Journal of Survey Statistics and Methodology (2022) ***10****, 115–130*

PARAMETRIC BOOTSTRAP CONFIDENCE INTERVALS FOR THE MULTIVARIATE FAY–HERRIOT MODEL

TAKUMI SAEGUSA
SHONOSUKE SUGASAWA*
PARTHA LAHIRI

Various multivariate extensions to the well-known Fay–Herriot model have been proposed in the small area estimation literature. Such multivariate models are quite effective in combining information through correlations among small area survey estimates of related variables or historical survey estimates of the same variable or both. Though the literature on small area estimation is already very rich, construction of second-order efficient confidence intervals from multivariate models has received little attention. In this article, we develop a parametric bootstrap method for constructing a second-order efficient confidence interval for a general linear combination of small area means using the multivariate Fay–Herriot normal model. The proposed parametric bootstrap method replaces difficult and tedious analytical derivations by the power of efficient algorithm and high speed computer. Moreover, the proposed method is more versatile than the analytical method because the parametric bootstrap method can be easily applied to any method of model parameter estimation and any specific structure of the variance–covariance

TAKUMI SAEGUSA is an Assistant Professor with the Department of Mathematics, The University of Maryland, College Park, MD 20742, USA. SHONOSUKE SUGASAWA is an Assistant Professor with the Center for Spatial Information Science, The University of Tokyo, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan. PARTHA LAHIRI is Professor with the Department of Mathematics and Joint Program in Survey Methodology, The University of Maryland, College Park, MD 20742, USA.

The second author's research was partially supported by Japan Society for Promotion of Science (KAKENHI) 18K12757. The third author's research was partially supported by the US National Science Foundation Grants SES-1758808.

*Address correspondence to Shonosuke Sugasawa, Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan; E-mail: sugasawa@csis.u-tokyo.ac.jp.

doi: 10.1093/jssam/smaa038 Advance access publication 3 December 2020

matrix of the multivariate Fay–Herriot model avoiding all the cumbersome and time-consuming calculations required in the analytical method. We apply our proposed methodology in constructing confidence intervals for the median income of four-person families for the fifty states and the District of Columbia in the United States. Our data analysis demonstrates that the proposed parametric bootstrap method, applied to both multivariate and univariate Fay–Herriot models, generally provides much shorter confidence intervals compared to the corresponding traditional direct method. Moreover, the confidence intervals obtained from the multivariate model are generally shorter than the corresponding intervals from the univariate model indicating the potential advantage of exploiting correlations of median income of four-person families with median incomes of three- and five-person families.

KEYWORDS: Empirical Best predictor; Higher-order asymptotics; Small area estimation.

1. INTRODUCTION

For the last few decades, there has been an increasing demand to produce reliable estimates for small geographic areas, commonly referred to as small areas, since such estimates are routinely used for fund allocation and regional planning. The primary data, usually survey data, are usually too sparse to produce reliable direct small area estimates that use data from the small area under consideration. To improve upon direct estimates, different small area estimation techniques that use multi-level models to combine information from relevant auxiliary data have been proposed in the literature. The readers are referred to Jiang and Lahiri (2006) and Rao and Molina (2015) for a comprehensive review of small area estimation.

In estimating per-capita income of small places (population less than 1,000), Fay and Herriot (1979) proposed an empirical Bayes method to improve on direct survey-weighted estimates by borrowing strength from administrative data and survey estimates from a bigger area. Their method uses a two-level normal model in which the first level captures the variability in the survey estimates and the second level links the true small area means to aggregate statistics from administrative records and survey estimates for a bigger area. Researchers working on small area estimation have found the Fay–Herriot model useful in investigating various theoretical properties as well as implementing methodology in different applied problems when we do not have access to micro-data because of confidentiality and other reasons. For a review on the Fay–Herriot model and the related empirical best predictions, readers are referred to Lahiri (2003b).

Following the pioneering paper by Fay and Herriot (1979), several multivariate extensions of the Fay–Herriot model have been considered to combine

information from small area estimates of related variables or from past small area estimates of the same variable or both. They are essentially special cases of the general multivariate random effects or two-level multivariate model. In the context of estimating median income of four-person families for the fifty states and the District of Columbia (small areas), Fay (1987) suggested a multivariate extension of the Fay–Herriot model, commonly referred to as the multivariate Fay–Herriot (MFH) model, to borrow strength from the corresponding survey estimates of median income of three- and five-person families for the small areas. Alternatively, in Fay's setting one could think of using survey estimates of median income for the three-person and five-person families as auxiliary variables in a univariate Fay–Herriot (UFH) model. But, unlike the UFH model, the MFH model incorporates sampling variance–covariance matrix of direct survey estimates of median income of the three-, four-, and five-person families for each small area. Moreover, the MFH model borrows strengths through correlations of the components of area-specific vector of random effects associated with the true median income of three-, four-, and five-person families. Inferences on the four-person median income for the small areas drawn from the MFH model are expected to be more efficient and reasonable when compared to the inferences drawn from an UFH model with survey estimates of the median income for three- and five-person families as auxiliary variables. This is because the UFH model would ignore the sampling variability of the survey estimates of median income for the three- and five-person families in the small areas.

In estimating median income of four-person families for the fifty states and the District of Columbia, Datta, Fay, and Ghosh (1991) used a bivariate Fay–Herriot model with a general structure for the variance–covariance matrix of the vector of area-specific random effects. However, in many small area applications, structured variance–covariance matrices for the vector of area-specific random effects arise naturally. For example, to combine information from the related past data, Rao and Yu (1994) proposed a stationary time series cross-sectional model while Datta, Lahiri, and Maiti (2002a) proposed a random walk time series and cross-sectional model. Although the time series cross-sectional models can be viewed as special cases of the MFH model, one can achieve greater efficiency in estimating the unknown variance–covariance matrix by reducing the number of parameters in the variance–covariance matrix through time series cross-sectional models.

