A Simple Computational Method for Estimating Mean Squared Prediction Error in General Small-Area Model

Snigdhansu Chatterjee and P. Lahiri
University of Minnesota and University of Maryland, College Park

Abstract

The basic requirements of second-order unbiasedness and non-negativity of the mean squared prediction error (MSPE) of an empirical best predictor (EBP) have led to different complex analytical adjustments to the naive parametric bootstrap technique for small area estimation. In this paper, we show a way to recover the basic simplicity in the parametric bootstrap method, i.e. replacement of laborious analytical calculations by computer-oriented simple techniques, without sacrificing the basic requirements in an MSPE estimator. The method works for a general class of mixed models and different techniques of parameter estimation.

Keywords: Bootstrap; Mean Squared Prediction Errors; General Linear Mixed Model.

1 Introduction

The use of an *empirical best prediction* (EBP) method is now common in solving a variety of small-area problems dealing with both discrete and continuous data. An EBP typically uses two or more levels of modeling to combine information from different relevant sources and to account for different sources of errors. The readers are referred to Rao (2003) and Jiang and Lahiri (2006) for a comprehensive review on small area estimation and EBP methods. Computation of an EBP is reasonably straightforward and is now available in many standard packages. However, a reliable *mean squared prediction error* (MSPE) estimation of an EBP that accounts for all sources of variation is a highly non-trivial problem. In a pioneering paper, Prasad and Rao (1990) proposed a Taylor linearization method to obtain a second-order (or nearly) unbiased estimator of the MSPE of an *empirical best linear unbiased predictor* (EBLUP) when the variance components are estimated by a simple method-of-moments. Following the work of Prasad and Rao (1990), several methods, which include methods based on higher-order asymptotic expansions, computationally oriented methods, and hybrid methods, have been proposed to obtain second-order unbiased MSPE estimators of an EBP. In recent years, computationally oriented methods, especially the parametric bootstrap method, have received considerable attention due to their simplicity and flexibility in solving real life problems with complex modeling.

An early application of the parametric bootstrap method to obtain a second-order unbiased MSPE estimators can be found in Butar (1997). The parametric bootstrap method has been pursued in different directions by a number of researchers, including Booth and Hobert (1998), Butar and Lahiri (2003), Pfeffermann and Glickman (2004), Pfeffermann and Tiller (2005), Hall and Maiti (2006a,b), and others. These methods rely on certain MSPE decomposition formulae, like the Kackar-Harville identity (see Kackar and Harville, 1984), and/or complex multi-stage corrections, which often use asymptotic approximations. As a result, the main attraction of the bootstrap method, as originally envisioned by Efron (1979), i.e. the replacement of an "old-fashioned" analytical approach by a user-friendly computer-oriented method, is lost. The basic dual requirements of second-order unbiasedness and non-negativity of the MSPE estimator necessitate such complex approaches.

A completely non-parametric bootstrap method for the MSPE estimation has not been very widespread in small area estimation. This is mainly because a naive resampling method that uses a simple random sample from the data fails to capture the complex dependency structures present in a typical small area model. Laird and Louis (1987) suggested a nonparametric bootstrap method for a simple random effects model to approximate a hierarchical Bayes solution; however, they did not use their nonparametric bootstrap to obtain a second-order MSPE estimator. Recently, Pfeffermann and Tiller (2005) and Hall and Maiti (2006a) pursued certain promising nonparametric bootstrap approaches. For a state-space model fitting using the Kalman filter, Pfeffermann and Tiller (2005) derived a useful representation of the model, which they used for a non-parametric bootstrap MSPE estimation. In the context of a nested error regression model, Hall and Maiti (2006a) noted that the MSPE is essentially a function of the first few moments of the sampling and random effects distributions, and used a moment matching approach to obtain a non-parametric bootstrap MSPE estimator. In small area studies and several other fields of application, reasonable multi-level parametric models are generally available and trustworthy; hence the need for completely nonparametric approaches is not dire. However, small area studies often require models with non-normal (but known) probability densities for the sampling and the random effects distributions that can adequately capture complex correlation structures in the data.

The traditional unconditional MSPE estimator averages over the marginal distribution of the study variable(s) and hence does not distinguish different areas even though the small area model does not hold equally well for different areas. In the small area context, the importance of using a suitable *conditional* mean squared prediction error that is area specific with respect to the study variable cannot be overstated. Fuller (1990) and Booth and Hobert (1998) brought out this aspect quite elegantly and advocated for a conditional mean squared prediction error.

In the context of the Small Area Poverty and Income Estimation (SAIPE) project of the U.S. Census Bureau, Bell (1999) noticed that the jackknife estimator of Jiang, Lahiri and Wan (2002) could produce negative estimates of MSPE of EBLUP. In fact, this problem can be observed in some Taylor linearization and bootstrap methods as well. This situation, however, is not very frequent; and occurs when the bias-correction term that appears in a second-order MSPE formula produces unduly large contribution to the MSPE formula. Chen and Lahiri (2002, 2007) and Hall and Maiti (2006a,b) addressed this issue.

