

## On Small Area Prediction Interval Problems

Snigdhanu Chatterjee, Parthasarathi Lahiri, Huilin Li  
University of Minnesota, University of Maryland, University of Maryland

### Abstract

Empirical best linear unbiased prediction (EBLUP) method uses a linear mixed model in combining information from different sources of information. This method is particularly useful in small area problems. The variability of an EBLUP is measured by the mean squared prediction error (MSPE), and interval estimates are generally constructed using estimates of the MSPE. Such methods have shortcomings like undercoverage, excessive length and lack of interpretability. We propose a resampling driven approach, and obtain coverage accuracy of  $O(d^3 n^{-3/2})$ , where  $d$  is the number of parameters and  $n$  the number of observations. Simulation results demonstrate the superiority of our method over the existing ones.

**Keywords:** Predictive distribution, prediction interval, general linear mixed model, bootstrap, coverage accuracy.

### 1 Introduction

Large scale sample surveys are usually designed to produce reliable estimates of various characteristics of interest for large geographic areas. However, for effective planning of health, social and other services, and for apportioning government funds, there is a growing demand to produce similar estimates for smaller geographic areas and for other sub-populations. To meet this demand, it is necessary to supplement the survey data with other relevant information that are often obtained from different administrative and census records. In many small area applications, mixed linear models are now routinely used in combining information from various sources and explaining different sources of errors. These models incorporate area specific random effects which explain the “between small area variations”, not otherwise explained by the fixed effects part of the model.

For a good review on small area research, the readers are referred to the review papers by Rao (1986, 1999, 2001, 2004), Ghosh and Rao (1994), Pfeiffermann (2002), Marker (1999), and the book by Rao (2003), among others. Point prediction using the empirical best linear unbiased (EBLUP) and the associated mean square prediction error (MSPE) estimation have been discussed extensively in the small area literature. But little advancement has been made in interval prediction problems. Needless to say that interval prediction is a useful data analysis tool, since this integrates the concepts of both point prediction and hypothesis testing in an effective manner. Prediction intervals are useful in small area studies in many ways. For example, prediction intervals may help establish if different counties have similar resources and needs, or if different ethnic or other sub-population groups are equally exposed to a particular disease.

In the small area context, prediction intervals are often produced using the standard  $EBLUP \pm z_{\alpha/2} \sqrt{mspe}$  rule, where  $mspe$  is an estimate of the true  $MSPE$  of the EBLUP and  $z_{\alpha/2}$  is the upper  $100(1 - \alpha/2)$  point of the standard normal distribution. These prediction intervals are asymptotically correct, in the sense that the coverage probability converges to  $1 - \alpha$  for large sample size  $n$ . However, they are not efficient in the sense they have either under-coverage or over-coverage problem for small  $n$ , depending on the particular choice of the MSPE estimator. In statistical terms, the coverage error of such interval is of the order  $O(n^{-1})$ , which is not accurate enough for most applications of small area studies, many of which involve small  $n$ . If the naive  $mspe$ , i.e. the one that does not account for the uncertainties involving the model parameter estimation, is used, the resulting prediction interval suffers from under-coverage problem. On the other hand, if the Prasad-Rao (1990) second-order unbiased MSPE estimator is used, the interval suffers from the over-coverage and over-lengthening problem.

For a general mixed linear model, Jeske and Harville (1988) proposed a prediction interval for a mixed effect. But, their method does not address the complexity introduced by the estimation of the variance components. In particular, they did not study the effect of estimated unknown variance components on the accuracy of the coverage error of their proposed interval.

Jiang and Zhang (2002) used a distribution-free method for constructing prediction intervals for a future observation under a non-Gaussian linear mixed model, based on the theory developed by Jiang (1998). This technique does not employ any area specific information and can be useful in constructing intervals when there is no survey data on the response variable. Jiang and Zhang (2002) proposed another method which can be applied to the situation when the sample size is large within each area. This is a technique of first obtaining the EBLUP for the random effects and the residuals. Then, under conditions sufficient to imply that the number of times each random effect is repeated (ie, number of observations in each small area) tends to infinity, the empirical distribution of random effects as well as the residuals converge appropriately. This technique fails when we do not have large samples for each small area, a situation that is common in many small area applications.

Other than the papers cited above, research on small area prediction intervals is limited except for some special cases of the Fay-Herriot model (see Fay and Herriot 1979), a well-celebrated mixed regression model. Varieties of Bayesian and empirical Bayes methods are used in the Fay-Herriot model for such intervals. Because of the extensive use of the Fay-Herriot model and its particular cases in the small area interval estimation, in Section 2 we review different approaches to

interval estimation for this model.

In order to study the MSPE of EBLUP predictors of linear combinations of fixed and random effects, Das, Jiang and Rao (2004) considered the following general linear mixed model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v} + \mathbf{e}_n, \tag{1}$$

where  $\mathbf{Y} \in \mathbb{R}^n$  is a vector of observed responses,  $\mathbf{X}_{n \times p}$  and  $\mathbf{Z}_{n \times q}$  are known matrices and  $\mathbf{v}$  and  $\mathbf{e}_n$  are independent random variables with dispersion matrices  $D(\psi)$  and  $R_n(\psi)$  respectively. Here  $\beta \in \mathbb{R}^p$  and  $\psi \in \mathbb{R}^k$  are fixed parameters. The mixed ANOVA model, longitudinal models including the Fay-Herriot model and the nested error regression model are special cases of the above framework.

In this paper we address the problem of obtaining prediction intervals of a mixed effect for the above general linear mixed model. Our approach is to employ parametric bootstrap. This results in a higher order coverage accuracy of the interval compared to the existing methods, most of which are applicable only to special cases like the Fay-Herriot model. It has been long recognized that health, economic activity and other measures of human well-being depend on a number of exogenous and endogenous factors, many of which must be measured at the individual level and incorporated in the model. In statistical terms, this translates to high dimensionality of  $\beta$  and  $\psi$ . In order to address the dimension asymptotics aspect of small area prediction, we allow parameter dimension  $d = p + k$  to grow with sample size  $n$ , and obtain coverage accuracy of the order  $O(d^3 n^{-3/2})$ .

Traditional small area interval estimates achieve  $O(n^{-1})$  for the Fay-Herriot model or its particular cases, for fixed values of  $p$  and  $k$ . As stated earlier, the  $O(n^{-1})$  rate is typically not adequate for related follow up applications. Hence, calibrations and corrections of these intervals have been proposed, some of which are discussed in Section 2. For clarity of our proposed methodology and technical conditions, and also for easy use in future applications, in Subsection 2.3 we discuss our proposed prediction interval in detail for the Fay-Herriot model.