Empirical best linear unbiased predictions and associated uncertainty measures for MFH models with or without structured variance–covariance matrices for the vector of random effects have been adequately studied (see, e.g., Datta et al. 1991; Rao and Yu 1994; Benavent and Morales 2016; Datta et al. 2002a; and others). However, the problem of constructing second-order efficient confidence intervals for the MFH model, that is, confidence intervals with coverage error $o(m^{-1})$, m being the number of small areas, received very little attention.

Datta, Ghosh, Smith, and Lahiri (2002b) obtained a second-order efficient confidence interval for a small area mean using an analytical method. To this end, they first obtained the exact expression for the term of order $O(m^{-1})$ in a higher-order expansion of coverage probability of a normality-based empirical Bayes confidence interval, originally proposed by Cox (1975). Then, using the $O(m^{-1})$ term in the expansion, they suggested an adjustment to the normal percentile to lower the coverage error to $o(m^{-1})$. The approach of Ito and Kubokawa (2021) in obtaining a second-order efficient confidence region for the vector of means for each area is essentially a multivariate generalization of Datta et al. (2002b). However, their results are specifically designed for the MFH model with an unstructured variance–covariance matrix when a method-of-moment estimator of the variance–covariance matrix of the vector of random effect is used. The derivation of the second-order efficient confidence intervals by the analytical method of Datta et al. (2002b) or Ito and Kubokawa (2021) is cumbersome, and one needs to go through the such derivation each time one changes the model [say, an MFH model with model variance–covariance structure suggested by the time series cross-sectional model of Rao and Yu (1994) or Datta et al. (2002a)] or estimation method for the model parameters.

Parametric bootstrap method for obtaining second-order unbiased mean squared error estimation was first proposed by Butar and Lahiri (2002). Construction of the second-order efficient confidence interval based on the empirical best linear predictor of a small area parameter for a general linear mixed model was proposed by Chatterjee, Lahiri, and Li (2008). For parametric bootstrap confidence intervals for the UFH model, see Lahiri (2003a) and Li and Lahiri (2010). In this paper, we develop a parametric bootstrap method for obtaining second-order efficient confidence intervals for small area parameters from an MFH model. Compared to the analytical method, our parametric bootstrap approach for constructing second-order confidence intervals for small area parameters is versatile and theoretically complete because our method applies to any variance estimator with minimal assumptions and theoretical justification is directly provided to the proposed method.

In section 2, we describe the multivariate model, associated estimation of the model parameters, and the proposed parametric confidence interval for a linear combination of small area means. We present our data analysis in section 3. An outline of the technical proof of our main result is deferred to the Appendix.

2. PARAMETRIC BOOTSTRAP CONFIDENCE INTERVALS FOR THE MFH MODEL

2.1 MFH Model

Let $\theta_i = (\theta_{i1}, \ldots, \theta_{is})^t$ and $y_i = (y_{i1}, \ldots, y_{is})^t$ be a vector of characteristics of interest and a vector of direct survey estimates of θ_i for area $i\,(i = 1, \ldots, m)$, respectively, where m is the number of small areas. The MFH model (Fay, 1987; Benavent and Morales, 2016) is given by

$$y_i = \theta_i + \varepsilon_i, \quad \theta_i = X_i\beta + v_i, \quad i = 1, \ldots, m, \tag{1}$$

where X_i is a $s \times p$ matrix of known explanatory variables; ε_i and v_i are vectors of area-specific sampling errors and random effects, respectively; and $\{\varepsilon_i,\ i = 1, \ldots, m\}$ and $\{v_i,\ i = 1, \ldots, m\}$ are all independent with $\varepsilon_i \sim N(0, D_i)$ and $v_i \sim N(0, A(\psi))$, D_i being the $s \times s$ known sampling variance–covariance matrix of y_i, $i = 1, \ldots, m$.

We assume that $A(\psi)$, the variance–covariance matrix of the random effects θ_i, depends on k unknown parameters $\psi = (\psi_1, \ldots \psi_k)$ with $1 \le k \le s(s+1)/2$. For the small area application considered by Datta et al. (1991), $A(\psi)$ is an unstructured variance–covariance matrix with $s = 2$ and $k = 3$. For the stationary time series cross-sectional model of Rao and Yu (1994), $A(\psi)$ is a structured variance–covariance matrix with s as the number of time points, and $k = 3$. For the random walk time series cross-sectional model of Datta et al. (2002a), $A(\psi)$ is a structured variance–covariance matrix with s as the number of time points, and $k = 2$. Let $\phi = (\beta, \psi)$ be a vector of all the unknown parameters.

We define $y = (y_1^t, \ldots, y_m^t)^t$, $X = (X_1^t, \ldots, X_m^t)^t$ and define v, ε, and θ in the same way as y. Then, the model can be expressed as

$$y = X\beta + v + \varepsilon,$$

where $v \sim N(0, \tilde{A}(\psi))$ with $\tilde{A}(\psi) = \mathrm{diag}(A(\psi), \ldots, A(\psi)) \in \mathbb{R}^{ms \times ms}$ and $\varepsilon \sim N(0, D)$ with $D = \mathrm{diag}(D_1, \ldots, D_m) \in \mathbb{R}^{ms \times ms}$. With this notation, we can write $\mathrm{Var}(y) \equiv \Sigma = \mathrm{diag}(A(\psi) + D_1, \ldots, A(\psi) + D_m)$. In this paper, we are interested in constructing confidence intervals for $T = c^t\theta$, where c is a ms-dimensional vector of known constants. For example, if we let $c = (1, 0, \ldots, 0)$, $T = \theta_{i1}$ is the small area mean of the first characteristics in the first area.

