction

for the i^{th} small area, observe that

$$\bar{Y}_i = f_i \bar{y}_i^{(s)} + (1 - f_i)\bar{Y}_i^{(ns)} \quad (2.1)$$

where $\bar{y}_i^{(s)} = (m_i)^{-1} \sum_{j=1}^{m_i} Y_{ij}$ and

$\bar{Y}_i^{(ns)} = (M_i - m_i)^{-1} \sum_{j=m_i+1}^{M_i} Y_{ij}$. Note that $\bar{y}_i^{(s)}$ is computed from the observed data and $\bar{Y}_i^{(ns)}$ is to be predicted.

Suppose there are $p-1$ covariates $x_{ij1}, x_{ij2}, \ldots, x_{ijM}$ associated with each individual. Following Battese, Harter and Fuller (1988) we consider the nested error regression model

$$Y_{ij} = \alpha_0 + \underline{x}_{ij}'\underline{\beta} + v_i + u_{ij}, j = 1, \ldots, M_i, i = 1, \ldots, c \quad (2.2)$$

where $\underline{x}_{ij} = (x_{ij1}, \ldots, x_{ij,p-1})'$, $\underline{\beta} = (\beta_1, \ldots, \beta_{p-1})'$ is the vector of regression parameters, $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2), u_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$, and the v_i are independent of the u_{ij}. Note that $\underline{\beta}$ is the same for all areas. In econometrics this model is known as an error component model and is widely used for combining cross-sectional and time series data. For a sample of size m_i from the ith area the model (2.2) holds and can be written as

$$Y_{ij} = \alpha_0 + \underline{x}_{ij}'\underline{\beta} + \epsilon_{ij}, j = 1, \ldots, m_i, i = 1, 2, \ldots, c \quad (2.3)$$

and in matrix notation as $\underline{y}_i = \alpha_0 \underline{1}_{m_i} + X_i\underline{\beta} + \underline{\epsilon}_i$, where $\underline{y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im_i})'$, $\underline{\epsilon}_i = [\epsilon_{i1}, \epsilon_{i2} \ldots, \epsilon_{im_i}]'$ with $\epsilon_{ij} = v_i + u_{ij}$ and X_i denoting the $m_i \times (p-1)$ matrix with rows $\underline{x}_{ij}', j = 1, \ldots, m_i$. First, we obtain the point predictor of $\bar{Y}_i$, the finite population mean of the i^{th} area. Let $\bar{x}_{i\cdot k}^{(s)} = m_i^{-1} \sum_{j=1}^{m_i} X_{ijk}$ and $\bar{X}_{i\cdot k}^{(ns)} = (M_i - m_i)^{-1} \sum_{j=m_i+1}^{M_i} X_{ijk}$ $i = 1, 2, \ldots, c$ and $k = 1, 2, \ldots, p-1$. Also let $\underline{\bar{x}}_i^{(s)} = (\bar{x}_{i\cdot 1}^{(s)}, \ldots, \bar{x}_{i\cdot p-1}^{(s)})'$ and $\underline{\bar{X}}_i^{(ns)} = (\bar{X}_{i\cdot 1}^{(ns)}, \ldots, (\bar{X}_{i\cdot p-1}^{(ns)})$. Assuming the model specification in (2.2)

$$E(\bar{Y}_i^{(ns)}) = \alpha_0 + \underline{\beta}'\underline{\bar{X}}_i^{(ns)}. \quad (2.4)$$

In parametric inference it is assumed that $\underline{\epsilon}_i \sim N(0, \sigma^2 V_i)$, where $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $V_i = [(1-\rho)I_{m_i} + \rho J_{m_i}]$, $\rho = \sigma_v^2/\sigma^2$ is the common intra-area correlation, I_{m_i} is the $m_i \times m_i$ identity matrix and J_{m_i} is the $m_i \times m_i$ unit matrix. In many situations in small area estimation, the assumption of multi-normality of $\underline{\epsilon}_i$ may not be valid. In such cases the survey statisticians may prefer an alternative non-parametric procedure. In this article we attempt a method which needs the following assumptions.