In Section 3, we present our prediction interval algorithm for the more general Das, Jiang, Rao model (1). In Section 4, our method is compared to two other existing methods using a Monte Carlo simulation study. The naive method, i.e. the one which uses the naive MSPE estimator, provides the shortest intervals among all the methods considered, but suffers from a severe under-coverage problem. The Prasad-Rao method provides an over-coverage at the expense of wider intervals. Our method is just about right in terms of coverage and is better than the Prasad-Rao method in terms of the length.

## 2 The Fay-Herriot model

In many practical applications, it is difficult to retrieve information at the respondent level for various reasons (including maintaining confidentiality) and researchers are provided only with design-based estimates and their associated estimated standard errors for the small-areas. In such cases, models are developed on the estimates themselves. In order to

produce per-capita income for small places (population less than 1,000), Fay and Herriot (1979) considered the following aggregate level model and used an empirical Bayes method which combines survey data from the U.S. Current Population Survey with various administrative and census data. Their model is now widely used in the small-area literature.

*The Fay-Herriot Model:*

1. Conditional on  $\theta = (\theta_1, \dots, \theta_n)^T$ ,  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$  follows a  $n$ -variate normal distribution with mean  $\theta$  and dispersion matrix  $D$  with diagonal entries  $D_{ii} = \sigma_i^2$  and off-diagonal entries 0. The sequence  $\{\sigma_i\}$  is a known sequence.

Here (and in the sequel) all vectors are taken to be column vectors, for any vector (matrix)  $a$  ( $A$ ), the notation  $a^T$  ( $A^T$ ) denotes its transpose.

2. The variable  $\theta$  follows a  $n$ -variate normal distribution with mean given by  $\mathbf{X}\beta$  for a known  $n \times p$  matrix  $X$  and unknown but fixed vector  $\beta \in \mathbb{R}^p$ . The dispersion matrix is given by  $\tau^2 \mathbf{I}_n$ , where the matrix  $\mathbf{I}_n$  is the  $n$  dimensional identity matrix and  $\tau$  is an unknown constant.

We are interested in obtaining a prediction interval for  $\theta_i = \mathbf{x}_i^T \beta + v_i$ . There are several options for constructing such intervals: one may use only the Level 1 model for the observed data, or only the Level 2 model for the borrowed strength component, or a combination of both.

The interval for  $\theta_i$  based only on the Level 1 model is given by  $I_i^D(\alpha) : Y_i \pm z_{\alpha/2} \sigma_i$ , where  $z_{\alpha/2}$  is the upper  $\alpha/2$  percent point of  $N(0, 1)$ . Obviously, for this interval, the coverage probability is  $1 - \alpha$ . However, it is not efficient, since its average length is too large to make any reasonable conclusion. This is due to the high variability of the point predictor  $Y_i$ .

An interval based only on Level 2 ignores the crucial area specific data that is modeled in Level 1, and hence falls short on two counts: it fails to be relevant to the specific small area under consideration, and it fails to achieve sufficient coverage accuracy. We show this latter property in a small example in Subsection 2.2.

Thus interval estimation techniques that combine both levels of the Fay-Herriot model are required. A popular approach is to employ empirical Bayes methodology, for example, by using empirical Bayes confidence intervals for prediction purposes. In the next subsection we review some of the important developments in empirical Bayes interval construction for the Fay-Herriot model or its particular cases.

### 2.1 Empirical Bayes intervals in the Fay-Herriot model

The posterior distribution of  $\theta$  given  $\mathbf{Y}$  is the  $n$ -variate normal distribution with the  $i^{th}$  mean vector  $(1 - B_i)Y_i + B_i \mathbf{x}_i^T \beta$ , where  $\mathbf{x}_i$  is the  $i^{th}$  row of  $\mathbf{X}$ . Here  $B_i = \sigma_i^2 / (\sigma_i^2 + \tau^2)$ . The posterior variance is the diagonal matrix whose  $i^{th}$  diagonal entry is  $\sigma_i^2 \tau^2 / (\sigma_i^2 + \tau^2)$ . Replacing  $\beta$  and  $\tau^2$  by their consistent estimators, we get the following empirical Bayes estimator of  $\theta_i$  :

$$\hat{\theta}_i^{EB} = (1 - \hat{B}_i)y_i + \hat{B}_i \mathbf{x}_i^T \hat{\beta},$$

where  $\hat{B}_i$  and  $\hat{\beta}$  are estimators of  $B_i$  and  $\beta$  respectively.

Cox (1975) initiated the idea of developing an empirical Bayes confidence interval. In the current context, his suggestion generates the following prediction interval:

$$I_i^C(\alpha) : \hat{\theta}_i^{EB} \pm z_{\alpha/2} \sigma_i (1 - \hat{B}_i)^{1/2}.$$

Under certain regularity conditions, it is easy to check that  $\mathbb{P}(\theta_i \in I_i^C(\alpha)) = 1 - \alpha + O(n^{-1})$ , where  $\mathbb{P}$  denotes a probability measure induced by the joint distribution of Level 1 and Level 2. Thus, this prediction interval attains the desired coverage probability asymptotically, but the coverage error is of the order  $O(n^{-1})$ , which is not accurate enough for many small area applications. Intuitively, this lack of accuracy could be due to the fact that the interval does not take into account the additional errors incurred by the estimation of the hyperparameters  $\beta$  and  $\tau^2$ .

For a special case of the Fay-Herriot model with common mean and equal sampling variances  $\sigma_i = \sigma$ , Morris (1983a) used a different measure of uncertainty for the empirical Bayes estimator, that incorporates the additional uncertainty due to the estimation of the hyperparameters. In a different work, Morris (1983b) considered a variation of his (1983a) empirical Bayes confidence interval where he used a hierarchical Bayes type point estimator in place of the previously used empirical Bayes estimator, and conjectured with some evidence that the coverage probability for his interval is at least  $1 - \alpha$ . He also noted that the coverage probability tends to  $1 - \alpha$  as  $n$  goes to  $\infty$  or  $\sigma$  goes to zero.

Basu, Ghosh and Mukerjee (2003) showed that the empirical Bayes interval proposed by Morris (1983a) does not improve upon Cox (1975), in the sense that the coverage error remains  $O(n^{-1})$ . However, using a Taylor series expansion of the coverage probability, they obtained an expression for the order  $O(n^{-1})$  term, which may then be used to calibrate Morris' interval to reduce the coverage error to  $o(n^{-1})$ . They also studied a prediction interval proposed by Carlin and Louis (1996, p.98), and showed that the Carlin-Louis interval has a coverage bias of the order  $o(n^{-1})$ . As a by product of this analytical result, they obtained an explicit expression for a new prediction interval with  $o(n^{-1})$  coverage bias and comparable expected length as the Carlin-Louis or the calibrated Morris intervals.

