

Sampling Variance Estimation Methods in Area Level of Small Area Estimation: A Narrative Review

Mojtaba **Soltani-Kermanshahi***

Social Determinants of Health Research Center, Semnan University of Medical Sciences, Semnan, Iran.

ABSTRACT

Introduction: In area level small area estimation (SAE) (Fay–Herriot and extension of it in spatial structure) sampling variances are assumed to be known. That is a powerful assumption and may be quite restrictive in some applications. The main in this field is the estimation of sampling variances for the small area parameters, which is necessary for obtaining reliable estimates and for evaluating the precision of the estimates.

Methods: The main objective of this article is to review some of the commonly used sampling variance estimation methods in area level of SAE. Information used to write this paper was collected from the sources of Google scholar, Scopus, MEDLINE and Web of Science and Hand searches of the references of retrieved literature.

Results: In the context of small area estimation, sampling variance estimation methods at the area level can be broadly classified into six categories: Direct Sampling Variance Estimator, Design-Based Variance Estimator, Generalized Variance Function, Model-based, Empirical Bayes and Extended methods.

Conclusion: According to the reviewed studies, no method can be reported as the best method for estimating the sampling variance in all conditions. Because each method needs different information, and on the other hand, simulation studies are needed to compare the methods.

Key words: Sampling variance; Estimation methods; Area level; Small area estimation; Fay herriot

***Corresponding Author:**
msoltani@semums.ac.ir



INTRODUCTION

Small area estimation (SAE) methods have been popularized due to the increasing demand for reliable estimates for small areas. Small area estimation (SAE) is a statistical technique used to estimate parameters in areas where direct estimation is difficult due to insufficient sample size. In SAE, the aim is to obtain accurate estimates for small areas by combining information from the sample and auxiliary sources.^{1,2}

However, lots of the popularity and utility of SAE methods can be ascribed to the model-based techniques which are now in wide utilization and have been extended in the past three decades. When area-level summary of ancillary variable exists, we can beneficially use area-level models. A fundamental area-level model was first suggested by Fay III and Herriot in 1979.³⁻⁶

Suppose that we have M areas in the population and only m areas are sampled. The SAE by the Fay–Herriot model has the following linear form:

$$\hat{\theta}_i = z_i^T \beta + b_i v_i + e_i \quad i = 1, \dots, m \quad (1)$$

In this model $\hat{\theta}_i = g(\hat{Y}_i)$ where $\hat{Y}_i$ is the direct estimator of the small area mean in area i , b_i 's are known positive constants (typically assumed to be $b_i = 1$), $\beta = (\beta_1, \dots, \beta_p)^T$ is the $p \times 1$ vector of regression coefficients. Moreover, v_i 's are area-specific random effects assumed to be independent and identically distributed (iid) as $N(0, \sigma_v^2)$ with $E_m(v_i) = 0$, $V_m(v_i) = \sigma_v^2 (\geq 0)$, Where E_m and V_m denote the model expectation and variance, respectively. e_i 's are independent sampling error with distribution of $N(0, \Psi_i)$, where sampling variances Ψ_i 's are assumed to be known. Also, v_i 's and e_i 's are assumed to be independent.¹⁻³

In numerous practical applications, data from neighbouring small areas present spatial correlation.^{2,6,7} More recently, an extension of the Fay–Herriot model through the Simultaneously Auto-Regressive (SAR) process has been considered. The model has the form (1) except that v be the result of a SAR process with unknown autoregression parameter ρ and proximity matrix W .^{8,9} Then v can be expressed as $v = (I_m - \rho W)^{-1} u$. Here, $u = (u_1, \dots, u_m)^T$ has mean 0 and covariance matrix $\sigma_u^2 I_m$, where I_m denotes the $m \times m$ identity matrix and σ_u^2 is an unknown parameter.^{7,8}

The conditional auto-regressive (CAR) structure is also a popular choice and in some papers the SAR structure is more common than the CAR.⁶

Also, Mehrabi et. al. (2020) proposed the exponential model in the Mat'ern function.