(1) The $\underline{\epsilon}_i$ $(i = 1, 2, \ldots, c)$ are independent and continuous random vectors, and the elements of $\underline{\epsilon}_i$ $(i = 1, 2, \ldots, c)$ are exchangeable random variables. Further, all the $\underline{\epsilon}_i$'s $(i = 1, 2, \ldots, c)$ have the same distributional form $F(., ., \ldots, .)$;

(2) Letting $\tau = \frac{1}{\gamma\sqrt{12}}$ where $\gamma = \int_{-\infty}^{\infty} f(\epsilon, \epsilon)\, d\epsilon\ (< \infty)$ and $f(.,.)$ is the bivariate p.d.f. of any two components of $\underline{\epsilon}_i$.

(3) $x_{ijk}^4 < B$, where B is finite and x_{ijk}'s do not depend on c; and

(4) $\text{Sup}_{i=1,\ldots,c} m_i = m < \infty$.

The assumptions (1)-(4) are all reasonable in small area estimation and they permit a rank-based approach. Our objective is, under assumptions (1) - (4) to obtain R-estimate $\hat{\beta}_k$ of the β_k $(k = 1, 2, \ldots p-1)$ which are used to construct a point predictor of the finite population mean of the ith area and its estimated mean-squared error.

Let $\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_{p-1})'$ be the partial regression coefficients corresponding to the model (2.3). Corresponding to the ith area we define a dispersion function as follows:

$$D_i(\underline{\beta}) = \sum_{j=1}^{m_i} [\frac{R(y_{ij} - \sum_{k=1}^{p-1} \beta_k x_{ijk})}{m_i + 1} - \frac{1}{2}][y_{ij} - \sum_{k=1}^{p-1} \beta_k\, x_{ijk}],$$

where $R(y_{ij} - \sum_{k=1}^{p-1} \beta_k x_{ijk})$ is the intra-area rank of the residual $y_{ij} - \sum_{k=1}^{p-1} \beta_k x_{ijk}$. The combined dispersion function for all small areas is

$$D(\underline{\beta}) = \sum_{i=1}^{c} D_i(\underline{\beta}).$$

Rashid and Nandram (1995) used an analogous dispersion function to make inference for one-way repeated measures designs with a changing covariate. Note that $D(\underline{\beta})$ is a location free measures, and is a linear function of the residuals with coefficients determined by the ranks of the residuals. Also $D(\underline{\beta})$ is a nonnegative, continuous and convex function of $\underline{\beta}$. Hence, it is expected that the R-estimates obtained by minimizing $D(\underline{\beta})$ will be more robust than least squares estimators since the the influence of the outliers have a linear, rather than a quadratic, effect. Thus, this approach will be an asset for small area practitioners. It is convenient that our predictor does not require an estimator of α_0 as inclusion or exclusion of α_0 in the dispersion function of model (2.3) does not alter the ranks of the residuals.

3. R-Predictor of the Finite Population Mean

First, we obtain the R-estimators of the parameters of model (2.3). Then we use these R-estimators to construct a R-predictor of the finite population mean and its estimated mean-squared error.

3.1 R-Estimation

Let $\hat{\underline{\beta}}$ minimize the dispersion function $D(\underline{\beta})$. See Jaeckel (1972) for further discussions on R-estimation. In order to develop the asymptotic distribution of $\hat{\underline{\beta}}$ we need the gradient of the dispersion function as well as its linear approximation.

As domain $\underline{\beta}$ of the space of $D(\cdot)$ consists of finite number of convex polygonal subsets, on each of which $D(\cdot)$ is a linear function of $\underline{\beta}$, the gradient vector exists and the negative of the gradient vector is

$$\underline{S}(\underline{\beta}) = [S_1(\underline{\beta}), S_2(\beta), \ldots, S_{p-1}(\beta)]', \tag{3.1}$$

everywhere, and its kth component is given by

$$S_k(\underline{\beta}) = \sqrt{12}\sum_{i=1}^{c}\sum_{j=1}^{m_i}[\frac{R(y_{ij} - \sum_{k=1}^{p-1}\beta_k x_{ijk})}{m_i+1} - \frac{1}{2}]x_{ijk}. \tag{3.1}$$