Under the model (1), the best linear unbiased predictor of θ_i with known model parameters is given by

$$\tilde{\theta}_i = y_i - D_i\{A(\psi) + D_i\}^{-1}(y_i - X_i\beta), \quad i = 1, \ldots, m,$$

which shrinks y_i toward the regression part $X_i\beta$. Note that each element in $\tilde{\theta}_i$ depends not only on the corresponding observation but also on other observations in the same area when D_i or $A(\psi)$ have nonzero off-diagonal elements.

Exploiting the information from the correlation structure, the best linear unbiased predictor can potentially provide more accurate estimates of θ_i than simple applications of the univariate FH models to each element. In fact, we have empirically shown that such advantage is inherited to interval lengths of confidence intervals. The MFH model provides more efficient confidence intervals than the UFH model by borrowing information from related components.

2.2 Estimation of Model Parameters

Because the best linear unbiased predictor $\tilde{\theta}_i$ depends on unknown model parameters, statistical inference on θ_i is carried out via the empirical best linear unbiased estimator given by

$$\widehat{\theta}_i = y_i - D_i\{A(\widehat{\psi}) + D_i\}^{-1}(y_i - X_i\widehat{\beta}),$$

where $A(\widehat{\psi})$ and $\widehat{\beta}$ are consistent estimators of $A(\psi)$ and β, respectively. We estimate β by the generalized least squares estimator

$$\widehat{\beta} = (X^t\widehat{\Sigma}^{-1}X)^{-1}X^t\widehat{\Sigma}y$$

once $A(\widehat{\psi})$ in $\widehat{\Sigma}$ is obtained. There are several different methodologies to estimate $A(\psi)$ (e.g., the restricted maximum likelihood estimator (Benavent and Morales, 2016) and moment-based estimators (Ito and Kubokawa, 2021)), but the proposed method to construct the empirical Bayes confidence interval does not depend on a specific variance estimator. For the data analysis in section 3, we adopt the maximum likelihood estimator that maximizes

$$\begin{aligned} L(\phi) = &-\frac{1}{2}\sum_{i=1}^{m}\log|A(\psi) + D_i| \\ &-\frac{1}{2}\sum_{i=1}^{m}(y_i - X_i\beta)^t\{A(\psi) + D_i\}^{-1}(y_i - X_i\beta). \end{aligned}$$

In this paper, we use the EM algorithm, described in the Appendix, for obtaining maximum likelihood estimates. Note that this method estimates β and $A(\psi)$ simultaneously and automatically yields the generalized least squares estimator $\widehat{\beta}$ as the maximum likelihood estimator.

2.3 Confidence Intervals via Parametric Bootstrap

We describe our methodology to construct the empirical Bayes confidence interval for $T = c^t\theta$. To motivate our method, we first consider a traditional approach to interval estimation. The key observation for this approach is that the

conditional distribution of $T = c^t\theta$ under model (1) is given by $T|y \sim N(\mu_T, \sigma_T^2)$, where

$$\mu_T \equiv \mu_T(y, \phi) = c^t D\Sigma^{-1}X\beta + c^t\tilde{A}(\psi)\Sigma^{-1}y,$$

$$\sigma_T^2 \equiv \sigma_T^2(\psi) = c^t \mathrm{diag}\Big((A(\psi)^{-1} + D_1^{-1})^{-1}, \ldots, (A(\psi)^{-1} + D_m^{-1})^{-1}\Big)c.$$

Since $\sigma_T^{-1}(T - \mu_T)$ follows the standard normal distribution, one can find z such that $P(\sigma_T^{-1}|T - \mu_T| \le z) = 1 - \alpha$ for a fixed $\alpha \in (0, 1)$. Because the resultant interval $\mu_T \pm z\sigma_T$ for T contains unknown parameters μ_T and σ_T, the traditional approach replaces these parameters by their consistent estimators $\widehat{\mu}_T$ and $\widehat{\sigma}_T$ to obtain the confidence interval $\widehat{\mu}_T \pm z\widehat{\sigma}_T$ for T. Though this interval has a correct coverage asymptotically, it tends to be too short or too long in practice. This undesirable phenomenon is due to the reliance on the rather crude approximation of the distribution of $\widehat{\sigma}_T^{-1}(T - \widehat{\mu}_T)$ by the standard normal distribution, which yields the coverage error of $O(m^{-1})$. Because μ_T and σ_T must be estimated, the issue of the asymptotic approximation is not avoidable. Instead, we consider the distribution of $\widehat{\sigma}_T^{-1}(T - \widehat{\mu}_T)$ from the beginning and consider a method to precisely approximate it. We achieve this goal through the parametric bootstrap.

We construct the bootstrap sample in a parametric way as follows. First, we independently generate $v_i^* \sim N(0, A(\widehat{\psi}))$ and $\varepsilon_i^* \sim N(0, D_i)$. Because $\theta_i = X_i\beta + v_i$ and $y_i = \theta_i + \varepsilon_i$ in model (1), we construct

$$\theta_i^* = X_i\widehat{\beta} + v_i^*,$$

$$y_i^* = \theta_i^* + \varepsilon_i^*.$$

The resultant bootstrap sample is $\{(y_1^*, X_1), \ldots, (y_m^*, X_m)\}$. To approximate the distribution of $\widehat{\sigma}_T^{-1}(T - \widehat{\mu}_T)$, we compute $T^* = c^t\theta^*$ with $\theta^* = ((\theta_1^*)^t, \ldots, (\theta_m^*)^t)^t$. Bootstrap estimates $\widehat{\mu}_T^*$ and $\widehat{\sigma}^*$ of μ_T and σ_T are obtained in the same way as $\widehat{\mu}_T$ and $\widehat{\sigma}_T$ by replacing the original sample y_i by the bootstrap sample y_i^*. For example, one can compute the bootstrap maximum likelihood estimate $\widehat{\phi}^* = (\widehat{\beta}^*, \widehat{\psi}^*)$ by maximizing

$$L^*(\phi) = -\frac{1}{2}\sum_{i=1}^{m} \log|A(\psi) + D_i| - \frac{1}{2}\sum_{i=1}^{m} (y_i^* - X_i\beta)^t\{A(\psi) + D_i\}^{-1}(y_i^* - X_i\beta).$$