Using a multi-level model, Nandram (1999) obtained an empirical Bayes confidence interval for a small area mean and showed that asymptotically it converges to the nominal coverage probability. However, accuracy results and other asymptotic properties are not known for his interval.

Datta, Ghosh, Smith and Lahiri (2002) used an approach similar to Basu, Ghosh and Mukerjee (2003) in order to calibrate the Cox-type prediction interval for a more general Fay-Herriot model with covariates but with equal sampling variances. The coverage error for their interval is of the order  $O(n^{-3/2})$ . Hill (1990) suggested a general framework for constructing an empirical Bayes confidence interval conditional on some suitable ancillary statistic. In a simple normal-normal setting this matches with an exact hierarchical Bayes confidence interval. Datta *et al.* (2002) followed up Hill's idea

in proposing an empirical Bayes confidence interval which is correct up to  $O(n^{-1})$ , a property not noted by Hill (1990).

## 2.2 The use of bootstrap in empirical Bayes intervals

It can be seen that empirical Bayes intervals require corrections in order to achieve high coverage accuracy. These corrections may be based on asymptotic expansions as in Basu, Ghosh and Mukerjee (2003), hierarchical Bayesian approach as in Hill (1990) or Morris (1983b), and the bootstrap.

Different bootstrap methods have been used to improve the naive empirical Bayes confidence intervals. The methods differ in the generation of the bootstrap samples and the type of correction made. For a special case of the Fay-Herriot model where  $Y_1, \dots, Y_n$  are independent and identically distributed, Laird and Louis (1987) proposed three different ways of generating bootstrap samples. They differ in the degree of the parametric assumptions involved. One problem with the nonparametric and semi-parametric methods in the small area context is that the bootstrap approximation of the distribution of the EBLUP is generally not even consistent. The Laird-Louis Type III bootstrap is the usual parametric bootstrap.

Once the bootstrap sample (nonparametric, semi-parametric or parametric) is generated, the next challenge is to find a method that will correct the naive empirical Bayes confidence intervals  $I_i^C(\alpha)$  in an effort to achieve better coverage. The method proposed by Laird and Louis (1987) is to consider an imitation of the hierarchical Bayes approach. Conditional on the values of  $\beta$  and  $\psi$ , let  $g(\theta_i|\mathbf{Y}, \beta, \psi)$  be the posterior density of  $\theta_i$  given the data  $\mathbf{Y}$ . If a prior  $\pi(\cdot)$  is available for  $\beta$  and  $\psi$ , then in the hierarchical Bayesian model one might consider the distribution

$$\int g(\theta_i|\mathbf{Y}, \beta, \psi) \pi(\beta, \psi|\mathbf{Y}), \tag{2}$$

where both the terms in the above integral are appropriate posterior distributions. Laird and Louis (1987) proposed to use

$$B^{-1} \sum_{j=1}^B g(\theta_i|\mathbf{Y}, \beta_j^*, \psi_j^*), \tag{3}$$

where  $B$  is the bootstrap sample size and  $\beta_j^*$  and  $\psi_j^*$  are the estimates of  $\beta$  and  $\psi$  from the  $j^{th}$  bootstrap sample. The discrete mixture distribution (3) is a Monte Carlo approximation of

$$\int g(\theta_i|\mathbf{Y}, s, t) d\mathcal{L}^*(s, t), \tag{4}$$

where  $\mathcal{L}^*(\cdot, \cdot)$  is the bootstrap approximation of the sampling distribution of the parameter estimators  $\hat{\beta}$  and  $\hat{\psi}$ . Formula (4) is motivated by (2), but it is also similar to the *bootstrap prediction* of Harris (1989).

Carlin and Gelfand (1990, 1991) point out another issue with the hierarchical Bayesian approach of (2) and methods like those of Laird and Louis (1987). These approaches necessarily lead to a lengthening of the interval that is obtained by the naive empirical Bayes predictor  $g(\theta_i|\mathbf{Y}, \hat{\beta}, \hat{\psi})$ . However,

a correction to the naive interval is needed, which is not the same as *lengthening*. Carlin and Gelfand (1990) discuss an example where increasing the length further exacerbates the coverage bias. These authors suggest to use parametric bootstrap to calibrate the naive empirical Bayes interval.

Calibration of intervals has been one of the major uses of bootstrap for some time, and can lead to considerable improvement of coverage accuracy. Coupled with use of bias correction, use of pivotal or nearly pivotal statistics, and Edgeworth corrections, improvements from calibration can sometimes be dramatic. See Abramovitch and Singh (1985), Beran (1990a, 1990b) and the book by Efron and Tibshirani (1993) and references therein for further details on these issues. On the other hand, calibration is both time and computational effort intensive, often requiring iterative searches; it typically increases variability; the results often lack straightforward interpretability; and successive calibrating steps typically have diminishing returns in terms of improvement of coverage.

Hall (2006) suggests an application of the nonparametric bootstrap confidence interval based on the generated  $\theta_i^*$ 's only. In the small area context, this may be applicable when the differences between the small areas are minor, or carried only in the fixed effects. In surveys, robustness is always an important issue, and the practitioners are always interested in efficient nonparametric methods. However, due to scarce data at the small area level, nonparametric estimators tend to underperform, often severely. This is because the nonparametric models typically permit the generation of bootstrap histograms based on a synthetic model or the regression model, but do not permit approximation of the conditional distribution of  $\theta_i$  given the data  $\mathbf{Y}$ . As a result, the nonparametric bootstrap prediction interval for  $\theta_i$  is likely to underweight the area specific data. Accurate weighting the area specific data is important for achieving good coverage properties, as the example below shows. Hall (2006) also pointed out the importance of parametric bootstrap in small area estimation and other related problems.

EXAMPLE. Consider the following special case of the Fay-Herriot model where  $\sigma_i \equiv 1$ , and  $\mathbf{x}_i^T \beta \equiv \mu$ . Thus, at Level 1,  $Y_i$ 's given the  $\theta_i$ 's are independently distributed as  $N(\theta_i, 1)$  random variables; and at Level 2 the  $\theta_i$ 's are independent, identically distributed as  $N(\mu, \tau^2)$  random variables. The estimators of  $\mu$  and  $\tau^2$  are given respectively by  $\hat{\mu} = \bar{Y}$ ,  $\hat{\tau}^2 = \max(0, s^2 - 1)$ , where  $s^2 = \sum (y_i - \bar{y})^2 / (n - 1)$ . Assume  $\hat{\tau}^2 > 0$ , a condition that is satisfied in many problems. The bootstrap procedure would require us to generate  $\theta_i^* \stackrel{iid}{\sim} N[\hat{\mu}, \hat{\tau}^2]$  and  $Y_i^* | \theta_i^* \stackrel{iid}{\sim} N[\theta_i^*, 1]$ . Then we have  $\hat{\mu}^* = \bar{Y}^*$ ,  $\hat{\tau}^{2*} = \max(0, s^{*2} - 1)$  where  $s^{*2} = \sum (y_i^* - \bar{y}^*)^2 / (n - 1)$ .