The model has the form (1) except that v is the vector by using an exponential model

$$\rho(d) = \exp(-d) \quad (2)$$

(where d and $\rho(\cdot)$ are the distance between two areas and correlation function, respectively.) in the Mat`ern function as follows:

$$v = (v_1, \dots, v_m) \sim \text{MvN}(0, S), \quad e_i \sim (0, \Psi_i), \quad s_{ij} = \sigma_v^2 \rho(d_{ij}), \quad i, j = 1, 2, \dots, m \quad (3)$$

where S is a variance-covariance matrix with elements s_{ij} and correlation function $\rho(\cdot)$; an isotropic correlation function that decays as the Euclidean distance $d_{ij} = \|l_i - l_j\|$ between two individuals increases. In the exponential model $\rho(\cdot)$ has the form describe by (2). Also, v_i 's and e_i 's are assumed to be independent.^{6, 10}

In area level SAE (Fay–Herriot and extension of it in spatial structure) sampling variances Ψ_i 's are assumed to be known. That is a powerful assumption and may be quite restrictive in some applications.^{1, 2, 5, 6} The main in this field is the estimation of sampling variances for the small area parameters, which is necessary for obtaining reliable estimates and for evaluating the precision of the estimates. Over the years, various sampling variance estimation methods for SAE have been proposed. The main objective of this article is to review some of the commonly used sampling variance estimation methods in area level of SAE.

METHODS

This narrative review study was conducted with the aim of investigating some common methods of sampling variance estimation at the level of the SAE area.

This research is the result of a research project approved by the research center for Social Determinants of Health and has an ethical code IR.SEMUMS.REC.1401.318.

Information used to write this paper was collected from the sources listed in table 1. Studies that met all the inclusion criteria (1. In particular, it has dealt with the estimation of sampling variance in small area estimation, 2. Access to the full text is possible) and 3. in any languages until March 2025 and did not meet any of the exclusion criteria (Lack of access to the full text) were included in the review.

Table 1. Sources Used for This Overview

Google scholar, Scopus, MEDLINE and Web of Science SEARCH til 2023 Key words: Sampling Variance, Area Level, Fay-Herriot, Spatial Structure, Estimation, Small Area, Mean Square Error.

Hand searches of the references of retrieved literature

Application of Methodology

In the context of small area estimation, sampling variance estimation methods at the area level can be broadly classified into six categories:

Direct Sampling Variance Estimator

The first method is the direct sampling variance estimator, discussed by Sørndal et al. (1992), which involves the computation of the sample variance of the data used to estimate the small area parameter. This method can be highly variable and requires assumptions to be made about the distribution of the data.^{2, 4-6, 11, 12}

Design-Based Variance Estimator

The Second method is the design-based variance estimator, which assumes that the sampling design is known and specifies a variance formula based on the design parameters. This method (e.g., jackknife, balanced repeated replication, etc by vide Wolter, 1985) is robust to assumptions about the distribution of the data, although it may not account for complex sampling designs.^{2, 3, 5, 6, 13-15}

Hernández et. al. in 2008 described a two-stage bootstrap method as well as a two-stage jackknife method to estimate variances of small area estimators of totals and rates in the SPANISH LABOUR FORCE SURVEY.¹⁶ Pratesi and Salvati (2008) in the SAR model estimated the sampling variances through a Jackknife procedure.¹⁷ On the other hand, You and Hidiroglou (2012) particularly studied the design effects methods for sampling variance smoothing for proportions.¹⁸

Generalized Variance Function (GVF)

The third method is the Generalized Variance Function (GVF) method which is based on the concept of the generalized variance function. This method can be used for nonlinear models and does not require prior information. However, it is computationally intensive and might be challenging to implement for large datasets.

The GVFs take on one of the forms in Table 2. ("We show the GVFs in terms of the variance, whereas Wolter (2007) gives them in terms of the relative variance".¹⁹) In that Y is the expectation of our estimator (such as mean , total etc...) and to estimate parameters such as α and β , we use a natural fitting methodology, namely ordinary least squares.¹⁹

Table 2. Generalized Variance Function MODELS

GVF Models	
I	γY^2
Ii	$\alpha + \beta Y$
Iii	$\alpha + \beta Y + \gamma Y^2$
Iv	$(\alpha + \beta/Y)^{-1}$
V	$(\alpha + \beta/Y + \gamma/Y^2)^{-1}$
Vi	αY^β

Note that for the poverty rates, which are proportions, the usual binomial distribution theory suggests

a model of the form

$$\alpha + \beta Y + \gamma Y^2$$

With $\alpha = 0$ and $\gamma = -\beta$ for GVF and Y is the expectation of our estimator (poverty rates).¹⁹⁻²²

Valliant (1987) connected use of a generalized variance function (GVF) with sample design by showing that a commonly used GVF is consistent with a particular class of prediction models for estimating totals from stratified, two-stage cluster samples.²³ Additional papers on GVF modeling of direct variance estimates include Huff, Eltinge, and Gershunskaya (2002), Cho et al. (2002), and Eltinge, Cho, and Hinrichs (2002) Shiferaw Y, Galpin J (2016).²⁴⁻²⁷

