Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Confidence interval estimation of small area parameters shrinking both means and variances

by Sarat C. Dass, Tapabrata Maiti, Hao Ren and Samiran Sinha

December 2012

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca,

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

To access this product

This product, Catalogue no. 12-001-X, is available free in electronic format. To obtain a single issue, visit our website, www.statcan.gc.ca, and browse by "Key resource" > "Publications."

This product is also available as a standard printed publication at a price of CAN$30.00 per issue and CAN$58.00 for a one-year subscription.

The following additional shipping charges apply for delivery outside Canada:

	Single issue	Annual subscription
United States	CAN$6.00	CAN$12.00
Other countries	CAN$10.00	CAN$20.00

All prices exclude sales taxes.

The printed version of this publication can be ordered as follows:

- Telephone (Canada and United States) 1-800-267-6677
- Fax (Canada and United States) 1-877-287-4369
- E-mail infostats@statcan.gc.ca
- Mail Statistics Canada
Finance
R.H. Coats Bldg., 6th Floor
150 Tunney's Pasture Driveway
Ottawa, Ontario K1A 0T6
- In person from authorized agents and bookstores.

When notifying us of a change in your address, please provide both old and new addresses.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Published by authority of the Minister responsible for Statistics Canada

Cette publication est aussi disponible en français.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Survey Methodology, December 2012
Vol. 38, No. 2, pp. 173-187
Statistics Canada, Catalogue No. 12-001-X

Confidence interval estimation of small area parameters shrinking both means and variances

Sarat C. Dass, Tapabrata Maiti, Hao Ren and Samiran Sinha [1]

Abstract

We propose a new approach to small area estimation based on joint modelling of means and variances. The proposed model and methodology not only improve small area estimators but also yield "smoothed" estimators of the true sampling variances. Maximum likelihood estimation of model parameters is carried out using EM algorithm due to the non-standard form of the likelihood function. Confidence intervals of small area parameters are derived using a more general decision theory approach, unlike the traditional way based on minimizing the squared error loss. Numerical properties of the proposed method are investigated via simulation studies and compared with other competitive methods in the literature. Theoretical justification for the effective performance of the resulting estimators and confidence intervals is also provided.

Key Words: EM algorithm; Empirical Bayes; Hierarchical models; Rejection sampling; Sampling variance; Small area estimation.

1. Introduction

Small area estimation and related statistical techniques have become a topic of growing importance in recent years. The need for reliable small area estimates is felt by many agencies, both public and private, for making useful policy decisions. An example where small area techniques are used in practice is in the monitoring of socio-economic and health conditions of different age-sex-race groups where the patterns are observed over small geographical areas.

It is now widely recognized that direct survey estimates for small areas are usually unreliable due to their typically large standard errors and coefficients of variation. Hence, it becomes necessary to obtain improved estimates with higher precision. Model-based approaches, either explicit or implicit, are elicited to connect the small areas and improved precision is achieved by "borrowing strength" from similar areas. The estimation technique is also known as shrinkage estimation since the direct survey estimates are shrunk towards the overall mean. The survey based direct estimates and sample variances are the main ingredients for building aggregate level small area models. The typical modeling strategy assumes that the sampling variances are known while a suitable linear regression model is assumed for the means. For details of these developments, we refer to reader to Ghosh and Rao (1994), Pfeffermann (2002) and Rao (2003). The typical area level models are subject to two main criticisms. First, in practice, the sampling variances are estimated quantities, and hence, are subject to substantial errors. This is because they are often based on equivalent sample sizes from which the direct estimates are calculated. Second, the assumption of known and fixed sampling variances of typical small area models does not take into account the uncertainty in the variance estimation into the overall inference strategy.

Previous attempts have been made to model only the sampling variances; see, for example, Maples, Bell and Huang (2009), Gershunskaya and Lahiri (2005), Huff, Eltinge and Gershunskaya (2002), Cho, Eltinge, Gershunskaya and Huff (2002),Valliant (1987) and Otto and Bell (1995). The articles Wang and Fuller (2003) and Rivest and Vandal (2003) extended the asymptotic mean square error (MSE) estimation of small area estimators when the sampling variances are estimated as opposed to the standard assumption of known variances. Additionally, You and Chapman (2006) considered the modelling of the sampling variances with inference using full Bayesian estimation techniques.

The necessity of variance modelling has been felt by many practitioners. The latest developments in this area are nicely summarized in a recent article by William Bell of the United States Census Bureau 2008. He carefully examined the consequences of these issues in the context of MSE estimation of model based small area estimators. He also provided numerical evidence of MSE estimation for Fay-Herriot models (given in Equation 1) when sampling variances are assumed to be known. The developments in the small area literature so far can be "loosely" viewed as (i) smoothing the direct sampling error variances to obtain more stable variance estimates with low bias and (ii) (partial) accounting of the uncertainty in sampling variances by extending the Fay-Herriot model.

As evident, lesser or no attention has been given to account for the sampling variances effectively while modeling the mean compared to the volume of research that has been done for modeling and inferring the means. There is a lack of systematic development in the small area literature that

1. Sarat C. Dass and Tapabrata Maiti, Department of Statistics & Probability, Michigan State University. E-mail: maiti@stt.msu.edu; Hao Ren, CTB/McGraw-Hill, 20 Ryan Ranch Rd, Monterey, CA 93940; Samiran Sinha, Department of Statistics, Texas A & M University.

includes "shrinking" both means and variances. In other words, we like to exploit the technique of "borrowing strength" from other small areas to "improve" variance estimates as we do to "improve" the small area mean estimates. We propose a hierarchical model which uses both the direct survey and sampling variance estimates to infer all model parameters that determine the stochastic system. Our methodological goal is to develop the dual "shrinkage" estimation for both the small area means and variances, exploiting the structure of the mean-variance joint modelling so that the final estimators are more precise. Numerical evidence shows the effectiveness of dual shrinkage on small area estimates of the mean in terms of the MSE criteria.

Another major contribution of this article is to obtain confidence intervals of small area means. The small area literature is dominated by point estimates and their associated standard errors; it is well known that the standard practice of [point estimate $\pm\, q \times$ standard error], where q is the Z (standard normal) or t cut-off point, does not produce accurate coverage probabilities of the intervals; see Hall and Maiti (2006) and Chatterjee, Lahiri and Li (2008) for more details. Previous work is based on the bootstrap procedure and has limited use due to the repeated estimation of model parameters. We produce confidence intervals for the means from a decision theory perspective. The construction of confidence intervals is easy to implement in practice.

The rest of the article is organized as follows. The proposed hierarchical model for the sample means and variances is developed in Section 2. The estimation of model parameters via the EM algorithm is developed in Section 3. Theoretical justification for the proposed confidence interval and coverage properties are presented in Section 4. Sections 5 and 6 present a simulation study and a real data example, respectively. Some discussion and concluding remarks are presented in Section 7. An alternative model formulation for small area as well as mathematical details are provided in the Appendix.

2. Proposed model

Suppose n small areas are in consideration. For the i^{th} small area, let (X_i, S_i^2) be the pair of direct survey estimate and sampling variance, for $i = 1, 2, \ldots, n$. Let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ip})^T$ be the vector of p covariates available at the estimation stage for the i^{th} small area. We propose the following hierarchical model:

$$\left.\begin{aligned} X_i \mid \theta_i, \sigma_i^2 &\sim \text{Normal}(\theta_i, \sigma_i^2) \\ \theta_i &\sim \text{Normal}(\boldsymbol{Z}_i^T \beta, \tau^2) \end{aligned}\right\} \qquad (1)$$

$$\left.\begin{aligned} \frac{(n_i - 1) S_i^2}{\sigma_i^2} \Bigg| \sigma_i^2 &\sim \chi^2_{n_i - 1} \\ \sigma_i^{-2} &\sim \text{Gamma}(a, b), \end{aligned}\right\} \qquad (2)$$

independently for $i = 1, 2, \ldots, n.$ In the model elicitation, n_i is the sample size for a simple random sample (SRS) from the i^{th} area, $\beta = (\beta_1, \ldots, \beta_p)^T$ is the $p \times 1$ vector of regression coefficients, and $\boldsymbol{B} \equiv (a, b, \beta, \tau^2)^T$ is the collection of all unknown parameters in the model. Also, $\text{Gamma}(a, b)$ is the Gamma density function with positive shape and scale parameters a and b, respectively, defined as $f(x) = \{b^a \Gamma(a)\}^{-1} e^{-x/b} x^{a-1}$ for $x > 0$, and 0 otherwise. The unknown σ_i^2 is the true variance of X_i and is usually estimated by the sample variance S_i^2. Although S_i^2's are assumed to follow a chi-square distribution with $(n_i - 1)$ degrees of freedom (as a result of normality and SRS), we note that for complex survey designs, the degree of freedom needs to be determined carefully [*e.g.*, Maples *et al.* 2009). More importantly, the role of the sample sizes in shrinkage estimation of σ_i^2 is as follows: For low values of n_i, the estimate of σ_i^{-2} is shrunk more towards the overall mean (ab) compared to higher n_i values. Thus, for variances, sample sizes play the same role as precision in shrinkage estimation of the small area mean estimates. We note that You and Chapman (2006) also considered the second level of the sampling variance modelling. However, the hyper-parameters related to prior of σ_i^2 are not data driven, they are rather chosen in such a way that the prior will be vague. Thus, their model can be viewed as the Bayesian version of the models considered in Rivest and Vandal (2003) and Wang and Fuller (2003). The second level modelling of σ_i^{-2} in (2) can be further extended to $\sigma_i^{-2} \sim \text{Gamma}(b, \exp(\boldsymbol{Z}_i^T \beta_2)/b)$ so that $E(\sigma_i^{-2}) = \exp(\boldsymbol{Z}_i^T \beta_2)$ for another set of p regression coefficients β_2 to accommodate covariate information in the variance modeling.

Although our model is motivated by Hwang, Qiu and Zhao (2009), we like to mention that Hwang *et al.* (2009) considered shrinking means and variances in the context of microarray data where they prescribed an important solution by plugging in a shrinkage estimator of variance into the mean estimator. The shrinkage estimator of the variance in Hwang *et al.* (2009) is a function of S_i^2 only, and not of both X_i and S_i^2; see Remarks 2 and 3 in Section 2. Thus, inference of the mean does not take into account the full uncertainty in the variance estimation. Further, their model does not include any covariate information. The simulation study described subsequently indicate that our method of estimation performed better than Hwang *et al.* (2009).