An obvious Level 2 based bootstrap prediction interval for  $\theta_i$  that is *not* area specific, is given by

$$\left( \hat{\mu} - t_1 \sqrt{\hat{\tau}^2}, \hat{\mu} + t_2 \sqrt{\hat{\tau}^2} \right), \tag{5}$$

where  $(t_1, t_2)$  are cutoff points satisfying

$$\mathbb{P}(\hat{\mu}^* - t_1 \sqrt{\hat{\tau}^{2*}} \leq \theta^* \leq \hat{\mu}^* + t_2 \sqrt{\hat{\tau}^{2*}}) = 1 - \alpha.$$

It can be shown that interval (5) has coverage of  $1 - \alpha + O(n^{-1/2})$  which makes it consistent, but hardly accurate enough. The lack of accuracy is due to the use of the Level 2 distribution only, so that the Level 1 data  $Y_i$  plays no special role in the interval construction.  $\square$

### 2.3 A parametric bootstrap prediction interval

Instead of using the bootstrap for calibration or as a surrogate for prior, our approach is to obtain a prediction interval directly from the bootstrap histogram. We do not attempt a nonparametric or semiparametric bootstrap, owing to their obvious consistency problems in all but the extremely simplified special cases.

For the general Fay-Herriot model, we begin by defining the following quantities:

$$\hat{\beta} = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T \mathbf{Y}, \tag{6}$$

$$\hat{\tau}^2 = (\mathbf{r}^T \mathbf{r} - \sum (1 - h_{ii}) \sigma_i^2) / (n - p), \tag{7}$$

$$\hat{\tau}^{2'} = \max(\hat{\tau}^2, \epsilon), \tag{8}$$

$$\hat{B}_i = \sigma_i^2 / (\sigma_i^2 + \hat{\tau}^{2'}). \tag{9}$$

Here  $\epsilon > 0$  is a fixed small number,  $h_{ii}$  is the  $i^{th}$  diagonal element of the projection matrix on the column space of  $\mathbf{X}$ , and  $\mathbf{r} = \mathbf{Y} - \mathbf{X}\hat{\beta}$  is the vector of residuals. Under the condition  $\tau^2 > 2\epsilon$ , the truncated estimator  $\hat{\tau}^{2'}$  is different from the unbiased estimator  $\hat{\tau}^2$  on a set of negligible probability. We use it to avoid some messy algebra, but it is entirely dispensable otherwise. Other estimators, for example, the weighted least squares estimator for  $\beta$  can also be chosen.

The parametric bootstrap procedure is as follows:

1. Conditional on  $\tilde{\theta}^* = (\theta_1^*, \dots, \theta_n^*)^T$ ,  $\mathbf{Y}^* = (Y_1^*, \dots, Y_n^*)^T$  follows a  $n$ -variate normal distribution with mean  $\tilde{\theta}^*$  and dispersion matrix  $\sigma^2 \mathbf{I}_n$ .
2.  $\tilde{\theta}^*$  follows a  $n$ -variate normal distribution with mean  $\mathbf{X}\hat{\beta}$  and variance  $\hat{\tau}^{2'} \mathbf{I}_n$ .

The parametric bootstrap prediction interval for  $\theta_i$  is given by:

$$I_i(\alpha) : \hat{\theta}_i^{EB} \pm t_i \sigma_i (1 - \hat{B}_i)^{1/2}.$$

Recall that the formula for  $\hat{\theta}_i^{EB}$  is given in Subsection 2.1. The quantities  $\hat{\beta}^*$ ,  $\hat{\tau}^{2*}$ ,  $\hat{\tau}^{2'*}$  and  $\hat{B}_i^*$  are computed using (6)-(9) with the resample  $\mathbf{Y}^*$  in place of  $\mathbf{Y}$ . The value of  $t_i$  is such that

$$\mathbb{P}^* \left[ \theta_i^* \in \{ \hat{\theta}_i^{EB*} \pm t_i \sigma_i (1 - \hat{B}_i^*)^{1/2} \} \right] = 1 - \alpha, \tag{10}$$

where  $\hat{\theta}_i^{EB*} = (1 - \hat{B}_i^*) Y_i^* + \hat{B}_i^* \mathbf{x}_i^T \hat{\beta}^*$ . The probability  $\mathbb{P}^*$  is computed conditional on the data.

We assume the following conditions:

- (A1) the matrix  $\mathbf{X}$  is full column rank, and the projection matrix on the columns of  $\mathbf{X}$  (the so called "hat" matrix) given by  $\mathbf{P}_x = \mathbf{X}(\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T$  has diagonal entries  $h_{ii}$  satisfying

$$\sup_i h_{ii} = O(p/n). \tag{11}$$

(A2) The sequence  $\{\sigma_i\}$  lies in a compact subset of  $(0, \infty)$ .

(A3) For a known  $\epsilon > 0$ ,  $\tau^2 > 2\epsilon$ .

Assumptions (A1) and (A2) are routine, while (A3) is a minor restriction that essentially simplifies the algebra.

Our result for the parametric bootstrap prediction interval is as follows:

**Theorem 2.1** *Under assumptions (A1)-(A3) for the Fay Herriot model, the following holds:*

$$\begin{aligned} \mathbb{P} \left[ \theta_i \in \{ \theta_i^{EB} \pm t_i \sigma_i (1 - \hat{B}_i)^{1/2} \} \right] \\ = 1 - \alpha + O(p^3 n^{-3/2}). \end{aligned}$$

A proof of Theorem 2.1 can be found in Chatterjee and Lahiri (2002).