Mohadjer et al. (2012) similarly use a GVF to produce variance estimates for use in an area level model (for estimating proportions), assuming the following model

$$\log\left(\frac{V_a}{p_a^2}\right) = \eta_0 + \eta_1 \log(\tilde{p}_a) + \eta_2 \log(1 - \tilde{p}_a) + \eta_3 \log(n_a) + \varepsilon_a$$

Where $\varepsilon_a \sim N(0, \sigma_\varepsilon^2)$ and $\tilde{p}_a$ denotes a predictor of p_a based on a model dependent solely on auxiliary covariate information and not explicitly on any direct estimates.²⁸

For sampling variance smoothing, you (2021) used a log-linear regression model on the direct sampling variance Ψ_i as suggested in You and Hidiroglou (2012), and the smoothing model is defined as:

$$\log(\Psi_i) = \eta_0 + \eta_1 \log(n_i) + \varepsilon_i, \quad i = 1, \dots, m$$

where the model error term is $\varepsilon_i \sim N(0, \Sigma^{\square})$, and $\Sigma^{\square}$ is unknown. Let $\hat{\eta}_0$ and $\hat{\eta}_1$ denote the ordinary least square estimates of the regression coefficients η_0 and η_1 , and $\hat{\Sigma}$ be the of the log-linear regression model.^{18, 29} Therefore, a smoothed estimator of the sampling variance Ψ_i can be obtained as

$$\hat{\Psi}_{i^{\square}} = \exp(\hat{\eta}_0 + \hat{\eta}_1 \log(n_i)) \exp\left(\frac{\hat{\Sigma}^{\square}}{2}\right)$$

Maples (2016) similarly proposed a GVF for producing variance estimates based on estimating the design effect using additional information about any unequal weighting or clustering in the sampling procedure.^{27, 30} Moreover, the findings imply that estimating sampling error variances using GVF models can be a very useful strategy in the application of Empirical Best Linear Unbiased Prediction (EBLUP) small area estimation, most particularly in poverty incidence estimation.³¹

Model-based

Model-based method is the fourth method. In this method, the sampling variance of the small area parameters is estimated using a model-based approach that relies on a statistical model to predict the small area parameters. The model can include both area-specific and household-level covariates that are related to the small area parameter of interest. The variance of the small area parameter is then estimated using the standard errors of the estimated model coefficients, as well as the residual

variance. This method is flexible and can be applied to a wide range of small area parameters, but it requires a considerable amount of expertise to specify and estimate the appropriate statistical model and may not be reliable.³²

Following Arora and Lahiri (1997) and Battese et. al. (1988) and Wang and Fuller (2003) and You and Chapman (2006), the estimators $\hat{S}_i^2$ assumed to have distributed via

$$\frac{d_i \hat{S}_i^2}{\Psi_i} \sim \chi_{d_i}^2 \quad (4)$$

where $d_i = n_i - 1$ and n_i is the sample size for the i 'th area.^{14, 32-35}

You and Chapman (2006) analyzed an essentially similar area level model to that used by Arora and Lahiri (1997), applying it also to their data, as well as to small area data on amount of land planted with corn and soybeans taken from the paper of Battese, Harter, and Fuller (1988). You and Chapman (2006) provided results both for the above model and for a model that assumed the $\hat{S}_i^2$ were equal to the Ψ_i . Comparing posterior standard deviations or coefficients of variation for the two models showed, for the corn and soybean example, substantially larger uncertainty surrounding the small area predictions from the model with the Ψ_i assumed unknown, but, for the milk consumption example, almost identical results for the two models. The latter result was due to the large small area sample sizes for that example, while for the corn and soybean example the small area samples were quite small.^{14, 32, 33}

Kleffe and Rao (1992), used an area level model with direct sampling error variance estimates assumed unbiased, independent of the direct survey point estimates, and distributed proportional to a chi-squared random variable with known degrees of freedom.³⁶

We can also model the sampling variance indirectly in the sense that the sampling variance can be expressed as a function of small area parameters or design effects. For example, You and Rao (2002) proposed a log-linear unmatched model with the sampling variance written as Ψ_i as $\Psi_i = \theta_i^2 \cdot (cv)_i^2$, where θ_i is the unknown small area mean of interest and cvi is the smoothed coefficient of variation (CV). In applications, the smoothed CV can be obtained over time series data for a given local area, as shown in the small area unemployment rate estimation of You, Rao and Gambino (2003). If the parameter of interest is proportion or rate, we can model the sampling variance using smoothed design effects as $\Psi_i = \theta_i (1 - \theta_i) \cdot deff_i / n_i$, where $deff$ is a known design effect for a given area and n_i is the small area sample size (e.g., You, 2008; Liu, Lahiri and Kalton, 2014)(37-41). Hawala and Lahiri (2018) also, proposed a similar GVF for count data.^{5, 6, 42}