The effect of estimation of high dimensional hyperparameters is relatively less studied in small area estimation problems with the exception of Chatterjee, Lahiri and Li (2007), although Jiang (1996) considered a similar problem in the linear mixed model context. Since the sample size for a small area is typically small, the effect of estimating a finite, but high dimensional, parameter can be quite significant in terms of the prediction accuracy. This is not reflected in the standard asymptotic setting that treats the hyper-parameter dimension as fixed. A better approach is to consider an asymptotic framework that assumes the parameter dimension to be a function of the sample size.

In this paper, we assume a *fully parametric model* (FPM), where the probability distributions at the both levels of a two-level model are known, up to finite dimensional parameters. The FPM covers most hierarchal models studied in the small area estimation, including linear mixed models (LMM), longitudinal models, analysis of variance models, generalized linear mixed models (GLMM), hierarchical generalized linear models (HGLM), state-space models of certain time series, kreiging models for spatial processes, nested error regression models, and point process models.

We obtain second-order unbiased, non-negative, conditional and unconditional MSPE estimators, and allow the hyperparameter dimension to grow with the sample size, thus bringing in the dimension asymptotic effect of estimating the hyperparameters. We retain the basic simplicity of the bootstrap methodology by replacing laborious analytical calculations by computer-oriented simple techniques, without sacrificing the theoretical properties of the MSPE estimators. We use a double bootstrap strategy as in Booth and Hobert (1998) and Hall and Maiti (2006b). However, apart from being applicable to a much broader collection of problems, our methodology is not driven by stepwise calibration ideas, and we have a one-step (conditional or unconditional) MSPE estimator. In each scenario, our resampling technique and the MSPE estimate formula is exactly the same for all situations; we do not require problem specific corrections.

The outline of the paper is as follows. In section 2, we state our model and discuss different special cases that have been considered in the small area literature. We also introduce an EBP in this section. In section 3, we propose our conditional and unconditional parametric bootstrap MSPE estimators of the proposed EBP, and discuss their properties. We present results from a Monte Carlo simulation study in section 4. To save space, we omit the proof of our main result.

2 A General Fully Parametric Model

In this section, we describe a general fully parametric model (FPM) and the associated EBP. Let Y_i be the vector of observations for the i^{th} area ($i = 1, \cdots, n$), n being the total number of small areas. The dimension of Y_i can be arbitrary, and may or may not depend on n and i. Each element of Y_i is an observation, though all observations are not necessarily independent. If n_i is the dimension of Y_i, we define $N = \sum_{i=1}^{n} n_i$, the total number of observations.

We propose the following two-level model:

$$\begin{array}{ll} \text{Level 1:} & Y_i|\theta_i \sim f_i(\cdot;\theta_i,\xi), \quad i=1,\ldots,n; \\ \text{Level 2:} & \theta_i \sim g_i(\cdot;\xi), \quad i=1,\ldots,n. \end{array}$$

We assume $\xi \in \Xi \subseteq \mathbb{R}^d$, where the parameter space Ξ is an open set in $\mathbb{R}^d$. Without loss of generality, we shall assume that both Level 1 and Level 2 can be described by a common vector of parameters ξ (if necessary, we simply define ξ to be the super-collection of all Level 1 and Level 2 unknown parameters.) Similarly, there is no loss of generality in considering the same parameters ξ appearing in all the distributions irrespective of i. In many applications, ξ is naturally split in two parts, with one part contributing to the mean of the Y_i's or θ_i's (the *location* parameters), while others (the *variance components*) contribute to the variances and covariances. Our set-up unifies the two components, thus allowing for very general models with multiple sources of heteroscedasticity, non-linear structures and dependencies.

The dimension of θ_i is arbitrary and can also depend on n and i. Although in the rest of this paper we concentrate on the case where θ_i is a scalar, the extension to the multi-dimensional case is obvious. We make occasional remarks to elucidate some of the technical points that arise for the MSPE estimation when θ_i is multi-dimensional.