Once $\widehat{\phi}^*$ is computed, we obtain $\widehat{\mu}_T^* = \mu_T(y^*, \widehat{\phi}^*)$ and $\widehat{\sigma}_T^* = \sigma_T(\widehat{\psi}^*)$.

The conditional distribution of

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

given the data y is the parametric bootstrap approximation to the distribution of $\widehat{\sigma}_T^{-1}(T - \widehat{\mu}_T)$. Because the random variable $\widehat{\sigma}_T^{-1*}(T^* - \widehat{\mu}_T^*)$ can be generated as described above, one can find (q_1, q_2) satisfying $P(q_1 \leq \widehat{\sigma}_T^{-1*}(T^* - \widehat{\mu}_T^*) \leq q_2) = 1 - \alpha$ as precisely as possible. Specifically, we compute the lower and upper $(\alpha/2)$ quantiles of the bootstrap distribution. Because parametric bootstrap provides a precise approximation, (q_1, q_2) is expected to yield a similar probability for $\widehat{\sigma}_T^{-1}(T - \widehat{\mu}_T)$. The proposed parametric bootstrap confidence interval is

$$\widehat{\mu}_T + \widehat{\sigma}_T q_1 \leq T \leq \widehat{\mu}_T + \widehat{\sigma}_T q_2.$$

The following theorem states that the proposed empirical Bayes confidence interval achieves correct coverage asymptotically with error $O(m^{-3/2})$.

Theorem 1 We assume the following conditions:

- The matrix X is of full rank satisfying $(X'\Sigma^{-1}X)^{-1} = O(m^{-1})$.
- $A(\widehat{\psi})$ is a strictly positive definite matrix satisfying $E||A(\widehat{\psi}) - A(\psi)||_F = O(m^{-1})$, where $||\cdot||_F$ is the Frobenius norm.
- There exists positive constants $\underline{\lambda}$ and $\bar{\lambda}$ such that the sampling variance–covariance matrix D_i satisfies $\underline{\lambda} I_s \leq D_i \leq \bar{\lambda} I_s$ *for* $i = 1, \ldots, m$.

Let $\alpha \in (0, 1)$. Suppose (q_1, q_2) satisfies

$$P(q_1 \leq \widehat{\sigma}_T^{-1*}(T^* - \widehat{\mu}_T^*) \leq q_2) = 1 - \alpha.$$

Then

$$P(\widehat{\mu}_T + \widehat{\sigma}_T q_1 \leq T \leq \widehat{\mu}_T + \widehat{\sigma}_T q_2) = 1 - \alpha + O(m^{-3/2}).$$

3. APPLICATION

In this section, we use old data used earlier by Datta et al. (1991) to compare three different confidence interval methods: direct method and parametric bootstrap confidence interval methods—one based on a UFH model and the other based on MFH model. The data contain direct survey estimates of median income of three-, four-, and five-person families and their associated standard errors for the fifty states and the District of Columbia during years 1979–1988. In addition, data contain the census median income of three-, four-, and five-person families obtained from the 1970 and 1980 decennial censuses. The US Department of Health and Human Services (HHS) administers a program of energy assistance to low-income families. Eligibility for the program is

determined by a formula where the most important variable is an estimate of the current median income for four-person families by states.

Let θ_{i1}, θ_{i2}, and θ_{i3} denote the true median income of three-, four-, and five-person families, respectively, for $i = 1, \ldots, m$, where $m = 51$ is the number of states and the District of Columbia in the United States. Let y_{i1}, y_{i2} and y_{i3} be the corresponding direct survey estimates. Our primary interest is the four-person family median income, θ_{i2}. We consider estimation of the four-person family median income by borrowing strength from not only area-specific auxiliary variables but also the direct survey estimates of median income for the three- and five-person families. As for the area-specific auxiliary variables, we consider the median income data obtained from the most recent decennial census and an "adjusted" census median income obtained by multiplying the most recent census median income by the ratio of per-capita income of the current year to the most recent decennial census year. The per-capita income information is available from administrative records maintained by the Bureau of Economic Analysis. The covariate matrix X_i is a 3×9 matrix given by

$$X_i = \text{diag}((1, x_{i1}, x_{i1}^*), (1, x_{i2}, x_{i2}^*), (1, x_{i3}, x_{i3}^*)),$$

where x_{ik} and x_{ik}^* denote the census and adjusted census median income for the k-person family ($k = 3, 4, 5$) in the ith area.

For each year during the period 1981–1988, we first applied the MFH model described by (1). We used 1979 median income data obtained from the 1980 decennial census data as auxiliary variables. For comparison, we also applied the UFH model only to the four-person family income data y_{i2} with the corresponding census data as auxiliary variables. We found that the maximum likelihood estimates of the random effects variance in the UFH model were 0 in 1982, 1983, and 1986 and thus for these years confidence interval of θ_{i2} cannot be obtained. On the other hand, we observed that the MFH model produces positive definite estimates for A in all the years, and correlations are quite high in some years. This indicates that the random effects variance in θ_{i2} can be adequately estimated by borrowing strength from other information such as y_{i1} and y_{i3} through the MFH model (1). For illustration, we focus on the results in 1984 and 1987 in which the estimated correlation matrices are given by

$$\begin{pmatrix} 1 & 0.171 & 0.938 \\ 0.171 & 1 & 0.200 \\ 0.938 & 0.200 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0.780 & 0.587 \\ 0.780 & 1 & 0.915 \\ 0.587 & 0.915 & 1 \end{pmatrix},$$

respectively. Note that the correlations are quite high in 1987 while relatively small in 1984.