### 3 Parametric bootstrap prediction interval for a general linear mixed model

We consider the model:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v} + \mathbf{e}_n \quad (12)$$

where  $\mathbf{X}$  is a known  $(n \times p)$  matrix,  $\mathbf{Z}$  is a known  $(n \times q)$  matrix,  $\mathbf{Y} \in \mathbb{R}^n$  is the vector of observed data,  $\beta \in \mathbb{R}^p$  is a fixed but unknown parameter vector, and  $\mathbf{v} \in \mathbb{R}^q$  and  $\mathbf{e}_n \in \mathbb{R}^n$  are random variables following the normal distributions  $N_q(0, D)$  and  $N_n(0, R_n)$  respectively. Assume  $\mathbf{v}$  and the sequence  $\{\mathbf{e}_n\}$  are independent. The first term  $\mathbf{X}\beta$  represents the fixed effects, and the second term  $\mathbf{Z}\mathbf{v}$  the random effects. Thus  $\mathbf{X}\beta + \mathbf{Z}\mathbf{v}$  constitute the *signal* component of the observed data, while  $\mathbf{e}_n$  is the *noise*. The properties of the signal are of interest, which depend on the unknown parameters  $\beta$ ,  $D$  and  $R_n$ .

Assume that the  $(q \times q)$  matrix  $D$  and the  $(n \times n)$  matrix  $R_n$  are known up to a  $(k \times 1)$  vector of unknown parameters, thus  $D = D(\psi)$  and  $R_n = R_n(\psi)$  for a fixed but unknown  $\psi = (\psi_1, \dots, \psi_k)^T \in \mathbb{R}^k$ . Note that the dispersion matrix of the observed data  $\mathbf{Y}$  is given by

$$\Sigma_n = \Sigma_n(\psi) = R_n(\psi) + \mathbf{Z}D(\psi)\mathbf{Z}^T.$$

We henceforth drop the  $n$  from  $\mathbf{e}_n$ ,  $R_n$  and  $\Sigma_n$  to simplify notations. We take  $d = p + k$ , the dimension of the parameter space. Let  $\theta = (\beta, \psi)$  denote the unknown parameters. We allow  $q$  to be arbitrary, and possibly vary with  $n$ .

Das, Jiang and Rao (2004) show that several linear mixed models, including Analysis of Variance (ANOVA) models and longitudinal models of both balanced and unbalanced nature are special cases of the the model (12). Unbalanced ANOVA models arise, for example, when  $R_n = \sigma_0^2 \mathbf{I}_n$ ; and  $D = \text{diag}(\sigma_1^2 \mathbf{I}_{r_1}, \dots, \sigma_{k-1}^2 \mathbf{I}_{r_{k-1}})$  where  $\mathbf{I}_r$  is the  $r \times r$  identity matrix. Here  $\psi$  is the vector of variance components  $\psi = (\sigma_0^2, \dots, \sigma_{k-1}^2)$ . Unbalanced longitudinal models arise when  $\Sigma$  has a block diagonal structure.

Let  $T = c^T(\mathbf{X}\beta + \mathbf{Z}\mathbf{v})$ , where  $c$  is a fixed and known  $(n \times 1)$  vector. The conditional distribution of  $T$  given  $\mathbf{Y}$  is  $N(\mu_T, \sigma_T^2)$ , where

$$\begin{aligned} \mu_T &= c^T \mathbf{X}\beta + c^T \mathbf{Z}D\mathbf{Z}^T \Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta) \\ &= c^T R \Sigma^{-1} \mathbf{X}\beta + c^T \mathbf{Z}D\mathbf{Z}^T \Sigma^{-1} \mathbf{Y}, \quad \text{and} \quad (13) \end{aligned}$$

$$\sigma_T^2 = c^T \mathbf{Z} (D - D\mathbf{Z}^T \Sigma^{-1} \mathbf{Z}D) \mathbf{Z}^T c. \quad (14)$$

Generally,  $\beta$  and  $\psi$  (and hence  $D$  and  $R_n$ ) are estimated from the data  $\mathbf{Y}$  by using the marginal distribution of  $\mathbf{Y}$ , given by  $N_n(\mathbf{X}\beta, \Sigma)$ . The resulting estimates  $\hat{\mu}_T$  and  $\hat{\sigma}_T$  of the mean and variance of  $T$  are expressions similar to (13) and (14), with  $\hat{\beta}$  and  $\hat{\psi}$  in place of  $\beta$  and  $\psi$ .

For algebraic simplicity, in the rest of this paper we assume that  $\mathbf{X}$  is full column rank and use the estimator

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T \mathbf{Y} \\ &= \beta + (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T (\mathbf{Z}\mathbf{v} + \mathbf{e}). \end{aligned}$$

This is the *ordinary least squares* estimator of  $\beta$ . Using the *weighted least squares* estimator with appropriate conditions on the weights is another possibility that makes little difference in the asymptotic analysis. Estimator  $\hat{\psi}$  is typically obtained by maximum likelihood or restricted maximum likelihood techniques.

Based on the fact that  $\sigma_T^{-1}(T - \mu_T)$  is a standard Normal pivot, a naive approach to interval estimation for  $T$  is to take  $(\hat{\mu}_T \pm z\hat{\sigma}_T)$  for the appropriate Normal quantile  $z$ . Unfortunately,  $\hat{\sigma}_T^{-1}(T - \hat{\mu}_T)$  is not a pivot, and the naive approach produces too short intervals, that typically under-cover. Let the distribution of  $\hat{\sigma}_T^{-1}(T - \hat{\mu}_T)$  be  $\mathcal{L}_n$ . Recognizing that  $\mathcal{L}_n$  is not the standard Normal distribution, we propose to estimate it using parametric bootstrap.

The parametric bootstrap algorithm is based on the estimates  $\hat{\beta}$  and  $\hat{\psi}$  of  $\beta$  and  $\psi$  respectively. Define

$$\mathbf{Y}^* = \mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{v}^* + \mathbf{e}^*$$

where  $\mathbf{v}^* \sim N_q(0, D(\hat{\psi}))$  and  $\mathbf{e}^* \sim N_n(0, R(\hat{\psi}))$  are independent of each other. From  $\mathbf{Y}^*$ , obtain  $\hat{\beta}^*$  and  $\hat{\psi}^*$  using the same techniques used to obtain  $\hat{\beta}$  and  $\hat{\psi}$  earlier. Next, obtain  $\hat{\mu}_T^*$  and  $\hat{\sigma}_T^*$  using  $\hat{\beta}^*$  and  $\hat{\psi}^*$  using (13) and (14). Define  $T^* = c^T(\mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{v}^*)$ . The distribution of

$$\hat{\sigma}_T^{*-1}(T^* - \hat{\mu}_T^*),$$

conditional on the data  $\mathbf{Y}$ , is the parametric bootstrap approximation  $\mathcal{L}_n^*$  of  $\mathcal{L}_n$ . We then obtain the interval estimate for  $T$  as  $(\hat{\mu}_T - q_1 \hat{\sigma}_T, \hat{\mu}_T + q_2 \hat{\sigma}_T)$ , where  $q_1$  and  $q_2$  are appropriate quantiles of the bootstrap approximation  $\mathcal{L}_n^*$  of  $\mathcal{L}_n$ .