We assume that the pairs $\{(Y_i, \theta_i),\ i = 1, \ldots, n\}$ are independent. Since there is no restriction on the dimension of Y_i's, the independence of the observations across different areas is not a strong assumption in most small area applications, and is a matter of nomenclature. However, our assumption rules out spatial models that assume correlations across the areas. The above model covers the following important models:

1. **The Fay-Herriot Model:** In order to estimate the per-capita income of small places (population less than 1000), Fay and Herriot (1979) used the following two-level model:

$$\text{Level 1 (sampling model): } Y_i|\theta_i \stackrel{\text{ind}}{\sim} N(\theta_i, D_i),$$
$$\text{Level 2 (linking model): } \theta_i \stackrel{\text{ind}}{\sim} N(x_i^T\beta, A),$$

where $i = 1, \cdots, n$. Here "ind" stands for independent random variables, and $N(\mu, \sigma^2)$ denotes the Normal distribution with mean μ and variance σ^2. In the Fay-Herriot model, Level 1 is used to account for the sampling variability of the regular survey estimates Y_i of true small area means. Level 2 links the true small area means θ_i to a vector of p known auxiliary variables x_i, often obtained from various administrative and census records. The parameters $\xi = (\beta\ A)$ of the linking model are generally unknown and are estimated from the available data. The sampling variabilities D_i are assumed to be known. In practice, D_i's are estimated using the generalized variance function (see Wolter 1985, Chapter 5) method that uses some external information from the survey.

For the Fay-Herriot model, the dimension of Y_i is 1 and so does not depend on n or i. Here $\xi = (\beta, A)^T$ so that $d = p + 1$. The parameter ξ is involved in Level 2 but not in Level 1. Several MSPE estimation methods, for example, Datta, Rao, Smith (2005), Pfeffermann and Glickman (2004), Chen, Lahiri, Rao (2007) are exclusively devised for the Fay-Herriot model.

2. **The Longitudinal Linear Mixed Normal Model:** This covers a wide variety of mixed models, including the Fay-Herriot model, nested error regression model (Battese et al. 1988), multi-level models (Moura and Holt, 1999) and time series model (Rao and Yu, 1994; Datta, Lahiri and Maiti, 2002; Pfeffermann and Tiller, 2005). Let X_i and Z_i be $n_i \times p$ and $n_i \times k_i$ matrices of known constants. Let $N = \sum_{i=1}^n n_i$ and $K = \sum_{i=1}^m k_i$. Then the model is described as follows:

$$\text{Level 1:} Y_i \mid U_i \stackrel{\text{ind}}{\sim} N_{n_i}(X_i\beta + Z_iU_i, R_i),$$
$$\text{Level 2: } U_i \stackrel{ind}{\sim} N_{k_i}(0,\ G_i),$$

$i = 1, \ldots, n$, where β is a $p \times 1$ column vector of unknown regression coefficients; $R_i = R_i(\psi)$ and $G_i = G_i(\psi)$ are respectively $n_i \times n_i$ and $k_i \times k_i$ matrices which possibly depend on ψ, a $s \times 1$ vector of unknown variance components. In this case $\xi = (\beta, \psi)'$ so that $d = p + s$. The dimension of Y_i depends on n and i. Butar and Lahiri (2003) (see also Butar, 1997) developed a parametric bootstrap for this model.

3. **The Das-Jiang-Rao (2005) Model:** This model includes the mixed ANOVA and the longitudinal models as special cases and can be described as

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

where $\mathbf{Y}_n \in \mathbb{R}^n$ is a vector of observed responses; $\mathbf{X}_{n\times p}$ and $\mathbf{Z}_{n\times q}$ are known matrices; $\mathbf{v}_q$ and $\mathbf{e}_n$ are independent random variables with dispersion matrices $D_q(\psi)$ and $R_n(\psi)$ respectively. Here $\beta \in \mathbb{R}^p$ and $\psi \in \mathbb{R}^k$ are fixed parameters.

4. **A conditionally independent two-level model**

Hall and Maiti (2006b) developed their parametric bootstrap method for the following model:

$$\text{Level 1: } Y_{ij}|\theta_i \stackrel{\text{ind}}{\sim} f(\cdot; \psi(\theta_i), \eta_i),$$
$$\text{Level 2: } \theta_i \sim g(\cdot; h_i(\beta), \xi),$$

$j = 1, \cdots, n_i,\ i = 1, \cdots, n$, where $h_i(\beta)$ is a known mean function involving known and fixed covariates $X_i = (X_{i1}, \ldots, X_{in_i})$ and $\psi(\cdot)$ is a known link function; f and g represent two densities not necessarily normal. The model allows for nonlinear and non-normal mixed effects models. However, their model has no dependency structure in Level 1. Special cases of this model include certain cases of generalized linear mixed models, including the Fay-Herriot and the nested error regression model.