In the above model formulation, inference for the small area mean parameter θ_i can be made based on the conditional distribution of θ_i given all of the data $\{(X_i, S_i^2, \boldsymbol{Z}_i),\ i = 1, \ldots, n\}$. Under our model set up, the conditional distribution of θ_i is a non-standard distribution and does not have a closed form, thus requiring numerical methods, such as Monte Carlo and the EM algorithm, for inference, and the details are provided in the next section.

3. Inference methodology

3.1 Estimation of unknown parameters via EM algorithm

In practice, $\boldsymbol{B} \equiv (a, b, \beta, \tau^2)^T$ is unknown and has to be estimated from the data $\{(X_i, S_i^2, \boldsymbol{Z}_i),\ i = 1, 2, ..., n\}$. Our proposal is to estimate $\boldsymbol{B}$ by the marginal maximum likelihood method: Estimate $\boldsymbol{B}$ by $\hat{\boldsymbol{B}}$ where $\hat{\boldsymbol{B}}$ maximizes the marginal likelihood $L_M(\boldsymbol{B}) = \prod_{i=1}^n L_{M,i}(\boldsymbol{B})$, where

$$L_{M,i} \propto \frac{\Gamma(n_i/2+a)}{\tau\Gamma(a)b^a}\int \exp\left\{-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2}\right\}\psi_i^{-(n_i/2+a)}d\theta_i, \quad (3)$$

and

$$\psi_i \equiv \left\{0.5(X_i - \theta_i)^2 + 0.5(n_i - 1)S_i^2 + \frac{1}{b}\right\}. \quad (4)$$

The marginal likelihood L_M involves integrals that cannot be evaluated in closed-form, and hence, one has to resort to numerical methods for its maximization. One such algorithm is the EM (Expectation-Maximization) iterative procedure which is used when such integrals are present. The EM algorithm involves augmenting the observed likelihood $L_M(\boldsymbol{B})$ with missing data; in our case, the variables of the integration, $\theta_i,\ i = 1, 2, ..., n$, constitute this missing information. Given $\theta \equiv \{\theta_1, \theta_2, ..., \theta_n\}$, the *complete data* log likelihood (ℓ_c) can be written as

$$\ell_c(\boldsymbol{B}, \theta) = \sum_{i=1}^n \Bigg[\log\{\Gamma(n_i/2 + a)\} - \log\{\Gamma(a)\} - a\log(b) - 0.5\log(\tau^2) - \frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2} - (n_i/2 + a)\log(\psi_i)\Bigg],$$

where the expression of ψ_i is given in Equation (4). Starting from an initial value of $\boldsymbol{B}, \boldsymbol{B}^{(0)}$ say, the EM algorithm iteratively performs a maximization with respect to $\boldsymbol{B}$. At the t^{th} step the objective function maximized is

$$Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)}) = E(\ell_c(\boldsymbol{B}, \theta)) = \sum_{i=1}^n \Bigg[\log\{\Gamma(n_i/2 + a)\} - \log\{\Gamma(a)\} - a\log(b) - 0.5\log(\tau^2) - \frac{E(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2} - (n_i/2 + a)E\{\log(\psi_i)\}\Bigg].$$

The expectation in $Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})$ is taken with respect to the conditional distribution of each θ_i given the data, $\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}^{(t-1)})$, which is

$$\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}) \propto \exp\{-0.5(\theta_i - \boldsymbol{Z}_i^T\beta)^2/\tau^2\}\psi_i^{-(n_i/2+a)}. \quad (5)$$

One challenge here is that the expectations are not available in closed form. Thus, we resort to a Monte carlo method for evaluating the expressions. Suppose that R iid samples of θ_i are available, say $\theta_{i,1}, \theta_{i,2}, ..., \theta_{i,R}$. Then, each expectation of the form $E\{h(\theta_i)\}$ can be approximated by the Monte Carlo mean

$$E\{h(\theta_i)\} \approx \frac{1}{R}\sum_{r=1}^R h(\theta_{i,k}). \quad (6)$$

However, drawing random numbers from the conditional distribution $\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}^{(t-1)})$ is also not straightforward since this is not a standard density. Samples are drawn using the accept-reject procedure (Robert and Casella 2004): For a sample from the target density f, sample x from the proposal density g, and accept the sample as a sample from f with probability $f(x)/\{M^* g(x)\}$ where $M^* = \sup_x\{f(x)/g(x)\}$. One advantage of the accept-reject method is that the target density f only needs to be known upto a constant of proportionality which is the case for $\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}^{(t-1)})$ in (5); due to the non-standard form of the density, the normalizing constant cannot be found in a closed form. For the accept-reject algorithm, we used the normal density $g(\theta_i) \propto \exp\{-0.5(\theta_i - \boldsymbol{Z}_i^T\beta)^2/\tau^2\}$ as the proposal density. The acceptance probability is calculated to be $[\{1/b + 0.5(n_i - 1)S_i^2\}/\{1/b + 0.5(n_i - 1)S_i^2 + 0.5(\theta_i - X_i)^2\}]^{n_i/2+a}$. One can choose a better proposal distribution to increase acceptance probability or different algorithm (such as the adaptive rejection sampling or envelope accept-reject algorithms) but our chosen proposal worked satisfactorily in the studies we conducted.

The maximizer of $Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})$ at the t^{th} step can be described explicitly. The solutions for β and τ^2 are available in closed form as

$$\beta^{(t)} = \left(\sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{Z}_i^T\right)^{-1}\left(\sum_{i=1}^n \boldsymbol{Z}_i E(\theta_i)\right)$$

and

$$(\tau^2)^{(t)} = \frac{1}{n}\sum_{i=1}^n E(\theta_i - \boldsymbol{Z}_i^T\beta)^2,$$

respectively. Also, $a^{(t)}$ and $b^{(t)}$ are obtained by solving $S_a = \partial Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})/\partial a = 0$ and $S_b = \partial Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})/\partial b = 0$ using the Newton-Raphson method where

$$S_a = \sum_{i=1}^{n} \frac{\partial}{\partial a} \log\{\Gamma(n_i/2 + a)\} - n\left\{\frac{\partial}{\partial a}\log\{\Gamma(a)\right\} - n\log(b) - \sum_{i=1}^{n} E\{\log(\psi_i)\}$$

and

$$S_b = -\frac{na}{b} + \sum_{i=1}^{n} \frac{(n_i/2 + a)}{b^2} E(\psi_i^{-1}).$$

We set $\boldsymbol{B}^{(t)} = (a^{(t)}, b^{(t)}, \beta^{(t)}, (\tau^{(t)})^2)$ and proceed to the $(t+1)$-st step. This maximization procedure is repeated until the estimate $\boldsymbol{B}^{(t)}$ converges. The MLE of $\boldsymbol{B}$, $\hat{\boldsymbol{B}} = \boldsymbol{B}^{(\infty)}$, once convergence is established.

3.2 Point estimate and confidence interval for θ_i

Following the standard technique, the small area estimator of θ_i is taken to be

$$\hat{\theta}_i = E(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})\Big|_{\boldsymbol{B}=\hat{\boldsymbol{B}}}, \tag{7}$$

the expectation of θ_i with respect to the conditional density $\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})$ with the maximum likelihood estimate $\hat{\boldsymbol{B}}$ plugged in for $\boldsymbol{B}$. The estimate $\hat{\theta}_i$ is calculated numerically using the Monte Carlo procedure (6) described in the previous section. Subsequently, all quantities involving the unknown $\boldsymbol{B}$ will be plugged in by $\hat{\boldsymbol{B}}$ although we still keep using the notation $\boldsymbol{B}$ for simplicity.

Further, we develop a confidence interval for θ_i based on a decision theory approach. Following Joshi (1969), Casella and Hwang (1991), Hwang *et al.* (2009), consider the loss function associated with the confidence interval C given by $(k/\sigma)L(C) - I_C(\theta)$ where k is a tuning parameter independent of the model parameters, $L(C)$ is the length of C and $I_C(\theta)$ is the indicator function taking values 1 or 0 depending on whether $\theta \in C$ or not. Note that this loss function takes into account both the coverage probability as well as the length of the interval; the positive quantity (k/σ) serves as the relative weight of the length compared to the coverage probability of the confidence interval. If $k = 0$, the length of the interval is not under consideration, which leads to the optimal C to be $(-\infty, \infty)$ with coverage probability 1. On the other hand, if $k = \infty$, then the coverage probability is 0, leading to optimal C to be a point set. The Bayes confidence interval for θ_i is obtained by minimizing the risk function (the expected loss) $E\{[(k/\sigma)L(C) - I_C(\theta)] \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})\}$. The optimal choice of C is given by

$$C_i(\boldsymbol{B}) = \{\theta_i\colon kE(\sigma_i^{-1} \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}) < \pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})\}. \tag{8}$$

Since $C_i(\boldsymbol{B})$ is obtained by minimizing the posterior risk, one may like to interpret this as a Bayesian credible set. However, following Casella and Berger (1990, page 470), we will continue naming $C_i(\boldsymbol{B})$ as a confidence interval. From an empirical Bayes perspective also, this terminology is more appropriate. How the tuning parameter k determines the confidence level of $C_i(\boldsymbol{B})$ will be shown explicitly in Section 3.3.

Assuming k is known for the moment, we follow the steps below to calculate $C_i(\boldsymbol{B})$. The conditional densities of σ_i^2 and θ_i are given by

$$\pi(\sigma_i^2 \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}) \propto \frac{\exp\left[\dfrac{-0.5(X_i - \boldsymbol{Z}_i^T\beta)^2}{(\sigma_i^2 + \tau^2) - \left\{0.5(n_i - 1)S_i^2 + \dfrac{1}{b}\right\}\left(\dfrac{1}{\sigma_i^2}\right)}\right]}{(\sigma_i^2)^{(n_i-1)/2+a+1}(\sigma_i^2 + \tau^2)^{1/2}} \tag{9}$$

and (5), respectively, which as mentioned before, are not available in closed form. Thus, similar to the case of θ_i, $E(\sigma_i^{-1} \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})$ is computed numerically using the Monte Carlo method by approximating the expected value with the mean $1/N \sum_{k=1}^{N} 1/\sigma_{i,k}$ where $\sigma_{i,r}^2$, $r = 1, 2, \ldots, R$ are R samples from the conditional density $\pi(\sigma_i^2 \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})$. The accept reject procedure is used to draw random numbers from $\pi(\sigma_i^2 \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})$ with a proposal density given by the inverse Gamma

$$\frac{\exp\left[-\left\{0.5(n_i - 1)S_i^2 + \dfrac{1}{b}\right\}\left(\dfrac{1}{\sigma_i^2}\right)\right]}{(\sigma_i^2)^{(n_i-1)/2+a+1}},$$

and the acceptance probability

$$\frac{\exp\left\{\dfrac{-0.5(X_i - \boldsymbol{Z}_i^T\beta)^2}{(\sigma_i^2 + \tau^2)}\right\}}{(\sigma_i^2 + \tau^2)^{1/2}} \times \exp(0.5) \times |X_i - \boldsymbol{Z}_i^T\beta|.$$

The next step is to determine the boundary values of $C_i(\boldsymbol{B})$ by finding two θ_i values that satisfy the equation $kE(\sigma_i^{-1} \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}) - \pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}) = 0$. This requires the normalizing constant in (5)

$$D_i = \int_{-\infty}^{\infty} \exp\{-0.5(\theta_i - \boldsymbol{Z}_i^T\beta)^2/\tau^2\}\,\psi_i^{-(n_i/2+a)}\,d\theta_i$$

to be evaluated numerically. This is obtained using the Gauss-Hermite integration with 20 nodes.