Our main result in this section is that  $\mathcal{L}_n^*$  approximates  $\mathcal{L}_n$  up to  $O(d^3 n^{-3/2})$  terms. In order to state the assumptions for our result, let us introduce some terminology and notation now. For any function  $f(\psi) : \mathbb{R}^a \rightarrow \mathbb{R}$ ,  $f'(\psi)$  denotes its first derivative written as a  $a \times 1$  column vector;  $f''(\psi)$  denotes the  $a \times a$  second derivative matrix. For a symmetric matrix  $A$ ,  $\lambda_{max}$  and  $\lambda_{min}$  respectively denote its maximum and minimum eigenvalue.

The following are the assumptions for our result in this section:

1. The following relations hold:

$$\|\mathbf{X}^T c\| = O(1), \quad (15)$$

$$\|\mathbf{X}^T \Sigma^{-1} \mathbf{Z} D \mathbf{Z}^T c\| = O(1), \quad (16)$$

$$c^T \mathbf{Z} D \mathbf{Z}^T c = O(1), \quad (17)$$

$$c^T \mathbf{Z} D \mathbf{Z}^T \Sigma^{-1} \mathbf{Z} D \mathbf{Z}^T c = O(1). \quad (18)$$

In addition,

$$\sigma_T^2 = c^T \mathbf{Z} (D - D \mathbf{Z}^T \Sigma^{-1} \mathbf{Z} D) \mathbf{Z}^T c > M > 0,$$

for some constant  $M > 0$ .

2. Assume that

$$\sup_{1 \leq i \leq n} \sum_{j=1}^p \left[ \sum_{a=1}^n X_{ja} \Sigma_{ai}^{1/2} \right]^2 = O(p/n), \quad (19)$$

$$\lambda_{\min} n^{-1} \mathbf{X}^T \mathbf{X} > M > 0, \quad (20)$$

for some constant  $M > 0$ .

3. The eigenvalues of the matrices  $D$  and  $R$  lie in  $(L^{-1}, L)$  for some  $L > 1$ . The eigenvalues of  $D(\hat{\psi})$  and  $R(\hat{\psi})$  lie in  $(L^{-1}/2, 2L)$ . The eigenvalues of  $\Sigma$  lie in a compact set on the positive half of the real line.

In the representations

$$D = D_1(\psi) D_1^T(\psi), \quad \hat{D} = D_1(\psi) \Lambda_D(\hat{\psi}) D_1^T(\psi), \quad (21)$$

$$R = R_1(\psi) R_1^T(\psi), \quad \hat{R} = R_1(\psi) \Lambda_R(\hat{\psi}) R_1^T(\psi), \quad (22)$$

where  $\Lambda_R$  and  $\Lambda_D$  are diagonal matrices, the following conditions are satisfied:

All the entries of the  $q \times q$  matrix  $\Lambda_D = \text{diag}(\Lambda_{D1}, \dots, \Lambda_{Dq})$  and the  $n \times n$  matrix  $\Lambda_R = \text{diag}(\Lambda_{R1}, \dots, \Lambda_{Rn})$  have three bounded continuous derivatives.

We denote by  $\Lambda'_D$  the  $k \times q$  matrix whose  $(j, i)^{th}$  entries are given by

$$\begin{aligned} ((\Lambda'_D))_{j,i}(\psi) &= \frac{\partial}{\partial \psi_j} \Lambda_{Di}(\psi), \\ j &= 1, \dots, k; \quad i = 1, \dots, q. \end{aligned}$$

The  $(j, i)^{th}$  entry of the  $k^2 \times q$  matrix  $\Lambda''_D$  is

$$\begin{aligned} ((\Lambda''_D))_{j,i}(\psi) &= \frac{\partial^2}{\partial \psi_{j_1} \partial \psi_{j_2}} \Lambda_{Di}(\psi), \\ j_1 + (j_2 - 1)k &= j, \quad j_1, j_2 = 1, \dots, k, \\ j &= 1, \dots, k^2, \quad i = 1, \dots, q. \end{aligned}$$

The  $(j, i)^{th}$  entry of the  $k^3 \times q$  matrix  $\Lambda^{(3)}_D$  is

$$\begin{aligned} \left( (\Lambda^{(3)}_D) \right)_{j,i}(\psi) &= \frac{\partial^3}{\partial \psi_{j_1} \partial \psi_{j_2} \partial \psi_{j_3}} \Lambda_{Di}(\psi), \\ j_1 + (j_2 - 1)k + (j_3 - 1)k^2 &= j, \\ j_1, j_2, j_3 &= 1, \dots, k, \quad j = 1, \dots, k^3, \quad i = 1, \dots, q. \end{aligned}$$

We define the  $k \times n$  matrix  $\Lambda'_R$ , the  $k^2 \times n$  matrix  $\Lambda''_R$  and the  $k^3 \times n$  matrix  $\Lambda^{(3)}_R$  along identical lines as above.

The following conditions are assumed:

$$\lambda_{\max} \Lambda_D^T(\psi) \Lambda'_D(\psi) = O(1), \quad (23)$$

$$\lambda_{\max} \Lambda_R^T(\psi) \Lambda'_R(\psi) = O(1), \quad (24)$$

$$\lambda_{\max} \Lambda_D^T(\psi) \Lambda''_D(\psi) = O(1), \quad (25)$$

$$\lambda_{\max} \Lambda_R^T(\psi) \Lambda''_R(\psi) = O(1), \quad (26)$$

$$\lambda_{\max} \Lambda_D^{(3)T}(\psi^*) \Lambda^{(3)}_D(\psi^*) < M = O(1), \quad (27)$$

$$\lambda_{\max} \Lambda_R^{(3)T}(\psi^*) \Lambda^{(3)}_R(\psi^*) < M = O(1), \quad (28)$$

for some constant  $M > 0$  for all  $\psi^*$  in a neighborhood of the true value  $\psi$ .

4. Let  $S = (k/n)^{1/2} (\hat{\psi} - \psi)$ . Assume that all the moments of  $\|S\|$  are  $O(1)$ . Moreover, the following relations are also satisfied:

$$\mathbb{E} S_j = O(\sqrt{k/n}), \quad (29)$$

$$\mathbb{E} S_a S_b = O(\sqrt{k/n}), \quad (30)$$

$$\mathbb{E} S_j (\mathbf{Z}\mathbf{v} + \mathbf{e})_i = O(\sqrt{k/n}), \quad (31)$$

$$\mathbb{E} S_a S_b (\mathbf{Z}\mathbf{v} + \mathbf{e})_i = O(\sqrt{k/n}), \quad (32)$$

We now state our main theorem for this section.