5. **Multivariate unbalanced generalized linear mixed models**

This model has the following structure:

$$\text{Level 1: } Y_i|\theta_i \sim f(\cdot; \psi(\theta_i), \xi),$$
$$\text{Level 2: } \theta_i \sim g(\cdot; \mathbf{x}_i^T\beta, \sigma_i^2),$$

$i = 1, \cdots, n$, where the density $g(\cdot)$ is a Normal density with mean $\mathbf{x}_i^T\beta$ and variance determined by σ_i^2. The Level 1 density or mass function $f(\cdot)$ is supported in $\mathbb{R}^{n_i}$, and depends on θ_i through the function $\psi(\theta_i)$. Examples of such models are multivariate logistic-Normal, multivariate Poisson-Normal and multivariate negative binomial-Normal models. A typical logistic-Normal model would be as follows:

The observed data from the i^{th} small area is $Y_i = (Y_{i1}, \ldots, Y_{in_i})$, where Y_{ij} follows Binomial (m_i, p_i) with known m_i's, and $\log(p_i/(1-p_i))$'s have a Normal distribution $N(\mathbf{x}_i^T\beta, \sigma_i^2)$. Here β and $\sigma_{n1}^2, \ldots, \sigma_{nn}^2$ are the parameters. One multivariate generalization of this model useful in applications where sampling stage (perhaps spatial) dependency of Y_{nij}'s across j need to be captured is given by Y_i following a Multinomial (m_i, p_i) distribution with known m_i's, and a probability vector $p_i = (p_{i1}, \ldots, p_{in_i})$. Here $U = (\log(p_{i1}/(1 - p_{i1})), \ldots, \log(p_{in_i}/(1 - p_{in_i})))$ follows a n_i-variate Normal distribution with mean $\mathbf{X}_i\beta$ and dispersion matrix $\Sigma(\psi)$, which depends on an unknown hyperparameter ψ.

6. **Multivariate unbalanced hierarchical generalized linear models**

   Same as above, except $G(\cdot)$ is conjugate, instead of being Normal. This model has been studied extensively by Lee and Nelder (1996), and others.

We write $\mathbf{Y}_n = (Y_{n1} \dots Y_{nn})$ to denote all the observed data. The marginal distribution of Y_i is written as

$$m_i(\cdot;\xi) = \int f_i(\cdot;t,\xi) g_i(t;\xi) dt.$$

Since we assume independence, the likelihood of ξ can be written as

$$L(\xi;\mathbf{Y}_n) = \prod_{i=1}^{n} m_i(Y_i;\xi).$$

We assume that the estimator $\hat{\xi}$ of ξ is obtained by minimizing the contrast function

$$\sum \Psi_i(\xi;Y_i).$$

This covers the classical maximum likelihood (ML) and residual maximum likelihood (REML) estimator (Jiang 1996), various method-of-moments estimators as developed in Fay and Herriot (1979), Pfefferman and Nathan (1981), Prasad and Rao (1990), Jiang (1998); decision theoretic estimators like the Bayesian or minimax estimators; and robust or otherwise estimators obtained by solving appropriate estimating equations or minimizing functions. We assume the functions $\Psi_i(\cdot;Y_i)$ have sufficient conditions to ensure that $\hat{\xi} = \xi + dn^{-1/2}T_n$, where $T_n = O_P(1)$ is a random variate that converges weakly to a Normal distribution and has sufficient moments. This is accomplished, for example, when $\Psi_i(\cdot;Y_i)$ has three continuous derivatives with respect to ξ (see Chatterjee and Bose, 2005, for details), or $\mathbb{E}\Psi_i(\cdot;Y_i)$ is convex (see Niemiro, 1992, Bose, 1998, and Bose and Chatterjee, 2003). The latter case allows for non-smooth $\Psi_i(\cdot;Y_i)$ functions, as long as their expectations are convex functions of ξ.

The conditional distribution of θ_i given Y_i (or the entire data $\mathbf{Y}_n$, owing to the independence of Y_{nj}'s) is given by

$$\pi_i(\cdot;Y_i,\xi) = \left[m_i(Y_i;\xi)\right]^{-1} f_i(Y_i;\cdot,\xi) g_i(\cdot;\xi).$$

The mean of this conditional distribution, is given by

$$\theta_{\pi i} \equiv \theta_{\pi i}(Y_i,\xi) = \int t\pi_i(t;Y_i,\xi)dt.$$

We are interested in predicting θ_i for some fixed i. Note that under the squared error loss, $\theta_{\pi i}(Y_i;\xi)$ is the *best predictor* (BP) of θ_i. However, this is not a statistic since ξ is unknown, and *empirical best predictor* (EBP) $\hat{\theta}_{\pi i} = \theta_{\pi i}(Y_i,\hat{\xi})$ is used as a predictor.

3 The MSPE estimation technique

The *conditional mean squared prediction error* (CMSPE) of $\theta_{\pi i}(Y_i,\hat{\xi})$ is defined as the conditional expectation

$$CMSPE \equiv CMSPE(Y_i,\xi) = \mathbb{E}\left[\left\{\theta_i - \theta_{\pi i}(Y_i,\hat{\xi})\right\}^2 \mid Y_i\right],$$

If $\theta_i \in \mathbb{R}^{q_i}$ (i.e., multivariate), then we define the above as the $q_i \times q_i$ matrix

$$CMSPE \equiv CMSPE(Y_i,\xi) = \mathbb{E}\left[\left\{\theta_i - \theta_{\pi i}(Y_i,\hat{\xi})\right\}\left\{\theta_i - \theta_{\pi i}(Y_i,\hat{\xi})\right\}^T \mid Y_i\right],$$