Based on 1,000 bootstrap replications, we computed 95 percent parametric bootstrap confidence intervals of θ_{i2} under both MFH and UFH models. We

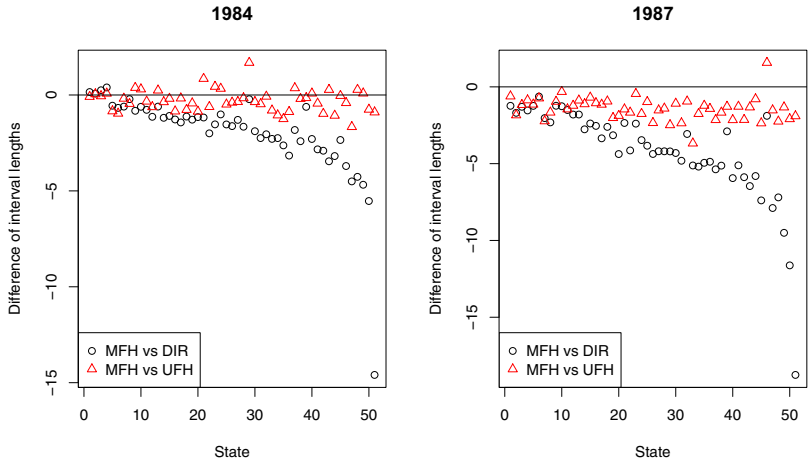

Figure 1. Plots of the Differences between the Lengths of 95 Percent Parametric Bootstrap Confidence Intervals Based on the MFH Model and Those of the Corresponding Parametric Bootstrap Confidence Intervals Based on the UFH Model [the Naive Confidence Intervals Based on Direct Estimates (DIR)] against States Arranged in the Ascending Order of Their Sampling Variances.

Table 1. Descriptive Statistics of Lengths of 95 Percent Confidence Intervals Constructed using the Parametric Bootstrap Method Based on the MFH and UFH Models and the Direct Method (DIR)

Year	Method	Minimum	25%	Median	Mean	75%	Maximum
1984	MFH	4.12	5.55	6.13	6.02	6.38	8.28
	UFH	4.21	5.78	6.42	6.32	6.72	8.44
	DIR	3.96	6.66	7.75	8.00	8.84	21.30
1987	MFH	4.54	6.08	6.78	6.74	7.23	12.87
	UFH	5.86	7.60	8.57	8.39	8.95	11.85
	DIR	6.18	8.23	10.71	10.84	12.15	32.39

NOTE.—The statistics are computed using data for all the states and the District of Columbia.

also computed 95 percent confidence intervals based on the direct estimator (denoted by DIR), given by $(y_{i2} - z_{0.025}\sqrt{D_{i22}}, y_{i2} + z_{0.025}\sqrt{D_{i22}})$, where $z_{0.025}$ is the upper 0.025 quantile of the standard normal distribution, and D_{i22} is the (2)-element of D_i. Figure 1 displays the differences between lengths of 95 percent parametric bootstrap confidence interval based on the MFH model and those based on the UFH model and the DIR method, where the states are arranged in the ascending order of sampling variances. A negative value of the difference indicates that the length of the confidence interval from the MFH

Table 2. Performance of 95 Percent Confidence Intervals Constructed using the Parametric Bootstrap Method Based on the MFH and UFH Models and the Direct Method (DIR)

Method	Three-person family			Four-person family			Five-person family		
	CR	Len1	Len2	CR	Len1	Len2	CR	Len1	Len2
MFH	100	3.21	3.13	96.1	2.40	2.29	100	4.52	4.36
UFH	74.5	1.77	1.69	–	–	–	–	–	–
DIR	86.3	5.57	5.41	86.3	5.88	5.91	86.3	9.25	8.98

NOTE.—The results for UFH in four- and five-person family incomes do not exist because of zero estimates of the random effects variances.

model is shorter than that from the UFH model or the DIR method. In table 1, we report descriptive statistics of lengths of confidence intervals constructed using these three methods. Comparing the intervals from the MFH model and the DIR method, the difference tends to increase as sampling variance increases. This is reasonable because we can improve the accuracy of inference on parameters in areas with large sampling variances by borrowing strength through the model. Comparing intervals from the MFH and UFH models, the lengths are comparable in 1984 possibly because the correlations among three median incomes are not so strong. The advantage of borrowing strength from the other incomes can be limited. On the other hand, in 1987, the MFH model produces shorter confidence intervals than the UFH model in almost all the areas due to the high correlations.