Note that

$$\underline{S}(\underline{\beta}) = \sum_{i=1}^{c}\underline{W}_i(\underline{\beta})$$

where

$$\underline{W}_i(\underline{\beta}) = [w_1^{(i)}, \ldots, w_{p-1}^{(i)}] \tag{3.3}$$

is contribution of the ith cluster in $\underline{S}(\underline{\beta})$, $\underline{W}_i(\underline{\beta})$, $(i = 1, 2, \ldots, c)$, are independent, and

$$w_l^i = \sqrt{12}[\sum_{j=1}^{m_i}\{\frac{R(y_{ij} - \sum_{k=1}^{p-1}\beta_k x_{ijk})}{m_i+1} - \frac{1}{2}\}$$

$, l = 1, 2, \ldots, p-1.$

Under assumption (1),

$$E[R_{ij}] = (m_i + 1)/2$$

and

$$\text{Cov}\,[R_{ij}, R_{i'j'}] = \begin{cases} (m_i^2 - 1)/12 & \text{if } i = i',\ j = j' \\ -(m_i+1)/12 & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases}$$

Thus, $E[\underline{W}_i(\underline{0})] = \underline{0}$ and

$$\text{Cov}[\underline{W}_i(\underline{0}] = \frac{m_i}{m_i+1}[\underline{x}_{i1}, \ldots, \underline{x}_{i\,p-1}]' H_i [\underline{x}_{i1}, \ldots, \underline{x}_{i\,p-1}]$$

$$= A_i = (a_{kk'}^{(i)}) \tag{3.4}$$

where

$$a_{kk'}^{(i)} = \begin{cases} \frac{m_i}{m_i+1}\sum_{j=1}^{m_i}(x_{ijk} - \bar{x}_{i.k})^2 & \text{if } k = k' \\ \frac{m_i}{m_i+1}\sum_{j=1}^{m_i}(x_{ijk} - \bar{x}_{i.k})(x_{ijk'} - \bar{x}_{i.'k}) & \text{if } k \neq k' \end{cases}$$

and $H_i = [I_{m_i \times m_i} - \frac{1}{m_i}1_{m_i}1'_{m_i}]$. Thus

$$E[\underline{S}(\underline{0})] = \underline{0}$$

and

$$\lim_{c\to\infty}\text{Cov}[\frac{1}{\sqrt{c}}\underline{S}(\underline{0})] = \lim_{c\to\infty}\frac{1}{c}\sum_{i=1}^{c}A_i = A = (a_{kk'}). \tag{3.5}$$

By using a multivariate central limit Theorem (See Rao 1973, p. 147 and Rashid and Nandram 1995),

$$\sqrt{c}[\frac{1}{c}\underline{S}(\underline{0})] \xrightarrow{\mathcal{D}} MVN(\underline{0}, A) \text{ as } c \to \infty$$

where A is given in (3.5).

It can be shown that

$$\hat{\underline{\beta}} = \underline{\beta} + \frac{\tau}{c}A^{-1}\underline{S}(\underline{\beta}) = o_p(\frac{1}{\sqrt{c}}) \tag{3.6}$$

where $\hat{\tau}$ of τ is a consistent estimate of $\hat{\tau}$ (to be determined).

Therefore using (3.6), $\underline{\beta}^0$ be the true value of $\underline{\beta}$,

$$\frac{\sqrt{c}}{\hat{\tau}}(\hat{\underline{\beta}} - \underline{\beta}^0) \xrightarrow{\mathcal{D}} \text{MVN}[\underline{0}, A^{-1}] \text{ as } c \to \infty. \tag{3.7}$$

That is, the R-estimators are asymptotically unbiased, consistent and asymptotically distribution free.

Even though the ranking of the residuals are done within each area separately, there is a borrowing strength from other areas since the R-estimates are obtained by minimizing the combined dispersion function.

3.2 R-Predictor

To obtain a R-predictor for $\bar{Y}_i^{(ns)}$ we substitute our rank-based estimators into (2.1). It follows that the R-predictor for the finite population mean of the i^{th} area is

$$\hat{\bar{y}}_i^{(R)} = \bar{y}_i^{(s)} + (1 - f_i)\hat{\underline{\beta}}'(\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)}). \tag{3.8}$$

We call the predictor in (3.8) a rank-based regression survey predictor. The estimator has an interpretable form. It is the sample mean adjusted by a fraction, the finite population correction, and the difference between $\bar{Y}^{(ns)}$ and $\bar{y}^{(s)}$ which is likely to be very small. It is also expected to be robust. One version of this estimator was considered by Sarndal (1984). Recall that the generalized least squares estimator and corresponding predictor under the model (2.3) are functions of σ_u^2 and σ_ν^2. Therefore, these variance components are unknown and have to be estimated, and at least in our case we do not need to estimate them for the point predictor.