**Theorem 3.1** Fix  $\alpha \in (0, 1)$ , and let  $q_1$  and  $q_2$  be real numbers such that

$$\mathcal{L}_n^*(q_2) - \mathcal{L}_n^*(-q_1) = 1 - \alpha. \quad (33)$$

Under the Assumptions(1)-(4), if  $d^2/n \rightarrow 0$ , we have

$$\begin{aligned} \mathbb{P} [\hat{\mu}_T - q_1 \hat{\sigma}_T \leq T \leq \hat{\mu}_T + q_2 \hat{\sigma}_T] \\ = 1 - \alpha + O(d^3 n^{-3/2}). \end{aligned} \quad (34)$$

We now discuss the assumptions leading to Theorem 3.1, and some additional features of our result.

**REMARK 1.** Note that the dimension  $q$  of the random effect  $\mathbf{v}$  is arbitrary which may or may not depend on  $n$ . Owing to this generalization, our analysis is for  $T = c^T (\mathbf{X}\beta + \mathbf{Z}\mathbf{v})$ , rather than the more traditional  $\tilde{T} = c_1^T \beta + c_2^T \mathbf{v}$ . Since  $\mathbf{X}$  is full column rank, the fixed effects in  $T$  and  $\tilde{T}$  are equivalent.

**REMARK 2.** In the development of all the assumptions above, we have preferred simplicity over generality. The requirement  $d^2 n^{-1} \rightarrow 0$  is standard in dimension asymptotics. Assumption 1 is in order to ensure  $T$  as a non-trivial quantity, that is, it ensures that both the fixed component and the variance of the random component of  $\mu_T$  are  $O(1)$ , and the variance  $\sigma_T^2$  is bounded away from zero and infinity. By suitably scaling the norm of the vector  $c$  this assumption is satisfied.

Assumption 2 is a standard assumption on the behavior  $\mathbf{X}$ . It ensures that the norm of each fixed effects covariate is of suitable order, and the fixed effects design is not singular. This

assumption can be modified to suit cases where  $\mathbf{X}$  is not full column rank, but such generalizations are routine.

Assumption 3 is on standard differentiability and eigenvalue conditions. Here again, we have tried to adopt simple conditions rather than the most general ones. Note that the existence of the representations (21) and (22) are not part of the assumptions, and these representations will be established in the proof of Theorem 3.1.

Also note that the eigenvalues of  $D(\hat{\psi})$  and  $R(\hat{\psi})$  are estimates of the variance components in typical applications. Note that we do not allow these to be zero, since these must always lie in  $(L^{-1}/2, 2L)$ . However,  $L$  may be arbitrarily large, consequently this assumption does not limit the applicability of our results.

In Assumption 4 we take all moments of  $S$  to exist in order to achieve simplicity. Our result involves computation of several terms involving  $S$ , and having all the moments of  $S$  available simplifies the algebra. In most applications, both  $\psi$  and  $\hat{\psi}$  lie in a compact set, hence this is not a strong condition. The other moment conditions on  $S$  given by (29)-(30) are routine. These hold when  $\hat{\psi}$  is obtained using either maximum likelihood or restricted maximum likelihood formulation, see Jiang (1998) for related developments.

The conditions (31)-(32) are interesting, since they effectively set a limit to the amount of dependency structure we can have in  $\Sigma$ . In order to visualize this, suppose  $\hat{\psi}^{(-i)}$  is the estimator of  $\psi$  obtained by using only those observations that are independent of  $Y_i$ ; and let  $S^{(-i)} = (k/n)^{1/2}(\hat{\psi}^{(-i)} - \psi)$ . Then, a sufficient condition for  $\mathbb{E}S_j(\mathbf{Z}\mathbf{v} + \mathbf{e})_i = O((k/n)^{1/2})$  is that  $S - S^{(-i)} = O_P((k/n)^{1/2})$ .

This is routinely achieved, and in particular, if  $Y_i$  is independent of all but a finite number of observations, we have  $S - S^{(-i)} = O_P((k/n)^{1/2})$ . This is the typical situation is almost all applications of small area studies. Thus, the effect of Assumption 4 is to restrict the complexity of the matrices  $D$  and  $R$ .

REMARK 3. Unlike the Fay-Herriot model, in the general Das-Jiang-Rao mixed effects model,  $n$  does not represent the number of small areas. In fact, it is the sum total of all observations made, counting each repeated measurement on each individual unit in each small area as a distinct observation. This allows Theorem 3.1 to be used with considerable flexibility, for example, when number of individual units in small areas are large, or when number of small areas are large, or both. However, requirements of asymptotic negligibility, as in (31)-(32), must still be met. Our assumptions are designed for the more realistic applications where number of small areas are large.

#### 4 Simulation example based on income and poverty statistics

We illustrate the performance of our proposed parametric bootstrap approach with a simulated application based on a data set on small-area income and poverty estimates for the 50 states and the District of Columbia of the United States of America. We use a Fay-Herriot model, since in that model we have viable competitors for our approach. We use three

Table 1: Coverage and average length of different intervals (nominal coverage=0.95)

State	Naive	Prasad-Rao	Bootstrap
1	0.43(4.12)	0.99(12.96)	0.96(11.25)
2	0.45(4.13)	0.99(12.38)	0.96(10.61)
3	0.43(4.10)	0.99(13.07)	0.96(11.20)
4	0.47(3.94)	0.99(11.87)	0.97(9.61)
5	0.47(2.55)	0.99(16.28)	0.95(5.72)
6	0.46(3.77)	0.99(12.72)	0.97(9.15)
7	0.42(4.06)	0.98(13.76)	0.95(11.51)
8	0.31(4.64)	0.96(20.41)	0.95(19.97)
9	0.46(4.21)	0.99(12.04)	0.96(1.76)
10	0.47(3.03)	0.99(14.54)	0.95(7.00)

major covariates and sampling variances  $D_i$  from the data set; and take estimates of the hyperparameters  $\beta$  and  $\psi$  obtained by linear regression analysis on the data as the true values of these hyperparameters; and artificially generate the response variable  $\mathbf{Y}$  from a Normal distribution. Along with an intercept term, we use as covariates two measures related to tax return data, and one measure related to food stamp data. The performance of the proposed method is more interesting for small values of  $n$ . Hence we took  $n = 10$  states arbitrarily.