The *mean squared prediction error* (MSPE) of $\theta_{\pi i}(Y_i,\hat{\xi})$ is defined as

$$MSPE \equiv MSPE(\xi) = \mathbb{E}\left[\theta_i - \theta_{\pi i}(Y_i,\hat{\xi})\right]^2,$$

where $\mathbb{E}$ denotes expectation with respect to the joint distribution of $\mathbf{Y}_n$ and θ_n. If $\theta_i \in \mathbb{R}^{q_i}$ (i.e., multivariate), then we define the above as the $q_i \times q_i$ matrix

$$MSPE \equiv MSPE(\xi) = \mathbb{E}\left[\theta_i - \theta_{\pi i}(Y_i,\hat{\xi})\right]\left[\theta_i - \theta_{\pi i}(Y_i,\hat{\xi})\right]^T.$$

It can be easily seen that the MSPE is the expectation of CMSPE. We consider the case $q_i \equiv 1$ for simplicity. The pointwise estimation for the matrix case follows along identical lines.

Our two-level parametric bootstrap algorithm for generating resamples is given below:

1. Resample $\mathbf{Y}^*_n = (Y^*_{n1},\dots,Y^*_{nn})$ using the following two-level model:

   $$\begin{aligned}&\text{Level } 1^*\text{:}\ \ Y^*_i|\theta^*_i \sim f_i(\cdot;\theta^*_i,\hat{\xi});\\ &\text{Level } 2^*\text{:}\ \ \theta^*_i \sim g_i(\cdot;\hat{\xi}),\end{aligned}$$

   $i = 1,\dots,n$. The expectation at this step, which is conditional on $\mathbf{Y}_n$, is denoted by $\mathbb{E}^*$. Recall that f_i is a density or mass function on $\mathbb{R}^{n_i}$, where n_i is the dimension of Y_i.

2. Obtain $\hat{\xi}^* = \hat{\xi}(\mathbf{Y}^*_n)$, the estimator of ξ based on the resample $\mathbf{Y}^*_n$, using the same technique used to obtain $\hat{\xi}(\mathbf{Y}_n)$.

3. Resample $\mathbf{Y}^{**}_n = (Y^{**}_{n1},\dots,Y^{**}_{nn})$ from $\mathbf{Y}^*_n$ using the following two-level model:

   $$\begin{aligned}&\text{Level } 1^{**}\text{:}\ \ Y^{**}_i|\theta^{**}_i \sim f_i(\cdot;\theta^{**}_i,\hat{\xi}^*);\\ &\text{Level } 2^{**}\text{:}\ \ \theta^{**}_i \sim g_i(\cdot;\hat{\xi}^*),\end{aligned}$$

   $i = 1,\dots,n$. The expectation at this step, which is conditional on $\mathbf{Y}_n$ and $\mathbf{Y}^*_n$, is denoted by $\mathbb{E}^{**}$.

4. We define $\hat{\xi}^{**} = \hat{\xi}(\mathbf{Y}_n^{**})$ and

$$\begin{aligned} M_{1a} &= \mathbb{E}^* \left[\theta_i^* - \theta_{\pi i}(Y_i, \hat{\xi}^*)\right]^2, \\ M_{2a} &= \mathbb{E}^*\mathbb{E}^{**} \left[\theta_i^{**} - \theta_{\pi i}(Y_i, \hat{\xi}^{**})\right]^2, \\ M_{3a} &= \mathbb{E}^* \left[\theta_i^* - \theta_{\pi i}(Y_i^*, \hat{\xi}^*)\right]^2, \\ M_{4a} &= \mathbb{E}^*\mathbb{E}^{**} \left[\theta_i^{**} - \theta_{\pi i}(Y_i^{**}, \hat{\xi}^{**})\right]^2. \end{aligned}$$

The conditional and unconditional MSPE estimators are given by:

$$\widehat{MSPE}_a = H(M_{1a}, M_{2a} - M_{1a}),$$

and

$$\widetilde{MSPE}_a = H(M_{3a}, M_{4a} - M_{3a})$$

respectively.