However, in order to compute the standard error of $\hat{\bar{y}}_i^{(R)}$, we need to assume that the covariance matrix of $\underline{\epsilon}_j$ exists. As our model (2.3) assumes exchangeability of the errors within each cluster,

$$\text{Cov}(\underline{\epsilon}_i) = \sigma^2[(1-\rho)I_{m_i\times m_i} + \rho J_{m_i\times m_i}] \tag{3.9}$$

where $\rho = \frac{\sigma_\nu^2}{\sigma_\nu^2+\sigma_u^2}$.

We also need the covariances between the non-sampled observations and $\hat{\underline{\beta}}$, which can be obtained by using a large sample apprximation to $\hat{\underline{\beta}}$ defined by $\tilde{\underline{\beta}} = \underline{\beta}^0 + \frac{\tau}{c}A^{-1}\underline{S}(\underline{\beta}^0)$. Also, note that these covariances depend on the scale parameter τ.

3.3 Mean-Squared Error (MSE) of the R-Predictor

Using the survey regression estimator, the prediction error is

$$\hat{\bar{y}}_i^{(R)} - \bar{Y}_i = (1-f_i)(\bar{y}_i^{(s)} - \bar{Y}_i^{(ns)}) + (1-f_i)\hat{\underline{\beta}}'(\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)}). \tag{3.10}$$

It follows that

$$\text{Var}[\hat{\bar{y}}_i^{(R)} - \bar{Y}_i] = g_1^i(\sigma,\rho) + g_2^i(\tau) + g_3^i(\sigma,\tau,\rho) \tag{3.11}$$

where letting $\eta = \frac{\tau^2}{c}(1-f)^2$

$$g_1^{(i)}(\sigma,\rho) = (1-f)^2\sigma^2(1-\rho)/m_i, \tag{3.12}$$

$$g_2^{(i)}(\tau \approx \eta[\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)}]'A^{-1}[\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)}] \tag{3.13}$$

and

$$g_3^{(i)}(\sigma,\tau,\rho) = 2(1-f)^2\text{Cov}[\bar{y}_i^{(s)} - \bar{Y}_i^{(ns)}, \hat{\underline{\beta}}'(\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)})]. \tag{3.14}$$

Prasad and Rao (1990) has a similar representation for the mean-squared error of the predictor under nested error regression model, see also Tam (1995). Under model (2.3) taking $M_i^{-1}\sum_{j=1}^{M_i}\epsilon_{ij} = 0$,

$$\bar{y}_i^{(s)} - \bar{Y}_i^{(ns)} = \underline{\beta}'[\bar{\underline{x}}_i^{(s)} - \bar{\underline{X}}_i^{(ns)}] + \frac{1}{1-f}\bar{\epsilon}_i^{(s)} \tag{3.15}$$

where $\bar{\epsilon}_i^{(s)} = \frac{1}{m_i}\sum_{j=1}^{m_i}\epsilon_{ij}$.

Thus,

$$g_3^{(i)}(\sigma,\tau,\rho) \approx 2(1-f)\text{Cov}[\bar{\epsilon}_i^{(s)}, \hat{\underline{\beta}}'(\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)})]. \tag{3.16}$$

It can be shown that

$$g_3^{(i)}(\sigma,\rho,\tau) \approx \eta_1\{[\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)}]'A^{-1}A_iA^{-1}[\bar{\underline{X}}_i^{(ns)} - \bar{\underline{x}}_i^{(s)}]\}^{\frac{1}{2}}. \tag{3.16}$$

where $\eta_1 = 2(1-f)(\sigma\tau/c)[\{\frac{1+(m_i-1)\rho}{m_i}\}$. Note that $g_3^{(i)}(\sigma,\rho,\tau) \to 0$ as $c \to \infty$. Therefore for large c it follows that

$$\text{Var}[\hat{\bar{y}}_i^{(R)} - \bar{Y}_i] \approx g_1^{(i)}(\sigma,\rho) + g_2^{(i)}(\tau). \tag{3.17}$$

It is worth noting that $g_2^{(i)}(\tau)$ contains a term τ^2, whose role is similar to $\sigma^2(1-\rho)$ in normal theory based predictive inference.

