Geographic differences in the prevalence of hypertension in Uganda

Lunyera, Joseph; Kirenga, Bruce; Stanifer, John W.; Kasozi, Samuel; van der Molen, Thys; Katagira, Wenceslaus; Kamya, Moses R.; Kalyesubula, Robert

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0201001

Empir Econ (2018) 55:147–167
https://doi.org/10.1007/s00181-018-1452-5

The spatial empirical Bayes predictor of the small area mean for a lognormal variable of interest and spatially correlated random effects

Dian Handayani[1,2,3] · **Henk Folmer**[2,4] · **Anang Kurnia**[3] · **Khairil Anwar Notodiputro**[3]

Received: 8 February 2017 / Accepted: 11 January 2018 / Published online: 26 May 2018

Abstract The standard small area estimator, the empirical best linear unbiased predictor (EBLUP), estimates small area parameters by way of linear mixed models. The EBLUP assumes normal and independent random small area effects as well as normal and independent random sampling errors. Under these assumptions, the variable of interest also follows a normal distribution. In practice, however, the above assumptions are often violated. The variable of interest is often non-normal and highly skewed, and the small areas are frequently spatially dependent. In this paper, we propose the spatial empirical Bayes predictor (SEBP) of the small area mean of a positively skewed variable of interest in the presence of spatial dependence among the random small area effects. We assume that the variable of interest follows a normal distribution after a log transformation and that its log transform is linked to some auxiliary variables by a nested error regression model. The SEBP is derived under the log-transformed nested error regression model. By way of simulation, we show that compared to its alterna-

✉ Khairil Anwar Notodiputro
khairil@apps.ipb.ac.id

Dian Handayani
dianh@unj.ac.id; dian99163@yahoo.com

Henk Folmer
h.folmer@rug.nl

Anang Kurnia
anangk@apps.ipb.ac.id

1 Department of Statistics, State University of Jakarta, Jakarta, Indonesia

2 Faculty of Spatial Sciences, University of Groningen, Groningen, The Netherlands

3 Department of Statistics, Bogor Agricultural University, Bogor, Indonesia

4 College of Economics and Management, Northwest Agriculture and Forestry University, Yangling, China

tives, i.e., the direct estimator which is solely based on the survey data for the small area under study, the EBLUP which does not take into account spatial dependence and skewness, the empirical Bayes predictor which takes into account skewness but not spatial dependence among the small areas, the SEBP has the smallest average relative bias and average relative root-mean-squared error for various combinations—though not all—of skewness and spatial correlation.

Keywords Small area estimation · Spatial empirical Bayes predictor · Spatial dependence · Skewed distribution

1 Introduction

Surveys usually only allow parameter estimation with an acceptable level of precision for large areas (for instance, the national, state, or provincial level). In recent years, the need for parameter estimates at lower administrative level (subpopulation or small areas) has increased. For example, to allocate funds to local governments, information about local parameters, such as poverty rates, unemployment rates, or mortality rates, is needed. However, data sets from surveys are often not large enough to allow direct estimation of subpopulation parameters with acceptable precision ("direct" refers here to an estimator based on the survey data for the subpopulation only). Such subpopulations are called "small areas" (Rao and Molina 2015), and the field of statistics dealing with estimation of parameter for such areas is called "small area estimation" (SAE).

Standard SAE method, EBLUP (empirical best linear unbiased predictor), is based on a linear mixed model (LMM) which relates the variable of interest to some fixed effects and random effects (McCulloch and Searle 2001). In EBLUP, the auxiliary information is considered as fixed effect and the small areas are considered as random effect. The variable of interest is assumed to follow normal distribution. The random area effect and the sampling errors are also assumed to follow normal distribution. Furthermore, there is no spatial dependence among the random area effects as well as among the sampling errors. In practice, however, these assumptions are frequently violated. In socioeconomic surveys, for example, the variable of interest is often skewed. Berg and Chandra (2014), Chandra and Chambers (2011), Bellow and Lahiri (2011), Wang and Fuller (2003), Slud and Maiti (2006), Kurnia and Chambers (2011) developed SAEs for non-normal, skewed, continuous variables of interest. Berg and Chandra (2014), Slud and Maiti (2006), Kurnia and Chambers (2011) proposed the empirical Bayes predictor (EBP) for a lognormally distributed variable of interest. Although the above references correct for non-normality and skewness, they assume independence among small areas.