Empirical Bayes

Empirical Bayes method is the fifth method. In this method, the sampling variance of the small area parameters is estimated using an empirical Bayes approach that borrows strength from neighbouring areas. Under this approach, the small area parameters are modeled as a combination of a random area effect and a fixed effect that is common to all areas. It is usually assumed that the area-specific random effect follows a normal distribution with a mean of zero and an unknown variance. The variance is

then estimated using the empirical Bayes approach that estimates the variance based on the variability of the random effects across neighbouring areas.²

Otto and Bell (1995) developed a model with state random effects for sampling covariance matrices of Current Population Survey's estimated poverty ratios. This model provides for a Bayesian or empirical Bayes smoothing of the direct sampling variance estimates.²²

Arora and Lahiri (1997) viewed a Bayesian extension of the small area model due to Kleffe and Rao (1992).³⁶ They then developed a Gibbs sampling scheme to make Bayesian inferences for the model. The model is related to that of Otto and Bell (1995), though with a much simpler model for the sampling variances.^{14, 22}

Bell (1995) investigated time series models with sampling error components, with the sampling error variances treated as unknown. A Bayesian approach was implemented via accept/reject sampling and used to produce posterior means and variances of the time series components, a prediction problem analogous to small area estimation. In the application considered the sampling errors were approximately uncorrelated over time, and were assumed to have constant relative variance over time, with logarithms taken of the time series so the relative variance was taken as an approximation to the sampling variance in the log scale. Thus, the sampling variances depended on this one unknown parameter, of which multiple estimates were available.⁴³

Nguyen, Bell, and Gomish (2002) applied a similar modeling approach to an application for which the sampling errors did not appear uncorrelated over time (a second order autoregressive model was used), and for which the sampling variances varied over time due to sample size fluctuations.⁴⁴

Fay and Herriot (1979) and Rao (2003) construct HB models spatial structure under the assumption that the sampling variance Ψ_i are assumed known and replaced by the smoothed estimate $\hat{\Psi}_i$.^{1, 3} Gershunskaya and Lahiri (2005) developed simple models for design-based variance estimates or their logarithms that permitted empirical Bayes smoothing of the variance estimates. They found the resulting variance estimates had low bias and were much more stable than the design-based variances.⁴⁵ You and Chapman (2006) also have shown that the posterior estimates and the corresponding CVs are stable and not sensitive to the choice of parameters in the vague proper priors.³²

The model of You and Chapman (2006) in which an inverse gamma model is used for the sampling variance Ψ_i with known vague parameter values.

$$y_i | \theta_i, \Psi_i \sim \text{ind } N(\theta_i, \Psi_i), \quad i = 1, \dots, m; \quad (5)$$

$$d_i s_i^2 | \Psi_i \sim \text{ind } \Psi_i \chi_{d_i}^2, \quad d_i = n_i - 1, \quad i = 1, \dots, m; \quad (6)$$

$$\theta_i | \beta, \sigma_v^2 \sim \text{ind } N(x_i' \beta, \sigma_v^2), \quad i = 1, \dots, m; \quad (7)$$

$$\pi(\Psi_i) \sim \text{IG}(a_i, b_i), \text{ where } a_i = 0.0001, b_i = 0.0001, i \quad (8)$$

Flat priors for unknown parameters: $\pi(\beta) \propto 1$, $\pi(\sigma_v^2) \propto 1$. The full conditional distributions for the Gibbs sampling procedure under both FH-HB and YCM can be found in You and Chapman (2006).³²

Hwang et al. (2009) and Dass et al. (2012) noted that assuming a common prior for Ψ_i for all areas i could induce shrinkage in the resulting variance estimates and produce improved interval estimates for the parameters of interest.^{46, 47} The shrinkage estimator of variances in Hwang et al. (2009) is a function of S_i^2 only, and not of both X_i and S_i^2 . Thus, the shortcomings of their method are as follows:

(i) sampling variance estimation is independent of means; (ii) inference on means does not take into account the full uncertainty in variance estimation; and (iii) their model does not include any covariate information.⁴⁶

Gómez-Rubio and You Y, Zhou (2011) showed that gamma as a prior distribution for each one of the variances (in CAR structure) was related to give vague prior information and let the model learn from the data, they used small values for the parameters of the gamma distribution.^{11, 12}

You (2016) Log-linear model on sampling variances, denoted as YLLM used (5), (6) and (7) and then

$$\begin{aligned} \theta_i | \beta, \sigma_v^2 &\sim \text{ind } N(x_i' \beta, \sigma_v^2), \quad i = 1, \dots, m; \\ \log(\Psi_i) &\sim N(\delta_1 + \delta_2 \log(n_i), \tau^2), \quad i = 1, \dots, m; \end{aligned}$$

Flat priors for unknown parameters: $\pi(\beta) \propto 1$, $\pi(\delta_1, \delta_2) \propto 1$, $\pi(\sigma_v^2) \propto 1$, $\pi(\tau^2) \propto 1$.