We next investigate the performance of the confidence intervals using the 1979 census data treating them as true values. We applied the MFH and UFH methods to the 1979 survey data using the 1969 census data as covariates. In this case, we applied the UFH method not only to y_{i2} but also to y_{i1} and y_{i3}. We observe that for the UFH model both maximum likelihood and restricted maximum likelihood estimates of the random effects variances are zero (for y_{i2} and y_{i3}) or very small (for y_{i1}). Based on 1,000 bootstrap replications and the maximum likelihood method, we obtained 95 percent confidence intervals of θ_{ik} ($k = 1, 2, 3$) for the MFH and UFH models. We calculated the mean and median lengths of confidence intervals, given by, (Len1) $m^{-1} \sum_{i=1}^{m} |\mathrm{CI}_i|$ and (Len2) Median$\{|\mathrm{CI}_1|, \ldots, |\mathrm{CI}_m|\}$, respectively, where CI_i is the confidence interval for the ith state. We also computed the empirical coverage rate (CR), given by $m^{-1} \sum_{i=1}^{m} I(\theta_i \in \mathrm{CI}_i)$, using the 1979 census data as true values θ_i. The results are reported in table 2. Although the UFH model provides shorter confidence intervals than the MFH model, the empirical CR is quite low when compared to the nominal level 95 percent. This suggests that the confidence intervals for the UFH model are too anti-conservative in this case, possibly

because of small estimates of random effects variance. On the other hand, the MFH model provides reasonable confidence intervals. Their CRs are quite high and their lengths are much shorter than those of the direct method.

4. CONCLUDING REMARKS

In this article, we introduce a parametric bootstrap method for constructing a second-order accurate confidence interval of a small area parameter from the MFH model. The proposed parametric bootstrap method is easy to implement and can be applied to different estimators of model parameters for a wide range of small area models. This is in a sharp contrast to the analytical calibration method proposed by Datta et al. (2002b) and Ito and Kubokawa (2021) where a different estimation method of model parameters requires cumbersome derivations of the correction terms and tedious checking of assumptions. We have demonstrated superiority of the proposed methodology over the corresponding parametric bootstrap method for the univariate model and direct method using a real-life data analysis. Better coverage and generally shorter length of the proposed interval are due to the effective use of the correlation structure in the same area and direct approximation of the distribution through the parametric bootstrap method.

We can extend the proposed methodology in several new directions. An immediate extension will be to construct the parametric bootstrap confidence region of a vector of small area parameters as an alternative to the analytical method studied by Ito and Kubokawa (2021). In this paper, we have focused on a linear combination of small area parameters. When more than three parameters are of interest, our simple and versatile parametric bootstrap method is expected to be a powerful alternative to the analytical calibration. Though the methodology proposed in this paper can be easily extended to construct confidence region for multiple parameters, theoretical justification of such method is expected to be a challenging problem because the problem involves multivariate integrals. Another research direction would be to extend the parametric bootstrap to a more general multivariate linear mixed models. Because theoretical arguments by Chatterjee et al. (2008) for the general univariate case is similar to that in this paper, its multivariate extension can be attempted in a similar way. Another interesting problem of future research interest would be to address the issue of possible nonpositive definiteness of the estimated variance–covariance matrices. Singularities of estimated variance–covariance matrices may occur in estimates derived from both the original sample and the bootstrap replicates. Because this issue compromises the validity of the parametric bootstrap procedure, it is important to develop reasonable adjustment methods such as the existing approaches in the univariate situation (e.g., Li and Lahiri, 2010).

APPENDIX

EM ALGORITHM

The EM algorithm for computing the maximum likelihood estimates starts with initial values $\beta^{(0)}$ and $\psi^{(0)}$. The algorithm then updates the values at the $(s+1)$th iteration from the sth iteration as follows:

$$\beta^{(s+1)} = \{X^t\Sigma(\psi^{(s)})^{-1}X\}^{-1}X^t\Sigma(\psi^{(s)})^{-1}y,$$

$$\psi^{(s+1)} = \text{argmin}_\psi \sum_{i=1}^{m} \log|\Sigma_i(\psi)| + \sum_{i=1}^{m}(y_i - X_i\beta^{(s+1)})^t\Sigma_i(\psi)^{-1}(y_i - X_i\beta^{(s+1)}),$$

where $\Sigma_i(\psi) = A(\psi) + D_i$ and $\Sigma(\psi) = \text{blockdiag}(\Sigma_1(\psi), \ldots, \Sigma_m(\psi))$. If the updated variance–covariance matrix $\Sigma(\psi)$ is nonpositive definite, we modify the matrix to make it positive definite. Let $\Sigma(\psi) = H^t\Lambda H$ be an eigenvalue decomposition, where H is an orthogonal matrix and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_s)$ with eigenvalues $\lambda_1, \ldots, \lambda_s$ of $\Sigma(\psi)$. Then, a modified version of $\Sigma(\psi)$ is defined as $\Sigma^*(\psi) = H^t\Lambda^*H$, where $\Lambda^* = \text{diag}(\max(\delta, \lambda_1), \ldots, \max(\delta, \lambda_s))$ for some $\delta > 0$.

PROOF OF THEOREM 1

For notational simplicity, we suppress the dependence of matrices on ψ. For example, we write A and $\widehat{A}$ for $A(\psi)$ and $A(\widehat{\psi})$, respectively.

Recall that the conditional distribution of θ given Y is a multivariate normal distribution with mean $\mu = (\mu_1^t, \ldots, \mu_m^t)^t$ and the variance–covariance matrix $\sigma^2 = \text{diag}\{(A^{-1} + D_1^{-1})^{-1}, \ldots, (A^{-1} + D_m^{-1})^{-1}\}$, where $\mu_i = A(A + D_i)^{-1}y_i + D_i(A + D_i)^{-1}X_i\beta$. Let $T = c^t\theta$. The conditional distribution of T given Y is then a normal distribution with mean $\mu_T = c^t\mu$ and variance $\sigma_T^2 = c^t\sigma^2c$. Let Φ and ϕ be the cumulative distribution function and density function for the standard normal random variable, respectively. Define $Q(Y) = \sigma_T^{-1}\{\widehat{\mu}_T - \mu_T + r(\widehat{\sigma}_T - \sigma_T)\}$. It follows that

$$
\begin{aligned}
P(\widehat{\sigma}_T^{-1}(T-\widehat{\mu}_T) \le r) &= E[\sigma_T^{-1}(T-\mu_T) \le r + Q(Y)|Y)] \\
&= E[\Phi(r+Q(Y))] \\
&= \Phi(r) + \phi(r)E[Q(Y] - \frac{1}{2}r\phi(r)E[Q(Y)^2] \\
&\quad + \frac{1}{2}E[\int_r^{r+Q}(r+Q-x)^2(x^2-1)\phi(x)dx] \\
&\equiv \Phi(r) + \phi(r)T_1 - \frac{1}{2}r\phi(r)T_2 + T_3.
\end{aligned}
$$