3.3 Choice of k

The choice of the tuning parameter k in (8) is taken to be

$$k = k(\boldsymbol{B}) = u_{i,0}\,\phi\left(t_{\alpha/2}\sqrt{\frac{n_i + 2a + 2}{n_i - 1}}\right) \qquad (10)$$

where ϕ is the standard normal distribution, $t_{\alpha/2}$ is $(1 - \alpha/2)^{\text{th}}$ percentile of t distribution with $(n_i - 1)$ degrees of freedom, and $u_{i,0} = \sqrt{1 + \sigma_i^2/\tau^2}$. Since $u_{i,0}$ involves σ_i^2 which is unknown, an estimated version $\hat{u}_{i,0}$ is obtained by plugging in the maximum a posteriori estimate

$$\hat{\sigma}_i^2 = \hat{\sigma}_i^2(\hat{\boldsymbol{B}}) = \arg\max_{\sigma_i^2} \pi(\sigma_i^2 \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B})\Big|_{\boldsymbol{B}=\hat{\boldsymbol{B}}} \qquad (11)$$

in place of σ_i^2. Also, $\boldsymbol{B}$ is replaced by $\hat{\boldsymbol{B}}$ in (11). We demonstrate that the coverage probability of $C_i(\hat{\boldsymbol{B}})$ with this choice of k is close to $1 - \alpha$. Theoretical justifications are provided in Section 4.

3.4 Other related methods for comparison

Our method will be denoted as Method I. Three other methods to be compared are briefly described below.

Method II: Wang and Fuller (2003) considered the Fay-Herriot small area estimation model given by (1). Their primary contribution is the construction of the mean squared error estimation formulae for small area estimators with estimated sampling variances. In the process, they had constructed two formulae denoted by $\widehat{\text{MSE}}_1$ and $\widehat{\text{MSE}}_2$. We use $\widehat{\text{MSE}}_1$ for our comparisons, which was derived following the bias correction approach of Prasad and Rao (1990). The basic difference with our approach is that they did not smooth the sampling variances, only taking the uncertainty into account while making inference on the small area parameters. The method of parameter estimation, which is moment based for all the model parameters, is also different from ours.

Method III: Hwang *et al.* (2009) considered the log-normal and inverse Gamma models for σ_i^{-2} in (2) for microarray data analysis. Their simulation study showed improved performance of confidence intervals for small area estimators under the log-normal model compared to the inverse gamma. We thus modified their log-normal model to add covariates and for unequal sample sizes n_i as follows:

$$\left.\begin{aligned} X_i \mid \theta_i, \sigma_i^2 &\sim \text{Normal}(\theta_i, \sigma_i^2) \\ \theta_i &\sim \text{Normal}(\boldsymbol{Z}_i^T\boldsymbol{\beta}, \tau^2); \end{aligned}\right\} \qquad (12)$$

$$\left.\begin{aligned} \log S_i^2 = \log(\sigma_i^2) + \delta_i;\ \delta_i &\sim N(m_i, \sigma_{ch,i}^2) \\ \log(\sigma_i^{-2}) &\sim N(\mu_v, \tau_v^2), \end{aligned}\right\} \qquad (13)$$

independently for $i = 1, 2, \ldots, n$. Note that the model for the means in (12) is identical to (1). The quantities τ^2, m_i and $\sigma_{ch,i}^2$ are assumed to be known and are given by $m_i = E[\log(\chi_{n_i-1}^2/(n_i - 1))]$ and $\sigma_{ch,i}^2 = \text{Var}[\log(\chi_{n_i-1}^2/(n_i - 1))]$. Thus, the sample size n_i's determine the shape of the χ^2 distribution via its degrees of freedom parameter. More importantly, as mentioned earlier, the different sample sizes account for different degrees of shrinkage for the corresponding true variance parameter. Similar to their estimation approach, the unknown model parameters μ_v and τ_v^2 are estimated using a moment based approach in an empirical Bayes framework giving $\hat{\mu}_v$ and $\hat{\tau}_v^2$, respectively. Note that in Hwang *et al.* (2009), these estimates are obtained based on the hierarchical model for σ_i^2 of (13) *only* without regard to the modelling (1) of the mean. We refer to the Section 5 of their paper for details of the estimation of the hyper-parameters. We follow the same procedure using only (13) to estimate μ_v and τ_v^2 in the case of unequal sample sizes.

The Bayes estimate of σ_i^2 is derived to be

$$\begin{aligned} \hat{\sigma}_{i,B}^2 &= \exp\left[E\{\ln(\sigma_i^2) \mid \ln(S_i^2)\}\right] \\ &= \left\{\frac{S_i^2}{\exp(m_i)}\right\}^{M_{v,i}} \exp\{\mu_v(1 - M_{v,i})\} \end{aligned}$$

where $M_{v,i} = \tau_v^2/(\tau_v^2 + \sigma_{ch,i}^2)$ and with estimates plugged in for the unknown quantities. The conditional distribution of θ_i given (X_i, S_i^2), is

$$\pi(\theta_i \mid X_i, S_i^2) = \int_0^\infty \pi(\theta_i \mid X_i, S_i^2, \sigma_i^2)\, \pi(\sigma_i^2 \mid X_i, S_i^2)\, d\sigma_i^2,$$

is approximated as $\pi(\theta_i \mid X_i, S_i^2) \approx \int_0^\infty \pi(\theta_i \mid X_i, S_i^2, \hat{\sigma}_{i,B}^2)$ $\pi(\sigma_i^2 \mid X_i, S_i^2)\, d\sigma_i^2 = \pi(\theta_i \mid X_i, S_i^2, \hat{\sigma}_{i,B}^2)$. This suggests the approximate Bayes estimator of the small area parameters given by

$$\hat{\theta}_i = E(\theta_i \mid X_i, \hat{\sigma}_{i,B}^2) = \hat{M}_i X_i + (1 - \hat{M}_i)\boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}, \qquad (14)$$

where $\hat{M}_i = \hat{\tau}_v^2/(\hat{\tau}_v^2 + \hat{\sigma}_{i,B}^2)$. The confidence interval for θ_i is obtained as

$$C_i^H = \left\{\theta_i : \frac{|\theta_i - \hat{\theta}_i|}{\hat{M}_i \hat{\sigma}_{i,B}^2} < -2\ln\{k\sqrt{2\pi}\} - \ln(\hat{M}_i)\right\}. \qquad (15)$$

In Section 3 of Hwang *et al.* (2009) pages 269-271, the interval C_i^H is matched with the $100(1 - \alpha)\%$ t-interval $[|\theta_i - X_i| < tS_i]$ to obtain the expression of k as $k \equiv k_i = \exp\{-t^2/2\}\exp\{m_i/2\}/(\sqrt{2\pi})$.

Method IV: This method comprises of a special case of the Fay-Herriot model in (1) but with the estimation of model parameters adopted from Qiu and Hwang (2007). Qiu and Hwang (2007) considered the model

$$\left.\begin{aligned} X_i \mid \theta_i, \sigma^2 &\sim \text{Normal}(\theta_i, \sigma^2) \\ \theta_i &\sim \text{Normal}(0, \tau^2), \end{aligned}\right\} \qquad (16)$$

independently for $i = 1, 2, ..., n,$ for analyzing microarray experimental data. When model parameters are known, they proposed the point estimator $\hat{\theta}_i = \hat{M}X_i, \hat{M} = (1 - ((n-2)\sigma^2 / |X|^2))_+$ where a_+ denotes $\max(0, a)$ for any number a and $|X| = (\sum_{i=1}^n X_i^2)^{1/2}$. The confidence interval for θ_i is $\hat{\theta}_i \pm v_1(\hat{M})$, where $v_1^2(\hat{M}) = \sigma^2 \hat{M}(q_1 - \ln(\hat{M}))$ with q_1 denoting the standard normal cut-off point corresponding to desired level of confidence coefficient and $v_1(0) \equiv 0$. Here For the purpose of comparisons with our method, the first level of the hierarchical model in (16) is modified as follows:

$$X_i = \boldsymbol{Z}_i^T \beta + v_i + e_i$$

where $v_i \sim \text{Normal}(0, \tau^2)$ and $e_i \sim \text{Normal}(0, S_i^2)$ independently for $i = 1, 2, ..., n,$ and S_i^2 is treated as known. Following Qiu and Hwang (2007), τ^2 is estimated by

$$\tilde{\tau}^2 = \frac{1}{n-p}\left[\sum_i \hat{u}_i^2 - \sum_i S_i^2 \left\{1 - \boldsymbol{Z}_i^T \left(\sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{Z}_i^T\right)^{-1} \boldsymbol{Z}_i^T\right\}\right]$$

and $\hat{\tau}^2 = \max(\tilde{\tau}^2, 1/n)$ where $\hat{u}_i = X_i - \boldsymbol{Z}_i^T \hat{\beta}$ and $\hat{\beta} = (\sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{Z}_i^T)^{-1} (\sum_{i=1}^n \boldsymbol{Z}_i X_i)$. Next, define $\hat{M}_{0i} = \hat{\tau}^2 / (\hat{\tau}^2 + S_i^2)$ and $\hat{M}_i = \max(\hat{M}_{0i}, M_1)$ where in the latter expression, $\hat{M}_{0i}$ is truncated by $M_{1i} = 1 - Q_\alpha / (n_i - 2)$, and Q_α is the α^{th} quantile of a chi-squared distribution with n_i degrees of freedom. This $\hat{M}_i$ is used in the formula of the confidence interval for θ_i given earlier. When applying this method in our simulation study and real data analysis, we modified the model to accommodate such unequal sample sizes and covariate information mentioned earlier.