We compare our method with (a) the naive plug-in method, and (b) the Prasad-Rao method. The naive method uses asymptotic normality of  $T$  with variance approximated by the estimate of  $E(\theta_i - \theta_i^B)^2$ , where  $\theta_i^B$  is the Bayes estimator of  $\theta_i$ . In Prasad-Rao method, an improved Taylor series based approximation of the asymptotic variance is used.

Based on 10,000 iterations, for the three methods we compute the coverage probabilities and average lengths for all the 10 selected states (small areas). The results are reported in Table 1. The naive method has severe undercoverage problem, and it could be as low as 0.31 compared to the nominal value of 0.95. The Prasad-Rao method is more conservative than our proposed method. In terms of the average length, it is much more inefficient than our bootstrap method (our method can cut down the average length of the Prasad-Rao interval by more than one-third). This example illustrates that using the proposed bootstrap based interval achieves desired coverage accuracy, without undue lengthening of the interval.

#### References

[1] ABRAMOVITCH, L.; AND SINGH, K., (1985) Edgeworth corrected pivotal statistics and the bootstrap, *Ann. Statist.*, **13**, 116–132.

[2] BASU, R., GHOSH, J. K., AND MUKERJEE, R., (2003), Empirical Bayes prediction intervals in a normal regression model: higher order asymptotics, *Statist. Probab. Lett.*, **63**, 197–203.

[3] BERAN, R. (1990a) Refining bootstrap simultaneous confidence sets. *J. Amer. Statist. Assoc.* **85**, 417–426.

- [4] BERAN, R. (1990b) Calibrating prediction regions. Refining bootstrap simultaneous confidence sets. *J. Amer. Statist. Assoc.* **85**, 715–723.
- [5] CARLIN, B.P. AND LOUIS, T.A. (1996), *Bayes and empirical Bayes methods for data analysis*, Chapman and Hall, London.
- [6] CARLIN, B. AND GELFAND, A. (1990), Approaches for empirical Bayes confidence intervals, *J. Amer. Statist. Assoc.*, **85**, 105-114.
- [7] CARLIN, B. AND GELFAND, A. (1991), A sample reuse method for accurate parametric empirical Bayes confidence intervals, *J. Roy. Statist. Soc. Ser. B*, **53**, 189-200.
- [8] CHATTERJEE, S. AND LAHIRI, P., (2002), Parametric bootstrap confidence intervals in small area estimation problems, *unpublished manuscript*.
- [9] COX, D.R. (1975), Prediction intervals and empirical Bayes confidence intervals, in *Perspectives in Probability and Statistics*, papers in honor of M.S. Bartlett. Ed: J. Gani, Academic Press, 47-55.
- [10] DAS, K., JIANG, J., AND RAO, J. N. K., (2004), Mean squared error of empirical predictor, *Ann. Statist.*, **32**, 818-840.
- [11] DATTA, G.S., GHOSH, M., SMITH, D., AND LAHIRI, P. (2002), On an asymptotic theory of conditional and unconditional coverage probabilities of empirical Bayes confidence intervals, *Scand. J. Statist.* **29**, 139-152.
- [12] EFRON, B. AND TIBSHIRANI, R. J. (1993) *An introduction to the bootstrap*, Chapman and Hall, New York.
- [13] FAY, R. E. AND HERRIOT, R. A. (1979) Estimates of income for small places: an application of James-Stein procedure to census data, *J. Amer. Statist. Assoc.*, **74**, 269-277.
- [14] GHOSH, M., AND RAO, J.N.K. (1994), Small Area Estimation: An Appraisal, *Statist. Sci.*, **9**, No. 1, 55-93, (with discussions).
- [15] HALL, P. (2006) Discussion of *Mixed model prediction and small area estimation*, *Test*, **15**, 1-96.
- [16] HARRIS, I. R. (1989) Predictive fit for natural exponential families, *Biometrika*, **76**, 675-684.
- [17] HILL, J.R. (1990), A general framework for model-based statistics, *Biometrika*, **77**, 115-126.
- [18] JESKE, D. R. AND HARVILLE, D. A. (1988), Prediction-interval procedures and (fixed-effects) confidence-interval procedures for mixed linear models, *Commun. Statist. - Theory Meth.* **17**, 1053-1087.
- [19] JIANG, J. (1998), Asymptotic Properties of the Empirical BLUP and BLUE in mixed linear models, *J. Amer. Statist. Assoc.*, **93**, 720-729.
- [20] JIANG, J. AND ZHANG, W. (2002), Distribution-free prediction intervals in mixed linear models, *Statistica Sinica* **12**, 537-553.
- [21] LAIRD, N. M. AND LOUIS, T. A. (1987) Empirical Bayes confidence intervals based on bootstrap samples (with discussion), *J. Amer. Statist. Assoc.*, **82**, 739-750.
- [22] MARKER, D.A. (1999), Organization of small area estimators using generalized regression framework, *J. Official Statist.* **15**, 1-24.
- [23] MORRIS, C.N., (1983a) Parametric empirical Bayes inference: theory and applications (with discussion), *J. Amer. Statist. Assoc.*, **78**, 47-65.
- [24] MORRIS, C.N., (1983b) Parametric empirical Bayes confidence intervals, in *Scientific Inference, Data Analysis and Robustness*, Academic Press, New York, 25-50.
- [25] NANDRAM, B., (1999), An empirical Bayes prediction interval for the finite population mean of a small area, *Statist. Sinica*, **9**, 325-343.
- [26] PFEFFERMANN, D. (2002), Small area estimation - new developments and directions, *Int. Statist. Rev.* **70**, 125-143.
- [27] PRASAD, N. G. N. AND RAO, J. N. K., (1990), The estimation of the mean squared error of small-area estimators, *J. Amer. Statist. Assoc.*, **85**, 163-171.
- [28] RAO, C. R., (1963), *Linear statistical inference and its applications*, Wiley Eastern, New Delhi.
- [29] RAO, J. N. K. (1986), Synthetic estimators, SPREE and the best model based predictors, *Proceedings of the Conference on Survey Methods in Agriculture*, 1-16, U.S. Dept. of Agriculture. Washington, D.C.
- [30] RAO, J. N. K., (1999) Some recent advances in model-based small area estimation, *Surv. Meth.* **25**, 175-186.
- [31] RAO, J. N. K., (2001), EB and EBLUP in small area estimation, in *Empirical Bayes and Likelihood inference, Lecture notes in Statistics No. 148* (ed. Ahmed, S. E. and Reid, N.), Springer, New York.
- [32] RAO, J. N. K., (2003), *Small Area Estimation*, Wiley, New York.
- [33] RAO, J. N. K., (2004), Small area estimation with applications to agriculture, *J. Indian Soc. Agricultural Statist.*, **57**, 159-170.