Note that model YLLM uses a log-linear model for the sampling variance Ψ_i , and extends the model proposed by Souza, Moura and Migon (2009) for sampling variances by using $\log(n_i)$ and adding a random effect to the regression part in the model.^{29, 48}

For sampling variance modeling, the HB models of You (2016) and Sugasawa, Tamae and Kubokawa (2017) perform equally well to improve the direct survey estimates. In this vein, Sugasawa et al. (2017) explore different priors for the sampling variances Ψ_i , adopting a fully Bayesian approach to estimation.^{29, 49}

Polettini (2017) induces shrinkage for the sampling variance estimates using a semiparametric Dirichlet process model with random variances.⁵⁰

Gharde et. al. (2012) used the SAR Model in hierarchical bayes framework such that:

$$\begin{aligned} y_{ij} | \beta + v_i, \Psi_i &\sim \text{ind } N(x_{ij}' \beta + v_i, \Psi_i), \quad i = 1, \dots, m, j = 1, \dots, n_i \\ \left(\frac{v}{\sigma_u^2} \right) &\sim N_m(0, G) \end{aligned}$$

Where $G = \sigma_u^2 [(I - \rho W)(I - \rho W)']^{-1}$ and n_i is the sample size for the i 'th area.

Assuming prior distribution for parameters are as

$$\beta \sim N_p(0, C), \\ \Psi_i \sim IG(a_0, b_0) \text{ and } \sigma_u^2 \sim IG(a_1, b_1)$$

Full conditional distribution of the sampling variance for the Gibbs sampler is as follows:

$$(\Psi_i | \beta, v, \sigma_u^2, Y) \sim IG[a_0 + \frac{n}{2}, b_0 + \frac{1}{2}[Y - (X\beta + Zv)]'[Y - (X\beta + Zv)]]$$

They used sensitive analysis and show that as the prior become more informative, posterior variances decrease.⁵¹

Soltani-Kermanshahi et. al. at 2017 in the Bayes method,⁵ assumed the distribution of HFE (Household Food Expenditure) at area i to be $N(\bar{y}_i, \Psi_i)$ Where $\bar{y}_i$ assumed to be known but $1/\Psi_i$ as random with prior distribution

$$r_i = \frac{1}{\Psi_i} \sim \text{Gamma}\left(1, \frac{1}{S_i^2}\right)$$

In which S_i^2 is the direct variance in area i .

Furthermore they calculated the posterior inverse variance as

$$r_i | \underline{y} \sim \text{Gamma}\left(1 + \frac{n_i}{2}, \left(S_i^2 + \frac{1}{2} \sum (y_{ij} - \bar{y}_i)^2\right)^{-1}\right)$$

and

$$\xrightarrow{\text{yields}} \hat{\Psi}_i = \frac{S_i^2 + \frac{1}{2} \sum (y_{ij} - \bar{y}_i)^2}{1 + \frac{n_i}{2}}$$

Also Mehrabi et. al. at 2023 in the Exponential Spatial Structure of SAE used the recent model to estimate the sampling variances in each area.⁶

Gao and Wakefield (2022), In order to account for variability in the estimated sampling variances, proposed a hierarchical Bayesian spatial area level model that smooths both the estimated means and sampling variances to produce point and interval estimates of small area proportions. Their model explicitly targets estimation of small area proportions rather than means of continuous variables and we consider examples of both moderate and low prevalence events. They demonstrate the performance of their approach via simulation and application to vaccination coverage and HIV prevalence data from the Demographic and Health Surveys.⁵²

Extended methods

Maiti et al. (2014) assume the same variance sampling model (4) but adopt an empirical Bayes approach, setting the prior

$$\Psi_i \sim \text{IG}(r, s)$$

And estimating $\{r, s, \sigma_u^2\}$ via maximum likelihood.⁵³

You and Chapman (2006) considered combined the sampling variance model (4) with the small area model Fay-Harriot to construct an integrated model. The integrated model borrows strength for small area estimates and sampling variance estimates simultaneously. The integrated HB modeling approach has thus been widely used in practice, for example, You (2008, 2016), Dass, Maiti, Ren and Sinha (2012), Sugawara, Tamae and Kubokawa (2017), Ghosh, Myung and Moura (2018), and Hidioglou, Beaumont and Yung (2019).^{29, 47, 49, 54-56}