Remark 1. Hwang *et al.* (2009) choose k by equating (15) to the t interval based on only X_i for the small area parameters θ_i. Note that X_i is the direct survey estimator. Consequently, this choice of k does not have any direct control over the coverage probability of the interval constructed under *shrinkage estimation*. On the other hand, our proposed choice of k has been derived to maintain nominal coverage under, specifically, shrinkage estimation.

Remark 2. Note that without any hierarchical modelling assumption, S_i and X_i are independent as S_i^2 and X_i are, respectively, ancillary and the complete sufficient statistics for θ_i. However, under models (1) and (2) the conditional distribution of σ_i^2 and θ_i involve both X_i and S_i^2 which is seen from (5) and (9).

Remark 3. In Hwang *et al.* (2009), the shrinkage estimator for σ_i^2 is based only on the information on S_i^2, and not of both X_i and S_i^2. The Bayes estimator of σ_i^2 is plugged into the expression for the Bayes estimator of small area parameters. Thus, Hwang *et al.*'s small area estimator is written as $E(\theta_i \mid X_i, \hat{\sigma}_{i,B}^2)$ in (14) where $\hat{\sigma}_{i,B}^2$ is the Bayes estimator of σ_i^2. Due to equation (9), the shrinkage estimator of σ_i^2 depends on $(X_i - \boldsymbol{Z}_i^T \beta)^2$ in addition to S_i^2 in contrast to Hwang *et al.* (2009). We believe this could be the reason for improved performance of our method compared to Hwang *et al.* (2009).

Remark 4. As mentioned previously, the degree of freedom associated with the χ^2 distribution for the sampling variance need not to be simply $n_i - 1$, n_i being the sample size for i^{th} area. There is no sound theoretical result for determining the degree of freedom when the survey design is complex. The article Wang and Fuller (2003) approximated the χ^2 with a normal based on the Wilson-Hilferty approximation. If one knows the exact sampling design then the simulation based guideline of Maples *et al.* (2009) could be useful. For county level estimation using the American Community Survey, Maples *et al.* (2009) suggested the estimated degrees of freedom of $0.36 \times \sqrt{n_i}$.

4. Theoretical justification

Theoretical justification for the choice of k according to equation (10) is presented in this section. As in Hwang *et al.* (2009), the conditional distribution of θ_i given X_i and S_i^2 can be approximated as $\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{B}) \approx \pi(\theta_i \mid X_i, S_i^2, \boldsymbol{B}, \hat{\sigma}_i^2)$, where $\hat{\sigma}_i^2$ as defined in (11). In a similar way, approximate $E(\sigma_i^{-1} \mid X_i, S_i^2, \boldsymbol{B})$ by $E(\sigma_i^{-1} \mid X_i, S_i^2, \boldsymbol{B}) \approx \hat{\sigma}_i^{-1}$. Based on these approximations, we have $C_i(\boldsymbol{B}) \approx \tilde{C}_i(\boldsymbol{B})$ where $\tilde{C}_i(\boldsymbol{B})$ is the confidence interval for θ_i given by $\tilde{C}_i(\boldsymbol{B}) = \{\theta_i : \pi(\theta_i \mid X_i, S_i^2, \boldsymbol{B}, \hat{\sigma}_i^2) \geq k\hat{\sigma}_i^{-1}\}$. From (1), it follows that the conditional density $\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{B}, \sigma_i^2)$ is a normal with mean μ_i and variance v_i, where μ_i and v_i are given by the expressions

$$\mu_i = w_i X_i + (1 - w_i) \boldsymbol{Z}_i^T \beta,$$

$$v_i = \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2}\right)^{-1} = \sigma_i^2 \left(1 + \frac{\sigma_i^2}{\tau^2}\right)^{-1}, \tag{17}$$

and

$$w_i = \frac{1/\sigma_i^2}{(1/\sigma_i^2 + 1/\tau^2)}.$$

Now, choosing

$$k = \hat{u}_0 \phi\left(t_{\alpha/2} \sqrt{\frac{n_i + 2a + 2}{n_i - 1}}\right)$$

as discussed, the confidence interval $\tilde{C}_i(\boldsymbol{B})$ becomes

$$\tilde{C}_i(\boldsymbol{B}) = \left\{\theta_i : \hat{u}_{0i} \frac{|\theta_i - \hat{\mu}_i|}{\hat{\sigma}_i} \leq t_{\alpha/2} \sqrt{\frac{n_i + 2a + 2}{n_i - 1}}\right\}, \tag{18}$$

where $\hat{\mu}_i$ is the expression for μ_i in (17) with σ_i^2 replaced by $\hat{\sigma}_i^2$. Now consider the behavior of $\hat{\sigma}_i^2 \equiv \hat{\sigma}_i^2(\boldsymbol{B})$ as τ^2 ranges between 0 and ∞. When $\tau^2 \to \infty$, $\hat{\sigma}_i^2$ converges to

$$\hat{\sigma}_i^2(\infty) \equiv \hat{\sigma}_i^2(a, b, \beta, \infty) = \frac{\frac{(n_i - 1)}{2} S_i^2 + \frac{1}{b}}{\frac{n_i - 1}{2} + a + 1} = \frac{(n_i - 1) S_i^2 + \frac{2}{b}}{n_i + 2a + 1}.$$

Similarly, when $\tau^2 \to 0$, $\hat{\sigma}_i^2$ converges to

$$\hat{\sigma}_i^2(0) \equiv \hat{\sigma}_i^2(a, b, \beta, 0) = \frac{(X_i - \boldsymbol{Z}_i^T \beta)^2 + (n_i - 1) S_i^2 + \frac{2}{b}}{n_i + 2a + 2}.$$

For all intermediate values of τ^2, we have $\min\{\hat{\sigma}_i^2(0), \hat{\sigma}_i^2(\infty)\} \le \hat{\sigma}_i^2 \le \max\{\hat{\sigma}_i^2(0), \hat{\sigma}_i^2(\infty)\}$. Therefore, it is sufficient to consider the following two cases: (i) $\hat{\sigma}_i^2 \ge \hat{\sigma}_i^2(\infty)$, where it follows that $(n_i + 2a + 2)\hat{\sigma}_i^2 = (n_i + 2a + 1)\hat{\sigma}_i^2 + \hat{\sigma}_i^2 \ge (n_i - 1) S_i^2 + 2/b + \hat{\sigma}_i^2 \ge (n_i - 1) S_i^2$, and (ii) $\hat{\sigma}_i^2 \ge \hat{\sigma}^2(0)$, where it follows that $(n_i + 2a + 2)\, \hat{\sigma}_i^2 = (X_i - \boldsymbol{Z}_i^T \beta)^2 + (n_i - 1) S_i^2 + 2/b \ge (n_i - 1) S_i^2$. So, in both cases (i) and (ii),

$$(n_i + 2a + 2)\, \hat{\sigma}_i^2 \ge (n_i - 1) S_i^2. \quad (19)$$

Since $\theta_i - \mu_i \sim N(0, \sigma_i^2 \tau^2 / (\sigma_i^2 + \tau^2))$ and $(n_i - 1) S_i^2 / \sigma_i^2 \sim \chi^2_{n_i - 1}$, the confidence interval

$$D_i = \left\{ \theta_i :\, u_{0i} \frac{|\theta_i - \mu_i|}{S_i} \le t_{\alpha/2} \right\} \quad (20)$$

has coverage probability $1 - \alpha$. Thus, if u_0 and μ_i are replaced by $\hat{u}_0$ and $\hat{\mu}_i$, it is expected that the resulting confidence interval $\tilde{D}_i$, say, will have coverage probability of approximately $1 - \alpha$. From (19), we have

$$P\{\tilde{C}_i(\boldsymbol{B})\} \ge P(\tilde{D}_i) \approx 1 - \alpha, \quad (21)$$

establishing an approximate lower bound of $1 - \alpha$ for the confidence level of $\tilde{C}_i(\boldsymbol{B})$.

In (21), $\boldsymbol{B}$ was assumed to be fixed and known. When $\boldsymbol{B}$ is unknown, we replace $\boldsymbol{B}$ by its marginal maximum likelihood estimate $\hat{\boldsymbol{B}}$. Since (21) holds regardless of the true value of $\boldsymbol{B}$, substituting $\hat{\boldsymbol{B}}$ for $\boldsymbol{B}$ in (21) will involve an order $O(1/\sqrt{N})$ of error where $N = \sum_{i=1}^n n_i$. Compared to each single n_i, this pooling of n_i's is expected to reduce the error significantly so that $\tilde{C}_i(\hat{\boldsymbol{B}})$ is sufficiently close to $\tilde{C}_i(\boldsymbol{B})$ to satisfy the lower bound of $1 - \alpha$ in (21).

5. A simulation study

5.1 Simulation setup

We considered a simulation setting using a subset of parameter configurations from Wang and Fuller (2003). Each sample in the simulation study was generated from the following steps: First, generate observations using the model

$$X_{ij} = \beta + u_i + e_{ij},$$

where $u_i \sim N(0, \tau^2)$ and $e_{ij} \sim N(0, n_i \sigma_i^2)$, independently for $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$. Then, the random effects model for the small area mean, X_i, is

$$X_i = \beta + u_i + e_i, \quad \text{independently for } i = 1, \ldots, n,$$

where $X_i \equiv \bar{X}_{i\cdot} \equiv n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ and $e_i \equiv \bar{e}_{i\cdot} \equiv n_i^{-1} \sum_{j=1}^{n_i} e_{ij}$. Therefore, $X_i \sim N(\theta_i, \sigma_i^2)$ where $\theta_i = \beta + u_i$, $\theta_i \sim N(\beta, \tau^2)$ and $e_i \sim N(0, \sigma_i^2)$. We estimated σ_i^2 with the unbiased estimator

$$S_i^2 = (n_i - 1)^{-1} n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2,$$

and it follows that $(n_i - 1) S_i^2 / \sigma_i^2 \sim \chi^2_{n_i - 1}$, independently for $i = 1, 2, \ldots, n$. Note that the simulation layout has ignored the second level modeling of sampling variances in (2). Thus, our result will indicate robustness with respect to the variance model misspecification.

The above steps produced the data $(X_i, S_i^2), i = 1, \ldots, n$. To simplify the simulation, we do not choose any covariate information $\boldsymbol{Z}_i$. Similar to Wang and Fuller (2003), we set all n_i's equal to m to ease programming efforts. However, the true sampling variances are still chosen to be unequal: One-third of the σ_i^2 are set to 1, another one-third are set to 4, and the remaining one-third are set to 16. We take $\beta = 10$ and three different choices of $\tau^2 = 0.25$, 1 and 4. These parameter values are chosen from Qiu and Hwang (2007). For each of τ^2, we generated 200 samples for the two combinations $(m, n) = (9, 36)$ and (18, 180).