Because of the facts that $|r+Q-x| \le |Q|$ for $x \in (r, r+Q)$ and $(x^2-1)\phi(x)$ is uniformly bounded, we have

$$
\begin{aligned}
T_3(r) &= \frac{1}{2}E[\int_r^{r+Q}(r+Q-x)^2(x^2-1)\phi(x)dx] \\
&\le \frac{1}{2}E[\int_r^{r+Q}|r+Q-x|^2|(x^2-1)\phi(x)|dx] \\
&\le CE[Q^2\textstyle\int_r^{r+Q}dx] \le CE|Q|^3,
\end{aligned}
$$

for some constant $C>0$. Thus, the evaluation of $P(\widehat{\sigma}_T^{-1}(T-\widehat{\mu}_T) \le r)$ reduces to the evaluation of EQ, EQ^2, and $E|Q|^3$. In particular, if $EQ = O(m^{-1})$, $EQ^2 = O(m^{-1})$ and $EQ^4 = O(m^{-2})$, then it follows that

$$
E|Q|^3 \le (EQ^4)^{-3/4} = O(m^{-3/2})
$$

by the Hölder's inequality so that

$$
P(\widehat{\sigma}_T^{-1}(T-\widehat{\mu}_T) \le r) = \Phi(r) + O(m^{-1})\gamma(r,\beta,\psi) + O(m^{-3/2}), \qquad (2)
$$

where γ is a smooth function of $O(1)$. A mathematical argument similar to the one used in the derivation of (2) leads to

$$
P(\widehat{\sigma}_T^{-1*}(T^*-\widehat{\mu}_T^*) \le r) = \Phi(r) + O(m^{-1})\gamma(r,\widehat{\beta},\widehat{\psi}) + O(m^{-3/2}).
$$

In the following, we provide a sketch of the proof of (2) by verifying $EQ^4 = O(m^{-2})$. Once we obtain this result, the theorem can be proved along the lines of Chatterjee et al. (2008).

To study the moment of Q, first consider the element of $\widehat{\mu}_T - \mu$. We have

$$\begin{aligned}
\widehat{\mu}_i - \mu_i &= A(A+D_i)^{-1}y_i + D_i(A+D_i)^{-1}X_i\beta - \widehat{A}(\widehat{A}+D_i)^{-1}y_i + D_i(\widehat{A}+D_i)^{-1}X_i\widehat{\beta} \\
&= D_i(A+D_i)^{-1}X_i(X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}(v+\varepsilon) \\
&+D_i(A+D_i)^{-1}X_i\{(X^t\widehat{\Sigma}^{-1}X)^{-1}X^t\widehat{\Sigma}^{-1} - (X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}\}(v+\varepsilon) \\
&+\left(\widehat{A}(\widehat{A}+D_i)^{-1} - A(A+D_i)^{-1}\right)(J_i - X_i(X^t\widehat{\Sigma}^{-1}X)^{-1}X^t\widehat{\Sigma}^{-1})(v+\varepsilon) \\
&+\left(D_i(\widehat{A}+D_i)^{-1} - D_i(A+D_i)^{-1}\right)X_i\beta \\
&+\left(\widehat{A}(\widehat{A}+D_i)^{-1} - A(A+D_i)^{-1}\right)X_i\widehat{\beta} \\
&= R_{1i} + R_{2i} + R_{3i} + R_{4i} + R_{5i},
\end{aligned}$$

where J_i is a $s \times s$ matrix with 1 in the (i, i)-element and 0 otherwise. Let $R_j = (R_{j1}^t, \ldots, R_{jm}^t)^t$, $j = 1, \ldots, 5$. Thus, we can write

$$Q(Y) = \sigma_T^{-1}\{c^tR_1 + c^tR_2 + c^tR_3 + c^tR_4 + c^tR_5 + q(\widehat{\sigma}_T - \sigma_T)\}.$$

We evaluate moments of c^tR_1. Clearly, $E[c^tR_1] = 0$. For the second moment, a general term of the matrix $E[R_1R_1^t]$ is given by

$$E[R_{1i}R_{1j}^t] = D_i(A+D_i)^{-1}X_i(X^t\Sigma^{-1}X)^{-1}X_j^t(A+D_j)^{-1}D_j^t.$$

Since $(X^t\Sigma^{-1}X)^{-1} = O(m^{-1})$ and c is fixed, we obtain $E(c^tR_1)^2 = O(m^{-1})$. For the eighth moment, note that the fourth moment of the sum is the sum of the fourth moments up to constant. Thus, we consider the fourth moment of $c_i^tR_{1i}$. We first note that $(X^t\Sigma^{-1}X)^{-1} = O(m^{-1})$ and

$$U \equiv X^t\Sigma^{-1}(v+\varepsilon) = \sum_{i=1}^{m} X_i\Sigma_i^{-1}(v_i+\varepsilon_i) = O_p(m^{1/2}).$$

The last result is obtained by an application of the central limit theorem since $\{v_i + \varepsilon_i\}_{i=1,\ldots,m}$ are independent. An application of the Cauchy–Schwartz inequality yields

$$E(c_i^tR_{1i})^4 \;\le\; (c_i^tc_i)^2\{E[R_{1i}^tR_{1i}]\}^2 = (c_i^tc_i)^2\{\mathrm{tr}(P_iE[UU^t])\}^2,$$

where $P_i = (X^t\Sigma^{-1}X)^{-1}X_i^tD_i^2(A+D_i)^{-2}X_i(X^t\Sigma^{-1}X)^{-1}$. Since $E[UU^t] = O_p(m)$, $P_i = O(m^{-2})$ from $(X^t\Sigma^{-1}X)^{-1} = O(m^{-1})$, and dimensions of c_i, P_i, and U are constant, we have $E(c_i^tR_{1i})^4 = O(m^{-2})$.