If $\bar{x}^{(s)} = \bar{X}^{(ns)}$,

$$\text{Var}[\hat{\bar{y}}_i^{(R)} - \bar{Y}_i] \doteq g_1^{(i)}(\sigma^2,\rho), \tag{3.18}$$

which is the optimal variance.

As the mean squared error of the predictor is a function of σ^2, ρ and τ, to maintain the spirit of our assumptions simple robust estimates of σ^2, ρ and τ are required.

An estimator of σ_ν is

$$\hat{\sigma}_\nu = 1.483\text{Med}_i\{|\hat{\nu}_i - \text{Med}_i\hat{\nu}_i|\}.$$

We construct an estimator $\hat{\sigma}_u^2$ of σ_u^2 as follows: The rank-based fitted value for model (2.2) of y_{ij} is
$\hat{y}_{ij} = \sum_{k=1}^{p-1}\hat{\beta}_k x_{ijk} + \hat{\nu}_i$, where $\hat{\nu}_i$ is taken to be

$$\hat{\nu}_i = \text{med}\left[y_{ij} - \sum_{k=1}^{p-1}\hat{\beta}_k x_{ijk} : j = 1,2,\ldots,m_i\right] \tag{3.19}$$

and $\hat{\underline{\beta}}$ is obtained from (2.3). Then, we obtain residuals of the model (2.2) (using all data) as

$$\hat{u}_{ij} = y_{ij} - \hat{y}_{ij}.$$

This practice of computing the residuals $\hat{u}_{ij}$ is recommended in MINITAB reference manual (1991, p. 10-12) for inferences concerning block designs based intra-subject ranks. A robust estimate is given by the mean absolute deviation (MAD) of $\hat{u}_{ij}$ which is

$$\hat{\sigma}_u = (m_T - c)^{-1}\sum_{i=1}^{c}\sum_{j=1}^{m_i}|\hat{u}_{ij} - \text{Med}_{ij}(\hat{u}_{ij})|$$

where $m_T = \sum_{i=1}^{c} m_i$. An estimate of ρ is $\hat{\rho} = \hat{\sigma}_\nu^2/\hat{\sigma}^2$, where $\hat{\sigma}^2 = \hat{\sigma}_u^2 + \hat{\sigma}_\nu^2$.

A consistent estimator $\hat{\tau}$ of τ is given as follows.
For fixed j and j' $(j \neq j')$ let

$$d_1^{(j,j')} = \epsilon_{1j} - \epsilon_{1j'},\ d_2^{(j,j')} = \epsilon_{2j} - \epsilon_{2j'}\ldots,\ d_c^{(j,j')} = \epsilon_{cj} - \epsilon_{cj'}$$

be difference of residuals. We adopt the estimate of Koul et al(1987) to get a consistent estimate of γ.

Thus,

$$\hat{\gamma}^{(j,j')} = \frac{1}{2c^{3/2}t^{(j,j')}_{(c,.80)}} \sum_{i=1}^{c}\sum_{i'=1}^{c} I_{(|d_i^{(j,j')} - d_{i'}^{(j,j')}| \le t^{(j,j')}_{(c,\alpha)}/\sqrt{c})}$$

where $0 < \alpha < 1$, $t^{(j,j')}_{(c,\alpha)}$ is the .8th quantile (recommended by Koul et al 1987) of the distribution of $|d_i^{(j,j')} - d_i^{(j,j')}|$.

Finally, the estimated mean-squared error of the R-predictor of the *ith* area population mean is

$$\widehat{\mathrm{Var}}(\bar{y}_i^{(R)}) = \hat{g}_1^{(i)}(\sigma^2, \rho) + \hat{g}_2^{(i)}(\tau^2) + \hat{g}_3^{(i)}(\sigma, \tau, \rho) \qquad (3.20)$$

where $\hat{g}_1^{(i)}$, $\hat{g}_2^{(i)}$, and $\hat{g}_3^{(i)}$ are obtained, respectively, from $g_1^{(i)}(\sigma^2, \rho)$, $g_2^{(i)}(\tau^2)$, and $g_3^{(i)}(\sigma, \tau, \rho)$ by substituting $\hat{\rho}$, $\hat{\sigma}^2$ and $\hat{\tau}$ respectively in (3.20).