In the simulation study, we compare the proposed method with the methods of Wang and Fuller (2003), Hwang *et al.* (2009) and Qiu and Hwang (2007) which are referred to as Methods I, II, III, and IV, respectively, based on bias, mean squared error (MSE), coverage probability (CP) of the confidence intervals and the length of the confidence intervals (ALCI). Table 1 contains the parameter estimates for a, b, β and τ^2. The numerical results indicate good performance of the maximum likelihood estimates for the model parameters; the estimated values of β and τ^2 are close to the true values indicating good robustness properties with respect to distributional misspecification in the second level of (2). Statistically significant estimates for both a and b indicate that "shrunk" sampling variances are incorporated in the proposed method. Tables 2, 3 and 4 provide numerical results averaged over areas within each group having the same true sampling variances. The results in the Tables are based on 200 replications.

Table 1
Simulation results for the model parameters, a (top left panel), b (top right panel), β (bottom left panel) and τ^2 (bottom right panel). Here SD represents the standard deviation over 200 replicates. We took $\beta = 10$ and $\tau^2 = 0.25$, 1 and 4

τ^2	$n = 36,\ m = 9$ Mean	SD	$n = 180,\ m = 18$ Mean	SD	τ^2	$n = 36,\ m = 9$ Mean	SD	$n = 180,\ m = 18$ Mean	SD
		a					b		
0.25	1.0959	0.1540	1.0328	0.0442	0.25	0.3992	0.0983	0.4249	0.0323
1	1.0937	0.1555	1.0325	0.0445	1	0.4030	0.1012	0.4253	0.0326
4	1.0996	0.1577	1.0339	0.0450	4	0.3999	0.1017	0.4245	0.0328
		β					τ^2		
0.25	10.0071	0.3618	9.9951	0.1853	0.25	0.2558	0.0605	0.2575	0.0097
1	10.0142	0.3311	9.9970	0.1743	1	0.9418	0.3333	1.0426	0.1264
4	10.0282	0.4639	10.0048	0.2254	4	3.5592	1.3316	4.0817	0.5551

Table 2
Simulation results for prediction when $\tau^2 = 0.25$. Here MSE, ALCI, CP represent the mean squared error, average confidence interval width, and coverage probability, respectively

		$n = 36,\ m = 9$ Method				$n = 180,\ m = 18$ Method			
	σ_i^2	I	II	III	IV	I	II	III	IV
Relative	1	0.0048	0.0198	0.0272	0.0018	-0.0051	-0.0086	-0.0112	-0.0111
bias	4	-0.0033	-0.0061	-0.0145	-0.0158	-0.0130	-0.0109	-0.0065	-0.0116
	16	0.0126	0.0370	0.0369	0.0096	-0.0046	-0.0045	-0.0080	-0.0061
MSE	1	0.3066	0.3890	0.6861	0.3805	0.2258	0.2680	0.4470	0.2922
	4	0.3281	0.5430	1.3778	0.7285	0.2595	0.3000	0.5805	0.3748
	16	0.3715	0.5240	1.6749	1.9316	0.2815	0.2850	0.4856	0.6383
ALCI	1	2.1393	2.5485	4.4906	3.0528	1.9220	1.6006	3.6466	2.4811
	4	2.2632	3.9574	6.8887	5.6842	2.0557	2.1524	5.2472	4.2160
	16	2.3221	4.5619	9.3335	11.1363	2.1046	2.3308	6.5273	7.8492
CP	1	0.9468	0.9770	0.9771	0.9708	0.9564	0.9710	0.9851	0.9631
	4	0.9468	0.9710	0.9829	0.9917	0.9555	0.9660	0.9967	0.9967
	16	0.9365	0.9660	0.9933	0.9975	0.9529	0.9610	0.9998	0.9999

Table 3
Simulation results for prediction when $\tau^2 = 1$. Here MSE, ALCI, CP represent the mean squared error, average confidence interval width and coverage probability, respectively

		$n = 36,\ m = 9$ Method				$n = 180,\ m = 18$ Method			
	σ_i^2	I	II	III	IV	I	II	III	IV
Relative	1	-0.0152	0.0205	0.0255	0.0051	-0.0064	-0.0085	-0.0111	-0.0101
bias	4	-0.0167	-0.0164	-0.0151	-0.0219	-0.0151	-0.0121	-0.0133	-0.0164
	16	-0.0323	0.0508	0.0515	0.0216	-0.0028	-0.0017	-0.0073	-0.0039
MSE	1	0.5645	0.6330	0.7238	0.6260	0.5288	0.5430	0.5673	0.6336
	4	0.8566	1.1100	1.5396	1.0992	0.8159	0.8770	0.9415	0.8948
	16	1.0482	1.3100	2.1059	2.3156	0.9786	1.0000	1.1024	1.1878
ALCI	1	3.4550	3.1822	4.4938	3.2117	3.1088	2.5094	3.6763	2.8676
	4	4.0321	5.8733	6.8984	5.7909	3.7844	4.2908	5.3323	4.5543
	16	4.4082	7.4286	9.3555	11.1555	4.1187	5.1590	6.6785	7.8937
CP	1	0.9704	0.9640	0.9762	0.9275	0.9660	0.9650	0.9786	0.8879
	4	0.9633	0.9560	0.9812	0.9808	0.9627	0.9680	0.9918	0.9740
	16	0.9533	0.9490	0.9912	0.9938	0.9613	0.9680	0.9974	0.9979

Table 4
Simulation results for prediction when $\tau^2 = 4$. Here MSE, ALCI, CP represent the mean squared error, average confidence interval length and the coverage probability, respectively

		$n = 36$, $m = 9$				$n = 180$, $m = 18$			
		Method				Method			
	σ_i^2	I	II	III	IV	I	II	III	IV
Relative bias	1	-0.0024	0.0248	0.0229	0.0180	-0.0084	-0.0098	-0.0122	-0.0106
	4	-0.0343	-0.0310	-0.0210	-0.0340	-0.0110	-0.0092	-0.0174	-0.0132
	16	-0.0147	0.0702	0.0767	0.0467	0.0016	0.0024	-0.0059	0.0012
MSE	1	0.8822	0.8590	0.8579	1.0559	0.8359	0.8180	0.8541	0.8605
	4	2.0577	2.2900	2.1818	2.2422	2.0424	2.1000	2.0935	2.1130
	16	3.4516	3.7600	3.9267	3.8981	3.3153	3.3500	3.3939	3.3631
ALCI	1	4.6318	4.1936	4.5369	3.7677	4.0256	3.5346	3.9626	3.7499
	4	6.2015	10.9093	7.0376	6.4314	5.9000	9.0913	6.2217	6.1540
	16	7.7221	18.0039	9.6718	11.3341	7.4430	14.6665	8.3908	8.7537
CP	1	0.9791	0.9670	0.9733	0.9029	0.9674	0.9570	0.9600	0.9468
	4	0.9556	0.9670	0.9725	0.9496	0.9592	0.9610	0.9633	0.9573
	16	0.9510	0.9670	0.9796	0.9858	0.9573	0.9650	0.9718	0.9776

Bias Comparisons: In most cases, the bias of the four methods are comparable. There is no clear evidence of significant differences between them in terms of the bias. High sampling variance gives more weight to the population mean by construction that makes the estimator closer to the mean at the second level. On the other hand, Methods I - III use shrinkage estimators of the sampling variances which would be less than the maximum of all sampling variances. Thus, Methods I - III tend to have little more bias. However, due to shrinkage in sampling variances, one may expect a gain in the variance of the estimators which, in turn, makes the MSE smaller. Among Methods I - III, Method I performed better compared to Methods II and III, which were quite similar to each other. The maximum gain using Method I compared to Method II is 99%.

MSE Comparisons: In terms of the MSE, Method I performed consistently better than the other three in all cases except when the ratio of σ_i^2 to τ^2 is the lowest: ($\sigma_i^2 =$ 1) / ($\tau^2 = 4$) = 0.25. In this case, the variance between small areas (model variance) is much higher than the variance within the areas (sampling variance). When using our method to estimate θ_i, the information "borrowed" from other areas may misdirect the estimation: The estimated mean of the Gamma distribution for σ_i^{-2} from the second level in (2) is $\hat{a}\hat{b}$ which equals 0.44 approximately for both the (m, n) combinations of (9, 36) and (18, 180) (the true value is $ab = 0.4$). Thus, $E(\sigma_i^{-2} | X_i, S_i^2, \hat{B})$ is significantly smaller than 1 due to shrinkage towards the mean for the group which has the true value of $\sigma_i^2 = 1$. Also, since σ_i^2 is smaller than τ^2, the weight of X_i should be much more compared to β, the overall mean. However, due to underestimation of σ_i^{-2} in this case, the resulting estimator puts less weight on X_i which leads to higher MSE. However, this underestimation will decrease for large sample sizes due to the consistency of Bayes estimators. This fact is actually observed when the sample size increases from $n = 36$ to $n = 180$ for the case $\sigma_i^2 = 1$ and $\tau^2 = 4$. Compared to Method II, Method I shows gains in most of the simulation cases; the maximum gain is 30% while the only loss is 9% for the combination $\sigma_i^2 = 1$ and $\tau^2 = 4$ for $n = 36$ and $m = 9$. Similarly, for Method III, the maximum gain of Method I is 77% and the only loss of 11% is for the same parameter and sample size specifications.

ACP Comparisons: We obtained confidence intervals with confidence level 95%. Methods I and III do not indicate any under-coverage. This is expected from their optimal confidence interval construction. Method I meets the nominal coverage rate more frequently than any other methods. Method II has some under coverage and can go as low as 82%.

ALCI Comparisons: Method I produced considerably shorter confidence intervals in general. Method IV produced comparable lengths as the other methods in all cases except when σ_i^2 was high, in which case, the lengths were considerably higher. The confidence interval proposed in Qiu and Hwang (2007) does not have good finite sample properties, particularly for small τ^2. To avoid low coverage, they proposed to truncate $M_0 = \tau^2/(\tau^2 + \sigma_i^2)$ with a positive number $M_1 = 1 - Q_\alpha/(\nu - 2)$ for known σ_i^2 where Q_α is the α^{th}-quantile of a chi-squared distribution with ν degrees of freedom. When the ratio of sampling variance to model variance, σ_i^2/τ^2, is high, M_1 tends to be higher than M_0. This results in a nominal coverage but with larger interval lengths. For example, in case of $(\sigma_i^2, \tau^2) = (16, 0.25)$, the ALCI is 11.13 for Method IV whereas ALCI is only 2.78 and 4.56 for Methods I and II.