We consider four different choices for the function $h(\cdot,\cdot)$. These are

$$\begin{aligned} H_1(x,b) &= (x-b)\, I_{\{x>b\}}, \\ H_2(x,b) &= (x-b)\, I_{\{b\le 0\}} \\ &\quad + x\exp\{-b/(x+b)\}\, I_{\{b>0\}}, \\ H_3(x,b) &= \left(x + n^{-1}\tan^{-1}\{-nb\}\right) I_{\{b\le 0\}} \\ &\quad + x^2\left(x + n^{-1}\tan^{-1}\{nb\}\right) I_{\{b>0\}}, \\ H_4(x,b) &= 2x/\left(1+\exp\{2b/x\}\right). \end{aligned}$$

The first function H_1 is a straightforward and a natural choice. The functions H_2 and H_3 are formula (2.17) and (2.18) from the MSPE calibration considered by Hall and Maiti (2006b). The function H_4 seems to work well both when $M_{2a} - M_{1a}$ (or $M_{4a} - M_{3a}$) is positive and negative, and is simple. The graphs of all the four functions are reasonably close in the region of $M_{2a} \approx M_{1a}$ (or $M_{4a} \approx M_{3a}$). However, H_4 seems to perform marginally better than H_2 and H_3 in simulations, and it is always positive, and reasonably close to the intuitive formula H_1.

Theorem 3.1 *The MSPE estimators* $\widehat{MSPE}_a = H_4(M_{1a}, M_{2a} - M_{1a})$ *and* $\widetilde{MSPE}_a = H_4(M_{3a}, M_{4a} - M_{3a})$ *are always positive, and have the following second order accuracy properties:*

$$\begin{aligned} &(Conditional): \\ &\mathbb{E}\left[\widehat{MSPE}_a - CMSPE(Y_i,\xi)\right] = o(d^2 n^{-1}) \\ &(Unconditional): \\ &\mathbb{E}\widetilde{MSPE}_a - MSPE(\xi) = o(d^2 n^{-1}) \end{aligned}$$

Further, they also satisfy

$$\begin{aligned} &(Conditional): \\ &\mathbb{E}\left[\widehat{MSPE}_a - CMSPE(Y_i,\xi)\right]^2 = O(d^2 n^{-1}) \\ &(Unconditional): \\ &\mathbb{E}\left[\widetilde{MSPE}_a - MSPE(\xi)\right]^2 = O(d^2 n^{-1}) \end{aligned}$$

Similar statements hold when function H_1, H_2, or H_3 are used.

4 Simulation Study

In this section, we present results from Monte Carlo simulations to compare the finite-sample performances of the proposed conditional MSPE estimators. In order to compare the MSPE estimators proposed in this paper, we adopt the simulation framework of Pfeffermann and Glickmann (2004) and Datta, Rao, Smith (2005). Thus, we consider the Fay-Herriot model with $m = 15, x_i^T\beta = 0, \psi = 1$, and consider the following pattern for the D_i's: 2.0, 0.6, 0.5, 0.4 and 0.2 [this is pattern (b) of Datta et al., 2005]. We consider five groups of small areas with three areas in each group having the same D_i's values. We discuss results only for our conditional MSPE estimators, although we use an unconditional criterion for comparison of estimators and use unconditional MSPE estimators from literature as competitors. The performance of the unconditional MSPE estimators (denoted by $\widetilde{MSPE}$) is similar and is not reported.

Our simulations, not reported here, tend to suggest that H_2 has an overestimation problem when D_i's are small, and H_3 can have an underestimation problem for larger D_i values, when compared to H_1 and H_4. Based on these observations, we consider $\widehat{MSPE}_a$ using H_1 and H_4 for the our simulation example. For the purpose comparison, we include the following alternative MSPE estimators in our study: (i) a naive MSPE estimator (denoted by "N"), (ii) the Prasad-Rao MSPE estimator (PR), (iii) a naive parametric bootstrap MSPE estimator (NB), (iv) the Fay-Herriot MSPE MSPE estimator (FH) and (v) a parametric bootstrap MSPE estimator (BL) due to Butar and Lahiri (2003) or Pfeffermann and Glickmann (2004) . The naive MSPE estimator (NB) corresponds to plugging in $\hat{\beta}$ and $\hat{A}$ in the formula for the conditional variance of θ_i given Y_i. This is available in all the FPM cases. The naive parametric bootstrap MSPE estimator (NB) is simply M_{1a}, , which can be computed in all cases of the FPM. Neither of these methods are second-order unbiased, but they are available in very wide ranging FPM cases, and can yield conditional MSPE estimators. Both the Prasad-Rao (PR) and the Fay-Herriot (FH) MSPE formula are based on asymptotic Taylor series based approximations. For the PR method, the PR method-of-moments estimator for β and A are used. For the FH method, we use the Fay-Herriot method of estimating β and A described in Datta, Rao and Smith (2005). The parametric bootstrap method due to Butar and Lahiri (2003) and Pfefferman and Glickmann (2004) contains corrective terms on the naive parametric bootstrap formula, to make it second order accurate.

We study three different combinations of the random effects v_i and sampling error e_i distributions: (i) the random effects and sampling errors are Normal random variables; (ii) the random effects are Normal random variables, while the sampling errors follow a Double exponential distribution; and (iii) the random effects are Normal random variables, while the sampling errors follow a location shifted exponential distribution.