4. Example

We consider an example described by Battese, Harter and Fuller (1988) which concerns the estimation of areas under corn and soybeans for each 12 counties in North-Central Iowa using farm-interview data in conjunction with LANDSAT satellite data. Each county was divided into area segments, and the areas under corn and soybeans were ascertained for a sample of segments by interviewing farm operators. The number of segments in the sampled counties ranged from 1 to 6. The total number of segments in the different counties ranged from 394 and 965. (Thus, this is, indeed an small area problem.) Auxiliary data in the form of number of pixels (a term used for "picture elements" of about 0.45 hectares) classified as corn and soybeans were also obtained for all the area segments, including the sample segments, in each county using the LANDSAT satellite readings.

Battese, Harter and Fuller (1988) employed a "nested error regression" model involving random small area effects and the segment-level data and then obtained the estimates of county areas under corn and soybeans using the classical components of variance approach. They obtained the following fitted regression equations:

$$\text{Corn} : y_{ij} = 51 + 0.329x_{1j1} - 0.134\ x_{ij2}$$

$$\text{Soybeans} : y_{ij} = -16 + 0.028x_{1j1} + 0.494\ x_{ij2}.$$

Datta and Ghosh (1991) applied the HB approach to these data and obtained similar results.

We apply the method developed in this article to the same data. We obtain the R-estimates of the partial regression coefficients by minimizing the dispersion function using the Nelder-Mead algorithm. The rank-based estimated regression equations are

$$\text{Corn} : y_{ij} = 47.64 + 0.3154x_{1j1} - 0.1019\ x_{ij2}$$

$$\text{Soybeans} : y_{ij} = -41.13 + 0.1029x_{1j1} + 0.5036\ x_{ij2}.$$

We estimate α_0, the intercept of model (2.3) by

$$\hat{\alpha_0} = \text{med}(\hat{\nu}_i, i = 1, 2, \ldots, c).$$

The best and the R-predictors along with their estimated standard errors, square root of the mean-squared error, (in parentheses) are displayed in Table 1.

It is also possible to assess the importance of the parameters of the model (2.2). Note that $D(\underline{\beta})$ is used as a criterion for fitting a nested regression model to the data. $D(\underline{\hat{\beta}})$ represents the minimum distance, as measured by $D(\underline{\beta})$, from the data vector to the subspace spanned by the nested regression model. Suppose we want to test $H_0 : \beta_k = 0$ versus $\beta_k \neq 0$. Let $\hat{\beta}^H_{k'}$ $(k \neq k')$ be the the R-estimate of $\beta_{k'}$. Then under $\beta_k = 0$

$$D^* = 2[D(\hat{\beta}^H_1, \ldots, \beta^H_{k-1}, \beta^H_{k+1}, \ldots \beta^H_{p-1}) - D(\underline{\hat{\beta}})]/\hat{\tau}$$

has an asymptotic chi -square distribution with 1 degree of freedom. Using the above test we find that both covariates are significant in explaining the variation for both corn and soybean individually.

We have plotted (plots are not shown) studentized residuals of the model (2.2) with respect to the corresponding fitted values. The points in the plots do not show any patterns. It appears that for the corn data the second farm in Hardin county is an outlier. Also for the soybean data the second farm in Pocahontas county is an outlier. (The studentized residuals in the first case is -6.72, and for the second case is -5.32). Otherwise, the fitted models are consistent with the data.

5. Concluding Remarks

We obtained a rank-based method to estimate the finite population mean of a small area. Our R-estimates are easy to compute using the Nelder-Mead simplex algorithm. We showed that the rank-based estimates of the regression parameters have reasonable properties. Moreover, our method is very simple and easy to compute. Although we do not assume normality, the estimate of the finite population mean of each area is close to the BLUP of Battese, Harter and Fuller (1988). While our mean-squared errors are expected to be larger than those of BHF, we obtain comparable estimates. We expect that this work would stimulate research in nonparametric methods for small area estimation.