5.2 Robustness study

In order to study the robustness of the proposed method with respect to departures from the normality assumption in the errors, we conducted the following simulation study. Data was generated as before but with e_{ij}'s drawn from a double-exponential (Laplace) and an uniform distribution. The estimators from Methods II and III had little effect. This is perhaps due to the fact that these methods used moment based estimation for model parameter estimation. Method IV resulted in larger relative bias, MSE and ALCI, and lower coverage probability. The MSE from Method I is always lower than that from Method II. For $\tau^2 = 0.25$ and 1, ALCI is smaller for Method I compared to Method II for $(n = 36,\ m = 9)$ but the results are opposite when $(n = 180,\ m = 18)$. In terms of CP, Method II has some under coverage (lowest is 80%). However, Method I did not have any under-coverage. In order to save space we only provide the results for parameters a, b, β and τ^2 under the Laplace errors (see Table 5).

6. Real data analysis

We illustrate our methodology based on a widely studied example. The data set is from the U.S. Department of Agriculture and was first analyzed by Battese (1988). The data set is on corn and soybeans productions in 12 Iowa counties. The sample sizes for these areas are small, ranging from 1 to 5. We shall consider corn only to save space. For the proposed model, the sample sizes $n_i > 1$ necessarily. Therefore, modified data from You and Chapman (2006) with $n_i \geq 2$ are used. The mean reported crop hectares for corn (X_i) are the direct survey estimates and are given in Table 6. Table 6 also gives the sample variances which are calculated based on the original data assuming simple random sampling. The sample standard deviation varies widely, ranging from 5.704 to 53.999 (the coefficient of variation varies from 0.036 to 0.423). Two covariates are considered in Table 6: Z_{i1}, the mean of pixels of corn, and Z_{i2}, the mean of pixels of soybean, from the LANDSAT satelite data.

Table 5
Simulation results for the model parameters, a (top left panel), b (top right panel), β (bottom left panel) and τ^2 (bottom right panel) when the errors follow a laplace distribution. Here SD represents the standard deviation over 200 replicates. We took $\beta = 10$ and $\tau^2 = 0.25$, 1 and 4

	$n = 36,\ m = 9$		$n = 180,\ m = 18$			$n = 36,\ m = 9$		$n = 180,\ m = 18$	
τ^2	Mean	SD	Mean	SD	τ^2	Mean	SD	Mean	SD
		a					b		
0.25	0.9624	0.1632	0.9471	0.0498	0.25	0.5793	0.1733	0.5279	0.0501
1	0.9628	0.1657	0.9476	0.0497	1	0.5816	0.1777	0.5275	0.0503
4	0.9689	0.1694	0.9487	0.0499	4	0.5758	0.1796	0.5263	0.0503
		β					τ^2		
0.25	9.9736	0.3775	9.9800	0.1773	0.25	0.2696	0.0882	0.2565	0.0074
1	9.9753	0.3709	9.9836	0.1662	1	1.0508	0.2501	1.0403	0.0668
4	9.9736	0.4835	9.9855	0.2161	4	3.9624	1.1719	4.1256	0.4201

Table 6
Corn data from You and Chapman (2006)

County	n_i	X_i	Z_{1i}	Z_{2i}	$\sqrt{S_i^2}$
Franklin	3	158.623	318.21	188.06	5.704
Pocahontas	3	102.523	257.17	247.13	43.406
Winnebago	3	112.773	291.77	185.37	30.547
Wright	3	144.297	301.26	221.36	53.999
Webster	4	117.595	262.17	247.09	21.298
Hancock	5	109.382	314.28	198.66	15.661
Kossuth	5	110.252	298.65	204.61	12.112
Hardin	5	120.054	325.99	177.05	36.807

The estimates of $\boldsymbol{B}$ are as follows: $a = 1.707$, $b =$ 0.00135, $\tau^2 = 90.58$ and $\beta = (-186.0, 0.7505, 0.4100)$. The estimated prior mean of $1/\sigma_i^2$ which is the mean of the Gamma distribution with parameters a and b is $ab =$ 0.002295 with a square root of 0.048 (note that $1/0.048 =$ 20.85 consistent with the range of the sample standard deviations between 5.704 and 53.999). The small area estimates and their confidence intervals are summarized in Table 7 and Figure 1. Point estimates of all 4 methods are comparable: the summary measures comprising of the mean, median, and range of the small area parameter estimates for Methods I, II, III, and IV are (121.9, 124.1, 122.2, 122.6), (125.2, 120.4, 115.0, 114.5) and (23.1, 53.0, 58.4, 56.6), respectively. The distribution of $\hat{\theta}_i$ (plotted based on considering all the i's) are summarized in Figure 2 which shows that there is a significant difference in their variability. Method I has the lowest variability and is superior in this sense. Further, smoothing sampling variances has strong implication in measuring uncertainty and hence in the interval estimation. The proposed method has the shortest confidence interval on an average compared to all other methods. Methods II and III provide intervals with negative lower limits. This seems unrealistic because the direct average of area under corn is positive and large for all the 12 counties (the crude confidence intervals ($x_i \pm t_{0.025} S_i$) do not contain zero for any of the areas either). Note that Method II does not have any theoretical support on its confidence intervals. Methods II and III produce wider confidence intervals when the sampling variance is high. For example, the sample size for both Franklin county and Pocahontas county is three, but sampling standard deviations are 5.704 and 43.406. Although the confidence interval under Method I is comparable, they are wide apart for Methods II and III. This is because although these methods consider the uncertainty in sampling variance estimates, the smoothing did not use the information from direct survey estimates, resulted the underlying sampling variance estimates remain highly variable (due to small sample size). In effect, the variance of the variance estimator (of the point estimates) is bigger compared to that in method I. This is further confirmed by the fact that the intuitive standard deviations of the "smoothed" small area estimates (one fourth of the interval) are smaller and less variable under method I compared to the others. Another noticeable aspect of our method is that the interval widths are similar for counties with same sample size. This could be an indication of obtaining equ-efficient estimators for equivalent sample sizes.

Model selection: For choosing the best fitting model, we used the Bayesian Information Criteria (BIC) which takes into account both the likelihood as well as the complexity of the fitted models. We calculated BICs for the models used in Methods I and III (Hwang *et al.* 2009). These two models have the same numbers of parameters with a difference in only the way the parameters are estimated. The model BIC for Method I is 210.025 and that for Method III is 227.372. This indicates superiority of our model. We could not compute the BIC for Wang and Fuller (2003) since they did not use any explicit likelihood.

Table 7
Results of the corn data analysis. Here CI and LCI represent the confidence interval and the length of the confidence interval, respectively

County	$\hat{\theta}_i$	CI	LCI	$\hat{\theta}_i$	CI	LCI
		I: Proposed method			II: Wang and Fuller (2003)	
Franklin	131.8106	104.085, 159.372	55.287	155.4338	124.151, 193.094	68.943
Pocahontas	108.7305	80.900, 136.436	55.536	102.3682	-38.973, 244.019	282.993
Winnebago	109.0559	81.430, 136.646	55.216	115.9093	-53.768, 279.314	333.083
Wright	131.6113	103.736, 159.564	55.828	131.0674	8.330, 280.263	271.932
Webster	113.1484	92.805, 133.348	40.543	109.4795	32.514, 202.675	170.161
Hancock	129.4279	111.781, 147.193	35.412	124.1028	56.750, 162.013	105.262
Kossuth	121.0071	103.451, 138.626	35.175	116.7147	68.049, 152.454	84.405
Hardin	130.2520	112.373, 148.114	35.741	137.7983	51.734, 188.373	136.638
		III: Hwang *et al.* (2009)			IV: Qiu and Hwang (2007)	
Franklin	158.4677	128.564, 188.370	59.805	157.7383	146.999, 168.477	21.478
Pocahontas	100.1276	-44.039, 244.295	288.334	101.1661	19.444, 182.887	163.442
Winnebago	114.1473	0.065, 228.228	228.163	113.7746	56.263, 171.286	115.022
Wright	140.3717	-24.119, 304.862	328.982	143.2244	41.559, 244.889	203.330
Webster	115.7865	50.297, 181.275	130.978	115.2224	75.124, 155.320	80.196
Hancock	111.3087	66.213, 156.403	90.189	113.1766	83.691, 142.661	58.970
Kossuth	110.9585	74.366, 147.550	73.184	112.3239	89.520, 135.127	45.607
Hardin	126.6093	40.040, 213.178	173.137	123.9049	54.607, 193.202	138.594