We generated 50,000 independent set of variates $\{v_i, e_i,\ \ i = 1, \cdots, m\}$ for each case with specified parameters. Simulated values of $MSPE$ and $E[MSPE$ estimator] were then computed from the 50,000 data sets $\{Y_i = v_i + e_i,\ i = 1, \cdots, m\}$ so generated. For each area, the RB of a MSPE estimator is calculated as

$$RB = [\text{E(MSPE estimator)} - \text{MSPE}]/(\text{MSPE})),$$

the average being taken over the three areas in each of the five groups.

Table 4.1 reports the percent average relative biases (RB) of different MSPE estimators when both the random effects and sampling errors are normally distributed. Table 4.2 reports the case when the random effects are Normal, but the sampling errors follow a double exponential distribution. Table 4.3 reports the case when the random effects are Normal, but the errors follow a location exponential distribution. For all the cases, the naive MSPE estimator (N) and the naive parametric bootstrap MSPE estimator (NB) lead to underestimation. The amount of underestimation can be as high as about 37% for the naive MSPE estimator and about 8% for the naive parametric bootstrap MSPE estimator - this is consistent with the theory, since they are not second-order unbiased.

For the small areas with small values of D_i, the normality-based Prasad-Rao MSPE estimator (PR) usually overestimates and the amount of overestimation could be as high as about 33% even for the normal-normal case. For the small areas with large D_i, the Prasad-Rao MSPE estimator could severely underestimate the true MSPE for the non-normal cases; e.g., when the sampling errors follow a location shifted exponential distribution and random effects follow normal distribution, for the group of small areas with the largest D_i (2.0), the Prasad-Rao MSPE estimator underestimates the true MSPE by about 19%. For the non-normal cases, the normality-based Butar-Lahiri and the Fay-Herriot MSPE estimators have a tendency for underestimation, the underestimation being more severe for the group of small areas with larger D_i values.

The two conditional parametric bootstrap methods proposed in this paper, namely, $\widehat{MSPE}_a$ using functions H_1 and H_4, perform extremely well for all the D_i values, and all combinations of random effect and error distributions, with relative bias below ±3% in all but one case.

References

[1] 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 American Statistical Association*, **80**, 28-36.

[2] Bell, W. (2001). Discussion with "Jackknife in the Fay-Herriott Model with An Example." *Proceedings of the Seminar on Funding Opportunity in Survey Research*, 98-104.

[3] Booth, J.G. and Hobert, J.P. (1998), Standard errors of predictors in generalized linear mixed models, *Journal of American Statistical Association* **93**, 262-272.

[4] Butar, B.F. (1997), Empirical Bayes Methods in Survey Sampling, Unpublished Ph.D. Thesis, Dept. of Mathematics and Statistics, University of Nebraska-Lincoln.

[5] Butar, F., and Lahiri, P. (2003), On measures of uncertainty of empirical Bayes small-area estimators, *Journal of Statistical Planning and Inference,* **112**, 63-76.

[6] Bose, A. and Chatterjee, S. (2003), Generalized bootstrap for estimators of minimizers of convex functionals, *Journal of Statistical Planning and Inference*, **117**, 225-239.

[7] Chatterjee, S. and Bose, A. (2005), Generalized bootstrap for estimating equations, *Annals of Statistics*, **33**, 414-436.

[8] Chatterjee, A. Lahiri, P., and Li, H. (2007), Parametric bootstrap approximation to the distribution of EBLUP, and related prediction intervals in linear mixed models, *Annals of Statistics*, in press.

[9] Chen, S. and Lahiri, P. (2007), On mean squared prediction error estimation in small area estimation problems, *Communications in Statistics,* in press.

[10] Chen, S. and Lahiri, P. (2002), A weighted jackknife MSPE estimator in small-area estimation, *Proceedings of the Survey Research Methods Section of the ASA.*

[11] Chen, S., Lahiri, P. and rao, J.N.K. (2007), Robust Mean Squared Prediction Error Estimators of EBLUP of a Small Area Total Under the Fay-Herriot Model, to appear in the Proceedings of the Statistics Canada.

[12] Das, K., Jiang, J. and Rao, J. N. K. (2004), Mean squared error of empirical predictor, *Annals of Statistics* **32**, 818-840.

[13] Datta, G.S., Lahiri, P. and Maiti, T. (2002), Empirical Bayes estimation of median income of four-person families by states using time series and cross-sectional data, *Journal of Statistical Planning and Inference,* **102**, 83-97.

[14] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005), On measuring the variability of small area estimators under a basic area level model, *Biometrika*, **92**, 183-196.

[15] Efron, B. (1979), Bootstrap methods: another look at the jackknife, *Annals of Statistics*, **7**, 1-26.

[16] Fay, R. E., and Herriot, R. A. (1979), Estimates of Income for Small Places: an Application of James-Stein Procedure to Census Data, *Journal of American Statistical Association* **74**, 269-277.