Table 1. A Comparison of the R-Predictor and Best Predictor of the Finite Population Mean

	Corn		Soybean	
County	Best	R	Best	R
Cerro Gordo	122.2	122.5	77.8	66.5
	(9.6)	(11.3)	(12.0)	(11.7)
Hamilton	126.3	126.0	94.8	104.5
	(9.5)	(11.3)	(11.8)	(11.7)
Worth	106.2	93.5	86.9	85.2
	(9.3)	(11.3)	(11.5)	(11.6)
Humboldt	108.0	106.8	79.7	65.4
	(8.1)	(9.1)	(9.7)	(9.5)
Franklin	145.0	149.7	65.2	59.6
	(6.5)	(6.8)	(7.6)	(7.1)
Pocahontas	12.6	114.4	113.8	116.3
	(6.6)	(7.5)	(7.7)	(7.9)
Winnebago	112.4	109.1	98.5	101.3
	(6.6)	(6.7)	(7.7)	(6.9)
Wright	122.1	123.9	112.8	111.2
	(6.7)	(8.1)	(7.8)	(8.5)
Webster	115.8	118.5	109.6	109.0
	(5.8)	(6.2)	(6.7)	(6.4)
Hancock	124.3	123.1	101.0	104.4
	(5.3)	(6.3)	(6.2)	(6.6)
Kossuth	106.3	104.2	119.9	121.8
	(5.2)	(5.4)	(6.1)	(5.6)
Hardin	143.6	143.8	74.9	78.1
	(5.7)	(7.9)	(6.6)	(8.3)

NOTE: Top entry: Mean; Bottom entry: Standard Deviation

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.

Calvin, J. and Sedransk, J. (1991), Bayesian and Frequentist Predictive Inference for Patterns of Care Studies, *Journal of the American Statistical Association, 86*, 36 -48.

Datta, G. and Ghosh, M. (1991), Bayesian Prediction in Linear Models: Applications to Small Area Estimation, *Annals of Statistics*, 19, 1748 - 1770.

Ghosh, M. and Rao, J. N. K. (1994), Small Area Estimation: An Appraisal, *Statistical Sciences, 9*, 55 - 93.

Ghosh, M. and Lahiri, P. (1987), Robust Empirical Bayes Estimation of Means of From Stratified Samples, *82*, Journal of the American Statistical Association, 82, 1153 - 1162

Jaeckel, L. (1972), Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals, *The Annals of Mathematical Statistics, 43*, 1449 - 1458.

Koul, H. L., Sievers, G. L. and McKean, J. W. (1987), An Estimator of the Scale Parameter for the Rank Analysis of Linear Models Under General Score Functions, *Scandinavian Journal of Statistics, 14*, 131-141.

McKean, J. W., Sheather, S. J. and Hettmansperger, T. P. (1990), Regression Diagnostics for Rank-Based Methods, *Journal of the American Statistical Association*, 85, 1018 - 1028.

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

Minitab Reference Manual (1991), Release 8, 10 - 12.

Nandram, B. and Sedransk, N (1993), Bayesian Predictive Inference for Longitudinal Sample Surveys, *Biometrics, 49, 1045 - 1055.*

Nandram, B. (1994), Bayesian Predictive Inference for Multivariate Sample Surveys, Journal of Official Statistics, 167 - 179.

Rashid, M. M. and Nandram, B. (1994), Rank-Based Predictive Inference For the Finite Population Mean of a Small Area, *Survey Methodology Section of ASA Proceedings*, 2, 1059 - 1064.

Rashid, M. M. and Nandram, B. (1995), A Rank Based Analysis of One-Way Repeated Measures Designs With a Changing Covariate, *Journal of Statistical Planning and Inference*, 46, 249 - 264.

Rao, C. R. (1973), Linear Statistical Inference and Its Applications, *John Wiley and Sons*, New York, 147.

Sarndal, C.E.(1984), Design-Consistent Versus Model-Dependent Estimation for Small Domains, *Journal of the American Statistical Association 79*, 624 - 631.

Tam, S. M. (1995), Optimal and Robust Strategies for Cluster Sampling, *Journal of the American Statistical Association 90*, 379 - 382.