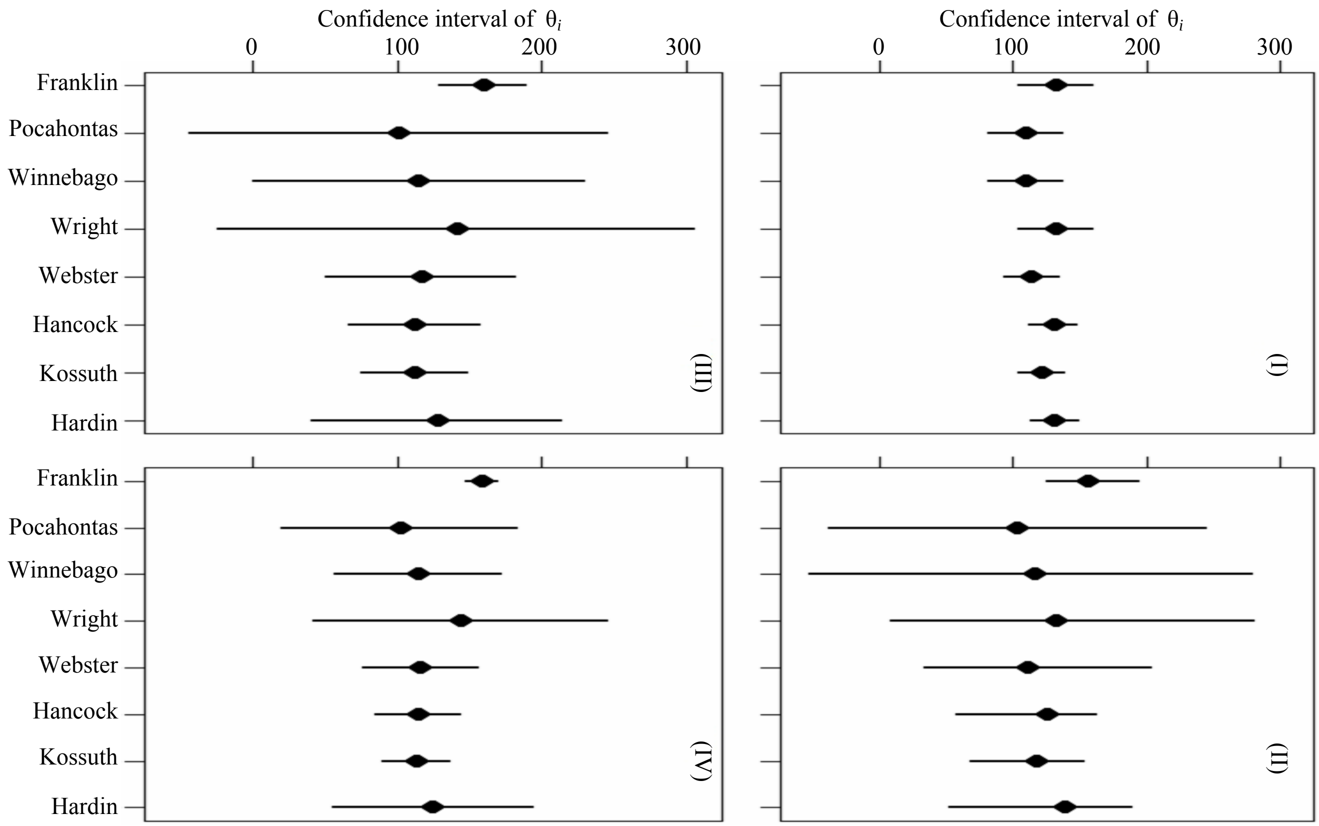

Figure 1 Corn hectares estimation. The horizontal line for each county displays the confidence interval of $\hat{\theta}_i$, with $\hat{\theta}_i$ marked by the circle, for (I) Proposed method, (II) Wang and Fuller (2003), (III) Hwang *et al.* (2009) and (IV) Qiu and Hwang (2007)

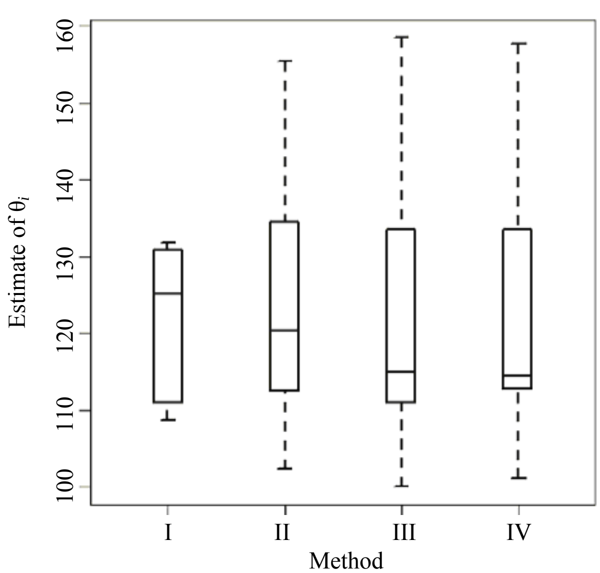

Figure 2 Boxplot of estimates of corn hectares for each county. (I) to (IV) are the 4 methods corresponding to Figure 1

7. Conclusion

In this paper, joint area level modeling of means and variances is developed for small area estimation. The resulting small area estimators are shown to be more efficient than the traditional estimators obtained using Fay-Herriot models which only shrink the means. Although our model is same as one considered in Hwang *et al.* (2009), our method of estimation is different in two ways: In the determination of the tuning parameter k and the use of $\pi(\sigma_i^2 \mid X_i, S_i^2, \boldsymbol{Z}_i)$ (which depends additionally on X_i), instead of $\pi(\sigma_i^2 \mid S_i^2, \boldsymbol{Z}_i)$, for constructing the conditional distribution of the small area parameters θ_i. We demonstrated robustness properties of the model when the assumption that σ_i^2 arise from a inverse Gamma distribution is violated. The borrowing of X_i information when estimating σ_i^2 as well as the robustness with respect to prior elicitation demonstrate the superiority of our proposed method. The parameter values chosen in the simulation study are different than in the real data analysis. The real data analysis given here is merely for illustration purposes. Our main aim was to develop the methodology of mean-variance modeling and contrast with some closely related methods to show its effectiveness. For this reason, we chose parameter settings in the simulation to be the same as in the well-known small area estimation article Wang and Fuller (2003).

Obtaining improved sampling variance estimators is a byproduct of the proposed approach. We have provided an innovative estimation technique which is theoretically justified and user friendly. Computationally, the method is much simpler compared to other competitive methods such as Bayesian MCMC procedures or bootstrap resampling methods. We need sampling from posterior distribution only once during the model parameter estimation, and the sampled values can be used subsequently for all other purposes. The software is available from the authors upon request.

Acknowledgements

The authors like to thank two referees and the Associate Editor for their constructive comments that have led to a significantly improved version of this article. The research is partially supported by NSF grants SES 0961649, 0961618 and DMS 1106450.

Appendix

A. Derivation of the conditional distributions

From Equation (1) and (2), the conditional joint distribution of $\{X_i, S_i^2, \theta_i, \sigma_i^2\}$, $\pi(X_i, S_i^2, \theta_i, \sigma_i^2 \mid a, b, \beta, \tau^2)$, is

$$\begin{aligned}
&\pi(X_i, S_i^2, \theta_i, \sigma_i^2 \mid \boldsymbol{Z}_i, \boldsymbol{B}) \\
&= \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(X_i-\theta_i)^2}{2\sigma_i^2}\right\} \frac{1}{\Gamma\left(\frac{n_i-1}{2}\right) 2^{\frac{n_i-1}{2}}} \\
&\times \left\{(n_i-1)\frac{S_i^2}{\sigma_i^2}\right\}^{\frac{n_i-1}{2}-1} \exp\left\{-\frac{(n_i-1)S_i^2}{2\sigma_i^2}\right\} \\
&\times \left(\frac{n_i-1}{\sigma_i^2}\right) \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2}\right\} \\
&\times \frac{1}{\Gamma(a) b^a} \left(\frac{1}{\sigma_i^2}\right)^{a+1} \exp\left(-\frac{1}{b\sigma_i^2}\right) \\
&\propto \exp\left[-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2} - \left\{\frac{(X_i-\theta_i)^2}{2} + \frac{(n_i-1)S_i^2}{2} + \frac{1}{b}\right\}\frac{1}{\sigma_i^2}\right] \\
&\times \left(\frac{1}{\sigma_i^2}\right)^{\frac{n_i}{2}+a+1} \left(\frac{1}{\tau^2}\right)^{\frac{1}{2}} \frac{1}{\Gamma(a) b^a}.
\end{aligned}$$

Therefore the conditional distribution of σ_i^2 and θ_i given the data and $\boldsymbol{B}$ are

$$\begin{aligned}
&\pi(\sigma_i^2 \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}) \\
&= \int \pi(X_i, S_i^2, \theta_i, \sigma_i^2 \mid \boldsymbol{Z}_i, \boldsymbol{B})\, d\theta_i \propto \frac{1}{(\sigma_i^2)^{(n_i-1)/2+a+1} (\sigma_i^2+\tau^2)^{1/2}} \\
&\exp\left[-\frac{(X_i - \boldsymbol{Z}_i^T, \beta)^2}{2(\sigma_i^2+\tau^2)} - \left\{\frac{1}{2}(n_i-1) S_i^2 + \frac{1}{b}\right\}\left(\frac{1}{\sigma_i^2}\right)\right],
\end{aligned}$$

$$\begin{aligned}
\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{Z}_i, \boldsymbol{B}) &= \int \pi(X_i, S_i^2, \theta_i, \sigma_i^2 \mid \boldsymbol{Z}_i, \boldsymbol{B})\, d\sigma_i^2 \\
&\propto \exp\left\{-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2}\right\} \psi_i^{-\left(\frac{n_i}{2}+a\right)}
\end{aligned}$$

where ψ_i is defined in Equation (4).

B. Details of the EM algorithm

The maximization of $Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})$ is done by setting the partial derivatives with respect to $\boldsymbol{B}$ to be zero, that is,

$$\frac{\partial Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})}{\partial \boldsymbol{B}} = 0. \qquad \text{(B.1)}$$

From the expression of $Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})$ in the text, we give explicit expressions for the partial derivates with respect to each component of $\boldsymbol{B}$. The partial derivative corresponding to β is

$$\begin{aligned}
&\frac{\partial Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})}{\partial \beta} \\
&= \sum_{i=1}^n \frac{\int \boldsymbol{Z}_i \left(\frac{\theta_i - Z_i^T\beta}{\tau^2}\right) \exp\left\{-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2}\right\} \psi_i^{-\left(\frac{n_i}{2}+a\right)} d\theta_i}{\int \exp\left\{-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2}\right\} \psi_i^{-\left(\frac{n_i}{2}+a\right)} d\theta_i} \\
&= \sum_{i=1}^n E\left\{\boldsymbol{Z}_i \left(\frac{\theta_i - \boldsymbol{Z}_i^T\beta}{\tau^2}\right)\right\}
\end{aligned}$$

where the expectation is with respect to the conditional distribution of θ_i, $\pi(\theta_i \mid X_i, S_i^2, \boldsymbol{B})$. The expression of the partial derivative corresponding to τ^2 is:

$$\begin{aligned}
&\frac{\partial Q(\boldsymbol{B} \mid \boldsymbol{B}^{(t-1)})}{\partial \tau^2} \\
&= -\frac{n}{2\tau^2} + \sum_{i=1}^n \frac{\int \frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2(\tau^2)^2} \exp\left\{-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2}\right\} \psi_i^{-\left(\frac{n_i}{2}+a\right)} d\theta_i}{\int \exp\left\{-\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2\tau^2}\right\} \psi_i^{-\left(\frac{n_i}{2}+a\right)} d\theta_i} \\
&= -\frac{n}{2\tau^2} + \sum_{i=1}^n E\left\{\frac{(\theta_i - \boldsymbol{Z}_i^T\beta)^2}{2(\tau^2)^2}\right\}.
\end{aligned}$$