The GVF approaches described thus far treat the resulting variance estimates as known, so the resulting Fay-Herriot estimates do not account for uncertainty in the variance model. Maples et al. (2016) address this (Mohadjer et al. (2012)) by combining a GVF with a sampling model for the direct variance estimates.^{28, 30}

Particularly You (2008) used smoothed design effects over time to model the sampling variance. Indirect modeling of sampling variance is actually a combination of smoothing and modeling.³⁹

Also, Maples et al. (2016) outline a resampling procedure for estimating degrees of freedom for their variance sampling model (4).³⁰

DISCUSSION

If the domain-specific sample size is not large enough to provide a direct and reliable estimate, a domain or an area is referred to as "small".¹ Then the biggest difference between small area studies and other studies was the inadequacy of samples in each area, but the rate of weakness was reduced using the SAE method. Therefore, in this method, any increase in the precision of estimates is very important. One of the things that increase the accuracy of the estimates is related to the high accuracy in estimating the sampling variance of the Fay-Herriot (and extension of it in spatial structure) method in area level small area. In our best knowledge, since there is no unique method for SV estimation, further study is required. We reviewed some of the commonly used sampling variance estimation methods in area level of SAE. According to the reviewed studies, no method can be reported as the best method for estimating the sampling variance in all conditions. Because each method needs different information, and on the other hand, simulation studies are needed to compare the methods but in general, we compared the discussed methods in Table 3.

Table 3. Pros And Cons of each Sampling Variance Estimation Methods in Area Level of Small Area Estimation

Methods	Advantage	Disadvantage
Direct Sampling Variance Estimator	Simple calculation	Highly variable and requires assumptions to be made about the distribution of the data
Design-Based Variance Estimator	Robust to assumptions about the distribution of the data.	May not account for complex sampling designs.
Generalized Variance Function (GVF)	• Can be used for nonlinear models and does not require prior information. • Consistent with a particular class of prediction models for estimating totals from stratified, two-stage cluster samples.	Computationally intensive and might be challenging to implement for large datasets.
Model-based	Flexible and can be applied to a wide range of small area parameters	Requires a considerable amount of expertise to specify and estimate the appropriate statistical model and may not be reliable
Empirical Bayes	• Stabilizes Variance Estimates • Improves Overall Model Performance • Handles Areas with Extremely Small Sample Sizes • Can Reduce MSE of the Final Estimates	• Increased Model Complexity and Risk of Misspecification • Computational Intensity • Sensitivity to Prior Distributions

We found limited studies comparing some methods with simulation. For example:

1. Rivest and Vandal (2003) compared method (4) with different degrees of freedom and showed that the improvements are more pronounced as the degrees of freedom gets small (they use a minimum of 4).³⁴

2. Gershunskaya and Lahiri (2005) examined alternative approaches to variance estimation for domain estimates in the Current Employment Statistics (CES) survey. In a simulation study with subsamples drawn from part of the CES sample they noted that design-based variance estimates have low bias but are unstable (very high CVs), while a synthetic model-based variance estimator was more stable but had substantial bias.⁴⁵

3. Soltani-Kermanshahi, et. al. (2017) compared four methods in Fay-Herriot method and showed that method (4) had approximately the best RRMSE (relative root mean square Error).⁵

4. Sugawara, Tamae and Kubokawa (2017), in order to account for variability in the estimated sampling variances, compared the proposed hierarchical Bayesian spatial area level model that smooths both the estimated means and sampling variances to produce point and interval estimates of small area proportions and through simulation, they have shown that inferences based on their model can improve upon those based on a model that only incorporates smoothing of means.⁴⁹

5. Mehrabi, et. al. (2023) compared four methods in the exponential structure of spatiality and showed that (4) and direct methods, respectively and approximately had the smallest and the highest Root Average Mean Square Errors (RAMSE) for all conditions.⁶

CONCLUSION

According to the reviewed studies, no method can be reported as the best method for estimating the sampling variance in all conditions. Because each method needs different information, and on the other hand, simulation studies are needed to compare the methods.

ACKNOWLEDGMENT

This research has been supported by Semnan University of Medical Sciences and health Services grant IR.SEMUMS.REC.1401.318.

Ethical Consideration

This research is the result of a research project approved by the research center for Social Determinants of Health and has an ethical code IR.SEMUMS.REC.1401.318.

Data Availability

The data presented in this study are openly available for readers upon request.

Conflict of Interest

The Author declares that there is no conflict of interest.

Funding Information

The Semnan University of Medical Sciences funded this research.