[17] Fuller, W. A. (1990), Prediction of true values for the measurement error model, *Contemporary Mathematics*, **112**, 41-57.

[18] Hall, P. and Maiti, T. (2006a), Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Annals of Statistics*, **34**, 1733-1750

[19] Hall, P. and Maiti, T. (2006b), On parametric bootstrap methods for small-area prediction. *Journal of the Royal Statistical Society* Ser. B, **68**, 221-238.

[20] Jiang, J. (1996), REML estimation: Asymptotic behavior and related topics, *Annals of Statistics* **24**, 255-286.

[21] Jiang, J. (1998), Asymptotic properties of the empirical BLUP and BLUE in mixed linear models, *Statistica Sinica* **8**, 861-885.

[22] Jiang, J., and Lahiri, P. (2006), Mixed model prediction and small area estimation (with discussions), *Test*, **15**, 1, 1-96.

[23] Jiang, J., Lahiri, P. and Wan, S. (2002), A unified jackknife theory for empirical best prediction with M-estimation, *Annals of Statistics* **30**, 1782-1810.

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

[25] Laird, N. M. and Louis, T. A. (1987), Empirical Bayes confidence intervals based on bootstrap samples, *Journal of the American Statistical Association* **82**, 739-757.

[26] Lee, Y. and Nelder, J. A. (1996), Hierarchical generalized linear models (with discussion), *Journal of the Royal Statistical Society*, B, **58**, 619-678.

[27] Lahiri, P., and Rao, J.N.K. (1995), Robust estimation of mean squared error of small area estimators, *Journal of the American Statistical Association* **90**, 758-766.

[28] Moura, F.A.S. and Holt, D. (1999), Small area estimation using multilevel models, *Survey Methodology* **25**, 73-80.

[29] Niemiro, W., 1992. Asymptotics for M-estimators de=ned by convex minimization. *Ann. Statist.* **20**, 15141533.

[30] Pfeffermann, D. and Nathan, G. (1981), Regression analysis of data from a cluster sample, *Journal of American Statistical Association* **76**, 681-689.

[31] Pfeffermann, D., and Glickman, H. (2004), Mean square error approximation in small area estimation by use of parametric and nonparametric bootstrap, *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

[32] Pfeffermann, D. and Tiller, R. (2005), Bootstrap approximation to prediction MSE for state-space models with estimated parameters, *Journal of Time Series Analysis*, **26**, 893-916.

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

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

[35] Rao, J.N.K. and Yu, M. (1994), Small area estimation by combining time series and cross-sectional data, *Canadian Journal of Statistics,* **22**, 511-528.

[36] Wolter, K. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.

Table 4.1: *Percent Relative biases of different mean squared prediction error estimators for the Fay-Herriot model with $m = 15$; the random effects and the sampling errors are normally distributed.*

Group	N	PR	NB	BL	FH	$\widehat{MSPE}_p$	
D_i						H_1	H_4
2.0	-22.87	-0.32	-7.02	0.42	-3.03	-0.75	-2.21
0.6	-24.66	6.74	-8.12	-1.05	-1.25	-1.18	-2.16
0.5	-24.54	9.05	-8.05	-1.54	0.51	-1.86	-2.57
0.4	-24.41	12.54	-6.51	-1.92	0.57	0.05	-.42
0.2	-24.53	32.90	-4.21	-0.28	3.44	1.50	1.49

Table 4.2: *Percent Relative biases of different mean squared prediction error estimators for the non-normal Fay-Herriot model with $m = 15$; random effects distribution is normal, but sampling error distribution is double exponential.*

Group	N	PR	NB	BL	FH	$\widehat{MSPE}_p$	
D_i						H_1	H_4
2.0	-35.90	-16.16	-8.04	-13.37	-13.38	-2.33	-3.12
0.6	-31.89	1.51	-7.09	-7.87	-5.55	0.22	-0.15
0.5	-32.14	4.49	-7.41	-8.16	-5.01	-0.29	-0.59
0.4	-32.63	9.04	-7.66	-8.19	-4.55	-0.74	-0.92
0.2	-31.29	46.74	-3.73	-3.15	1.86	2.83	2.65

Table 4.3: *Percent Relative biases of different mean squared prediction error estimators for the non-normal Fay-Herriot model with $m = 15$; random effects distribution is normal, but sampling error distribution is location shifted exponential.*

Group	N	PR	NB	BL	FH	$\widehat{MSPE}_p$	
D_i						H_1	H_4
2.0	-36.89	-18.63	-5.22	-11.55	-16.58	0.46	-0.07
0.6	-32.06	-2.51	-5.71	-7.45	-7.74	2.46	2.02
0.5	-32.70	-1.14	-7.02	-7.89	-7.84	0.97	0.36
0.4	-30.54	5.62	-3.34	-4.00	-3.60	4.99	4.56
0.2	-31.64	25.12	-3.17	-2.09	-0.93	3.97	3.83