Similarly for a and b, we get the solutions by setting $S_a = 0$ and $S_b = 0$ where S_a and S_b are, respectively, the partial derivatives of $Q(\boldsymbol{B} | \boldsymbol{B}^{(t-1)})$ with respect to a and b with expressions given in the main text. These equations are solved using the Newton-Raphson method which requires the matrix of second derivatives with respect to a and b. These are given by the following expressions:

$$S_{aa} = \sum_{i=1}^{n}\left[\log''\left\{\Gamma\left(\frac{n_i}{2}+a\right)\right\} - \log''\{\Gamma(a)\} + \mathrm{Var}\{\log(\psi_i)\}\right]$$

$$S_{ab} = \sum_{i=1}^{n}\left[-\frac{1}{b}+\frac{1}{b^2}E\left(\frac{1}{\psi_i}\right)-\left(\frac{n_i}{2}+a\right)\frac{1}{b^2}\mathrm{Cov}\left\{\frac{1}{\psi_i}, \log(\psi_i)\right\}\right],$$

and (B.2)

$$S_{bb} = \sum_{i=1}^{n}\left\{\frac{a}{b^2} - (n_i+2a)\frac{1}{b^3}E\left(\frac{1}{\psi_i}\right) + \left(\frac{n_i}{2}+a\right)\frac{1}{b^4} E\left(\frac{1}{\psi_i^2}\right) + \left(\frac{n_i}{2}+a\right)^2\frac{1}{b^4}\mathrm{Var}\left(\frac{1}{\psi_i}\right)\right\}$$

with $S_{ba} = S_{ab}$. At the u^{th} step, the update of a and b are given by

$$\begin{bmatrix} a^{(u)} \\ b^{(u)} \end{bmatrix} = \begin{bmatrix} a^{(u-1)} \\ b^{(u-1)} \end{bmatrix} - \begin{bmatrix} S_{aa}^{(u-1)} & S_{ab}^{(u-1)} \\ S_{ba}^{(u-1)} & S_{bb}^{(u-1)} \end{bmatrix}^{-1} \begin{bmatrix} S_a^{(u-1)} \\ S_b^{(u-1)} \end{bmatrix}, \quad (B.3)$$

where the superscript $(u-1)$ on S_{aa}, S_{ab}, S_{ba}, S_{bb}, S_a and S_b denote these quantities evaluated at the values of a and b at the $(u-1)^{\text{th}}$ iteration. Once the Newton Raphson procedure converges, the value of a and b at the t^{th} step of the EM algorithm is set as $a^{(t)} = a^{(\infty)}$ and $b^{(t)} = b^{(\infty)}$.

C. An alternative small area model formulation

It is possible to reduce the width of the confidence interval $\tilde{C}(\boldsymbol{B})$ based on an alternative hierarchical model for small area estimation which has some mathematical elegance. The constant term $n_i + 2a + 2$ in (19) becomes $n_i + 2a$ in this alternative model formulation. The model is given by

$$X_i | \theta_i, \sigma_i^2 \sim N(\theta_i, \sigma_i^2), \quad (C.1)$$

$$\theta_i | \sigma_i^2 \sim N(\boldsymbol{Z}_i\beta, \lambda\sigma_i^2), \quad (C.2)$$

$$\left.\frac{(n_i-1)S_i^2}{\sigma_i^2}\right| \sigma_i^2 \sim \chi^2_{n_i-1}, \quad (C.3)$$

$$\sigma_i^2 \sim \mathrm{Inverse-Gamma}(a,b), \quad (C.4)$$

independently for $i = 1, 2, \ldots, n$. Note that in the above formulation, it is assumed that the conditional variance of θ_i is proportional to σ_i^2 whereas the marginal variance is constant (by integrating out σ_i^2 using (C.4). In (1) and (2), the variance of θ_i is a constant, τ^2, independent of σ_i^2, and there is no conditional structure for θ_i depending on σ_i^2. The set of all unknown parameters in the current hierarchical model is $\boldsymbol{B} = (a, b, \beta, \lambda)$. The inference procedure for this model is given subsequently. The model essentially assumes that the true small area effects are not identically distributed even after eliminating the known variations.

C.1 Inference methodology

By re-parameterizing the variance as in (C.2), some analytical simplifications are obtained in the derivation of the posteriors of θ_i and σ_i given X_i, S_i^2 and $\boldsymbol{B}$. We have

$$\pi(\sigma_i^2 | X_i, S_i^2, \boldsymbol{B}) = IG\left(\frac{n_i}{2}+a, \left[\frac{(n_i-1)S_i^2}{2} + \frac{(X_i - \boldsymbol{Z}_i\beta)^2}{2(1+\lambda)} + \frac{1}{b}\right]^{-1}\right)$$

where $IG(a,b)$ stands for the inverse Gamma distribution with shape and scale parameters a and b, respectively. Given $\boldsymbol{B}$ and σ_i^2, the conditional distribution of θ_i is

$$\pi(\theta_i | X_i, \sigma_i^2, \boldsymbol{B}) = \mathrm{Normal}\left(\boldsymbol{Z}_i^T\beta, \frac{\lambda\sigma_i^2}{1+\lambda}\right).$$

Integrating out σ_i^2, one obtains the conditional distribution of θ_i given X_i, S_i^2 and $\boldsymbol{B}$,

$$\pi(\theta_i | X_i, S_i^2, \boldsymbol{B}) = \int_0^\infty \pi(\theta_i | X_i, \sigma_i^2, B)\, \pi(\sigma_i^2 | X_i, S_i^2, \boldsymbol{B})\, d\sigma_i^2 \propto \left\{\frac{(1+\lambda)}{2\lambda}(\theta_i - \boldsymbol{Z}_i^T\beta)^2 + \frac{\delta^2}{2}\right\}^{-(n_i+2a+1)/2}, \quad (C.5)$$

where $\delta^2 = (n_i - 1)S_i^2 + (X_i - \boldsymbol{Z}_i\beta)^2/(1+\lambda) + 2/b$. We can rewrite (C.5) as

$$\pi(\theta_i | X_i, S_i^2, \boldsymbol{B}) = \frac{\Gamma((n_i+1)/2+a)\sqrt{1+\lambda}}{\delta^*\Gamma(n_i/2+a)\sqrt{(n_i+2a)\lambda\pi}} \left\{1 + \frac{(\theta_i - \mu_i)^2}{(n_i+2a)\delta^{*2}\lambda/(1+\lambda)}\right\}^{-(n_i+2a+1)/2}$$

which can be seen to be a scaled t-distribution with $n_i + 2a$ degrees of freedom and scale parameter $\delta^*\sqrt{\lambda/(1+\lambda)}$ with $\delta^{*2} = \delta^2/(n_i + 2a)$. Hence,

$$E(\sigma_i^{-1} | X_i, S_i, \boldsymbol{B}) = \frac{\Gamma((n_i+1)/2+a)(\delta^2/2)^{-\{(n_i+1)/2+a\}}}{\Gamma(n_i/2+a)(\delta^2/2)^{-(n_i/2+a)}} = \frac{\Gamma((n_i+1)/2+a)}{\Gamma(n_i/2+a)}\,\frac{\sqrt{2}}{\delta^*\sqrt{n_i+2a}}.$$

In this context, choosing

$$k = k(\boldsymbol{B}) = \left\{1 + \frac{t_{\alpha/2}^2}{n_i - 1}\right\}^{-(n_i+2a+1)/2} \sqrt{\frac{1+\lambda}{\lambda}} \frac{1}{\sqrt{2\pi}},$$

the confidence interval in (8) simplifies to

$$C_i(\boldsymbol{B}) \equiv \left\{\theta_i : \frac{|\theta_i - \mu_i|}{\sqrt{\frac{\lambda}{1+\lambda} \frac{(n_i + 2a)\delta^{*2}}{n_i - 1}}} \le t_{\alpha/2}\right\}. \quad \text{(C.6)}$$

Using the similar arguments as before and noting that $(n_i + 2a)\delta^{*2} \ge (n_i - 1)S_i^2$, we have $P\{C_i(\boldsymbol{B})\} \ge P(D_i) = 1 - \alpha$ where D_i is the confidence interval in (20). When $\boldsymbol{B}$ is unknown, we replace $\boldsymbol{B}$ by its marginal maximum likelihood estimate $\hat{\boldsymbol{B}}$. It is expected that the pooling technique will result in an error small enough so that $P\{C_i(\hat{\boldsymbol{B}})\} \approx P\{C_i(\boldsymbol{B})\} \ge 1 - \alpha$.

References

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satelite data. *Journal of the American Statistical Association*, 95, 28-36.

Bell, W. (2008). Examining sensitivity of small area inferences to uncertainty about sampling error variances. Technical Report U.S. Census Bureau.

Casella, G., and Hwang, J. (1991). Evaluating confidence sets using loss functions. *Statistica Sinica*, 1, 159-173.

Chatterjee, S., Lahiri, P. and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *Annals of Statistics*, 36, 1221-1245.

Cho, M., Eltinge, J., Gershunskaya, J. and Huff, L. (2002). Evaluation of generalized variance function estimators for the U.S. current employment survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 534-539.

Fay, R., and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Gershunskaya, J., and Lahiri, P. (2005). Variance estimation for domains in the U.S. current employment statistics program. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 3044-3051.

Ghosh, M., and Rao, J. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 54-76.

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

Huff, L., Eltinge, J. and Gershunskaya, J. (2002). Exploratory analysis of generalized variance function models for the U.S. current employment survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1519-1524.

Hwang, J., Qiu, J. and Zhao, Z. (2009). Empirical Bayes confidence intervals shrinking both mean and variances. *Journal of the Royal Statistical Society*, B, 71, 265-285.

Joshi, V. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *The Annals of Mathematical Statistics*, 40, 1042-1067.

Maples, J., Bell, W. and Huang, E. (2009). Small area variance modeling with application to county poverty estimates from the american community survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 5056-5067.

Otto, M., and Bell, W. (1995). Sampling error modelling of poverty and income statistics for states. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 160-165.

Pfeffermann, D. (2002). Small area estimation - New developments and directions. *International Statistical Review*, 70, 125-143.

Prasad, N., and Rao, J. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

Qiu, J., and Hwang, J. (2007). Sharp simultaneous intervals for the means of selected populations with application to microarray data analysis. *Biometrics*, 63, 767-776.

Rao, J. (2003). Some new developments in small area estimation. *Journal of the Iranian Statistical Society*, 2, 145-169.

Rivest, L.-P., and Vandal, N. (2003). Mean squared error estimation for small areas when the small area variances are estimated. *Proceedings of the International Conference on Recent Advances in Survey Sampling*.

Robert, C., and Casella, G. (2004). *Monte Carlo Statistical Methods* (Second edition).

Valliant, R. (1987). Generalized variance functions in stratified two-stage sampling. *Journal of the American Statistical Association*, 82, 499-508.

Wang, J., and Fuller, W. (2003). The mean squared error of small area predictors constructed with estimated error variances. *Journal of the American Statistical Association*, 98, 716-723.

You, Y., and Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology*, 32, 1, 97-103.