Author Contributions

Study concept and design, Acquisition of data, Analysis and interpretation of data, drafting of the manuscript, Critical revision of the manuscript for important intellectual content and Study supervision, all done by Dr Mojtaba Soltani. The author reviewed the manuscript.

REFERENCES

1. Rao J. Small Area Estimation.–New York: A John Wiley & Sons. Inc; 2003.
2. Rao JN, Molina I. Small area estimation: John Wiley & Sons; 2015.

3. Fay III RE, Herriot RA. Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*. 1979;74(366a):269-77.
4. Soltani-Kermanshahi M, Mehrabi Y, Kavousi A, Baghestani A-R, Mohammadi-Nasrabadi F, Aliakbari-Saba R. Households' food expenditures in urban areas of Iran: An application of small area estimation. *Epidemiology, Biostatistics and Public Health*. 2017;14(4).
5. Soltani-Kermanshahi M, Mehrabi Y, Kavousi A, Baghestani A-R, Mohammadi-Nasrabadi F. Effects of sampling variance estimation methods on precision of small area estimation. *Jp J Biostat*. 2017;14(2):93-106.
6. Mehrabi Y, Kavousi A, Soltani-Kermanshahi M. Sampling Variance Estimation Method and Precision of Small Area Estimation in the Exponential Spatial Structure. *Journal of Biostatistics and Epidemiology*. 2023.
7. Molina I, Salvati N, Pratesi M. Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics*. 2009;24:441-58.
8. Anselin L. *Spatial econometrics: methods and models*: Springer Science & Business Media; 1988.
9. Cressie N. *Statistics for spatial data*: John Wiley & Sons; 2015.
10. Mehrabi Y, Kavousi A, Baghestani A-R, Soltani-Kermanshahi M. Spatiality in small area estimation: A new structure with a simulation study. *Geospatial Health*. 2020;15(2).
11. You Y, Zhou QM. Hierarchical Bayes small area estimation under a spatial model with application to health survey data. *Survey Methodology*. 2011;37(1):25-37.
12. Gómez-Rubio V, Best N, Richardson S, Li G, Clarke P. *Bayesian statistics small area estimation*. 2010.
13. Särndal C-E, Swensson B, Wretman J. *Model assisted survey sampling*: Springer Science & Business Media; 2003.
14. Arora V, Lahiri P. On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistica Sinica*. 1997:1053-63.
15. Verma V, Betti G. *Sampling errors and design effects for poverty measures and other complex statistics*: Università di Siena, Dipartimento di metodi quantitativi; 2005.
16. Hernández F, Herrador M, Morales D. VARIANCE ESTIMATION OF SMALL AREA ESTIMATORS IN THE SPANISH LABOUR FORCE SURVEY.

17. Pratesi M, Salvati N. Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical methods and applications*. 2008;17:113-41.
18. You Y, Hidiroglou M. Sampling variance smoothing methods for small area proportion estimators. *Methodology Branch Working Paper, SRID-2012-08E*, Statistics Canada, Ottawa, Canada 2012.
19. Wolter KM, Wolter KM. *Introduction to variance estimation*: Springer; 2007.
20. Dick P. Modelling net undercoverage in the 1991 Canadian Census. *Survey Methodology*. 1995;21(1):45-54.
21. Jiang J, Lahiri P. Mixed model prediction and small area estimation. *Test*. 2006;15:1-96.
22. Otto MC, Bell WR, editors. *Sampling error modelling of poverty and income statistics for states*. American Statistical Association, Proceedings of the Section on Government Statistics; 1995.
23. Valliant R. Generalized variance functions in stratified two-stage sampling. *Journal of the American Statistical Association*. 1987;82(398):499-508.
24. Cho MJ, Eltinge JL, Gershunskaya J, Huff L, editors. *Evaluation of generalized variance function estimators for the US Current Employment Survey*. Proceedings of the American Statistical Association, Survey Research Methods Section; 2002: Citeseer.
25. Huff L, Eltinge J, Gershunskaya J, editors. *Exploratory analysis of generalized variance function models for the US Current Employment Survey*. Proceedings of the American Statistical Association, Survey Research Methods Section; 2002.
26. Eltinge J, Cho M, Hinrichs P. Use of generalized variance functions in multivariate analysis. In: *Proceedings of the American Statistical Association, Survey Research Methods Section* 2002:904-12.
27. Shiferaw Y, Galpin J. Area specific confidence intervals for a small area mean under the Fay-Herriot model. *Journal of the Iranian Statistical Society*. 2016;15(2):1-43.
28. Mohadjer L, Rao J, Liu B, Krenzke T, Van de Kerckhove W. Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *Journal of the Indian Society of Agricultural Statistics*. 2012;66(1):55-63.
29. You Y. Small area estimation using Fay-Herriot area level model with sampling variance smoothing and modeling. *Survey Methodology*. 2021;47(2):361-71.
30. Maples JJ, editor *Estimating Design Effects in Small Areas/Domains through Aggregation*.