To evaluate the fourth moment of rest of the terms in $Q(Y)$, we need to evaluate the moment of $G_i \equiv (\widehat{A}+D_i)^{-1} - (A+D_i)^{-1}$ and $H \equiv \widehat{\Sigma}^{-1} - \Sigma^{-1}$. To see this, we have, for example,

$$
\begin{aligned}
R_{2i} = {} & D_i G_i X_i K X^t H(v+\varepsilon) + D_i G_i X_i K X^t \Sigma^{-1}(v+\varepsilon) \\
& + D_i G_i X_i (X^t \Sigma^{-1} X)^{-1} X^t H(v+\varepsilon) + D_i G_i X_i (X^T \Sigma^{-1} X)^{-1} X^t \Sigma^{-1}(v+\varepsilon) \\
& + D_i (A+D_i)^{-1} X_i K X^t H(v+\varepsilon) + D_i (A+D_i)^{-1} X_i K X^t \Sigma^{-1}(v+\varepsilon) \\
& + D_i (A+D_i)^{-1} X_i (X^t \Sigma^{-1} X)^{-1} X^t H(v+\varepsilon),
\end{aligned}
$$

where $K = (X^t \widehat{\Sigma}^{-1} X)^{-1} - (X^t \Sigma^{-1} X)^{-1}$. Note that the evaluation of $(\widehat{A}+D_i)^{-1} - (A+D_i)^{-1}$ and $\widehat{\Sigma}^{-1} - \Sigma^{-1}$ involves asymptotic expansions of matrix entries. As pointed out by Chatterjee et al. (2008, p. 1240), this computation involves numerous elementary calculations. In the end, both fourth moments reduce to the fourth moment of the Frobenius norm of $\widehat{A} - A$, which is $O(m^{-2})$. We omit these details and refer to Chatterjee et al. (2008).

REFERENCES

Benavent, R., and D. Morales (2016), "Multivariate Fay–Herriot Models for Small Area Estimation," *Computational Statistics & Data Analysis*, 94, 372–390.

Butar, F., and P. Lahiri (2003), "On the Measures of Uncertainty of Empirical Bayes Small-Area Estimators," *Journal of Statistical Planning and Inference*, 112, 63–76.

Chatterjee, S., P. Lahiri, and H. Li (2008), "Parametric Bootstrap Approximation to the Distribution of Eblup and Related Prediction Intervals in Linear Mixed Models," *The Annals of Statistics*, 36, 1221–1245.

Cox, D. (1975), "Prediction Intervals and Empirical Bayes Confidence Intervals," in *Perspectives in Probability and Statistics. Papers in Honor of M. S. Bartlett*, ed. J. Gani, pp. 47–55, Sheffield: Applied Probability Trust, Univ. Sheffield.

Datta, G., R. Fay, and M. Ghosh (1991), "Hierarchical and Empirical Multivariate Bayes Analysis in Small Area Estimation," Proceedings of 1991 ARC of Census Bureau, pp. 63–79.

Datta, G., P. Lahiri, and T. Maiti (2002a), "Empirical Bayes Estimation of Median Income of Four- Person Families by State Using Tie Series and Cross-Sectional Data," *Journal of Statistical Planning and Inference*, 102, 83–97.

Datta, G. S., M. Ghosh, D. D. Smith, and P. Lahiri (2002b), "On an Asymptotic Theory of Conditional and Unconditional Coverage Probabilities of Empirical Bayes Confidence Intervals," *Scandinavian Journal of Statistics*, 29, 139–152.

Fay, R. (1987), "Application of Multivariate Regression to Small Domain Estimation," in *Small Area Statistics*, eds. R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh, pp. 91–102, New York: Wiley.

Fay, R., and R. Herriot (1979), "Estimators of Income for Small Area Places: An Application of James–Stein Procedures to Census," *Journal of the American Statistical Association*, 74, 269–353.

Ito, T., and T. Kubokawa (2021), "Corrected Empirical Bayes Confidence Region in a Multivariate Fay–Herriot Model," *Journal of Statistical Planning and Inference*, 211, 12–32.

Jiang, J., and P. Lahiri (2006), "Mixed Model Prediction and Small Area Estimation, Editor's Invited Discussion Paper," *Test*, 15, 1–96.

Lahiri, P. (2003a), "On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation," *Statistical Science*, 18, 199–210.

———. (2003b), "A Review of Empirical Best Linear Unbiased Prediction for the Fay–Herriot Small-Area Model," *The Philippine Statistician*, 52, 1–15.

Li, H., and P. Lahiri (2010), "Adjusted Maximum Method for Solving Small Area Estimation Problems," *Journal of Multivariate Analysis*, 101, 882–892.

Rao, J. N. K., and I. Molina (2015), *Small Area Estimation* (2nd ed.), New York: Wiley.

Rao, Y., and M. Yu (1994), "Small Area Estimation by Combining Time Series and Cross-Sectional Data," *Canadian Journal of Statistics*, 22, 511–528.