- Proceedings of the Section on Survey Research Methods, Alexandria, VA, American Statistical Association; 2016.
31. Milla N. Small area estimation of poverty incidence with sampling error variances through generalized variance function. *Am J Theor Appl Statist.* 2017;6:72-8.
 32. You Y, Chapman B. Small area estimation using area level models and estimated sampling variances. *Survey Methodology.* 2006;32(1):97.
 33. Battese GE, Harter RM, Fuller WA. An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association.* 1988;83(401):28-36.
 34. Rivest L-P, Vandal N, editors. 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*; 2003: Laboratory for Research in Statistics and Probability, Carleton University.
 35. Wang J, Fuller WA. The mean squared error of small area predictors constructed with estimated area variances. *Journal of the American Statistical Association.* 2003;98(463):716-23.
 36. Kleffe J, Rao J. Estimation of mean square error of empirical best linear unbiased predictors under a random error variance linear model. *Journal of Multivariate Analysis.* 1992;43(1):1-15.
 37. You Y, Rao J. Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics.* 2002;30(1):3-15.
 38. You Y, Rao JN, Gambino J. Model-based unemployment rate estimation for the Canadian Labour Force Survey: a hierarchical Bayes approach. *Survey Methodology.* 2003;29(1):25-32.
 39. You Y, editor *Small area estimation using area level models with model checking and applications. Proceedings of the Survey Methods Section, Statistical Society of Canada*; 2008.
 40. Liu B, Lahiri P, Kalton G. Hierarchical Bayes modeling of survey-weighted small area proportions. *Survey Methodology.* 2014;40(1):1-13.
 41. Bell WR, editor *Examining sensitivity of small area inferences to uncertainty about sampling error variances. Proceedings of the American Statistical Association, Survey Research Methods Section*; 2008: Citeseer.
 42. Hawala S, Lahiri P. Variance modeling for domains. *Statistics and Applications.* 2018;16(1):399-409.
 43. Bell WR, editor *Bayesian sampling error modeling with application. Statistical Society of Canada, Proceedings of the Survey Methods Section*; 1995.

44. Nguyen TTT, Bell WR, Gomish JM, editors. Investigating model-based time series methods to improve estimates from monthly value of construction put-in-place surveys. Proceedings of the American Statistical Association; 2002: Citeseer.
45. Gershunskaya JB, Lahiri P, editors. Variance estimation for domains in the US Current Employment Statistics Program. Proceedings of the American Statistical Association, Survey Research Methods Section; 2005.
46. Gene Hwang J, Qiu J, Zhao Z. Empirical Bayes confidence intervals shrinking both means and variances. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2009;71(1):265-85.
47. Dass SC, Maiti T, Ren H, Sinha S. Confidence interval estimation of small area parameters shrinking both means and variances. *Survey Methodology*. 2012;38(2):173-87.
48. Souza DF, Moura FA, Migon HS. Small area population prediction via hierarchical models. *Survey Methodology*. 2009;35(2):203-14.
49. Sugasawa S, Tamae H, Kubokawa T. Bayesian estimators for small area models shrinking both means and variances. *Scandinavian Journal of Statistics*. 2017;44(1):150-67.
50. Polettini S. A generalised semiparametric Bayesian Fay–Herriot model for small area estimation shrinking both means and variances. 2017.
51. Gharde Y, Rai A, Chandra H. Hierarchical Bayes Small area approach for spatial data. *Indian Society of Agricultural Statistics*. 2012;66(2):259-68.
52. Gao PA, Wakefield J. A spatial variance-smoothing area level model for small area estimation of demographic rates. *arXiv preprint arXiv:220902602*. 2022.
53. Maiti T, Ren H, Sinha S. Prediction error of small area predictors shrinking both means and variances. *Scandinavian Journal of Statistics*. 2014;41(3):775-90.
54. You Y. An integrated modeling approach to unemployment rate estimation for sub-provincial areas of Canada. *Survey Methodology*. 2008;34(1):19.
55. Hidioglou MA, Beaumont J-F, Yung W. Development of a small area estimation system at Statistics Canada. *Survey Methodology*. 2019;45(1):101-26.
56. Ghosh M, Myung J, Moura FA. Robust Bayesian small area estimation. *Survey Methodology*. 2018;44(1):101-16.