In practice, the assumption of independence among the small areas is also frequently violated; there is often spatial dependence among them. Petrucci and Salvati (2004a, b), Salvati (2004), Petrucci and Salvati (2006), Pratesi and Salvati (2008), Molina et al. (2009) introduced the spatial empirical best linear unbiased predictor (SEBLUP) which relaxes the assumption of independence among the small areas. The above references showed that the estimated SEBLUP standard errors tend to be smaller than the estimated EBLUP standard errors, if there is strong spatial dependence among the small areas. Although the above references take spatial dependence among the small areas into account, they still assume a normal distribution for the variable of interest.

In practice, both a skewed variable of interest and spatial dependence among the small areas usually occur. Therefore, we propose the spatial empirical Bayes predictor (SEBP) which can be applied to account for both skewness and spatial dependence among the small areas. Our study assumes that the variable of interest is normally distributed after log transformation. Furthermore, the relationship between the log-transformed variable of interest and the auxiliary variables is assumed linear and is modeled by a nested error regression model.

The paper is organized as follows. Section 2 summarizes the EBLUP and EBP, and Sect. 3 presents the SEBP. Section 4 presents a Monte Carlo simulation of the performance of the SEBP compared to the EBP, EBLUP, and the direct estimator. Section 5 concludes.

2 A review of small area estimators based on a linear mixed model

2.1 The empirical best linear unbiased predictor (EBLUP)

Before going into detail, we note that population characteristics will be denoted by capitals and sample characteristics by lower cases. Consider a population U divided into M non-overlapping small areas denoted $U_i, i = 1, 2 \dots M$. U_i consists of N_i elements so that population U has N elements ($N = N_1 + N_2 + \cdots + N_m$). Let for element j in the ith small area, p auxiliary data $x_{ij} = \left(x_{1ij}, x_{2ij}, \dots, x_{pij}\right)^T$ be available for variable of interest y_{ij} through the nested error regression model:

$$y_{ij} = x_{ij}^T \beta + z_{ij} v_i + e_{ij} \quad j = 1, 2 \dots N_i;\ i = 1, 2 \dots M \tag{1}$$

where β is the vector of regression parameters, z_{ij} is a known positive constant, and v_i and e_{ij} are random small area effects and random sampling errors, respectively, with distribution $v_i \sim N\left(0, \sigma_v^2\right)$ and $e_{ij} \sim N\left(0, \sigma_e^2\right)$. Furthermore, v_i and e_{ij} are assumed mutually independent. Hence, the variable of interest follows the normal distribution $y_{ij} \sim N\left(x_{ij}^T \beta, z_{ij}^2 \sigma_v^2 + \sigma_e^2\right)$.

Model (1) can be written in matrix notation as:

$$y_i = X_i \beta + Z_i v_i + e_i, \quad i = 1, 2 \dots M \tag{2}$$

where $y_i = \left[y_{i1}, y_{i2}, \dots y_{iN_i}\right]^T$ is the $(N_i \times 1)$ vector of the variable of interest, $X_i = \left[x_{i1}, x_{i2} \dots x_{iN_i}\right]^T$ the $(N_i \times p)$ matrix of unit-specific auxiliary variables, β the $(p \times 1)$ vector of regression coefficients, Z_i the $(N_i \times 1)$ vector with elements equal to positive constant, v_i the random small area effect associated with small area i, and e_i the $(N_i \times 1)$ vector of random sampling error. We assume that $v_i \sim N\left(0, \sigma_v^2 1_{N_i} 1_{N_i}^T\right)$; $e_i \sim N\left(0, \sigma_e^2 I_{N_i}\right)$, respectively. Note that 1_{N_i} is a column vector of dimension N_i

with all elements equal to one and I_{N_i} is $(N_i x N_i)$ identity matrix. For the population U, model (2) can be written in matrix notation as:

$$Y = X\beta + Zv + e \tag{3}$$

where Y is the $(N \times 1)$ vector of the variable of interest, X the $(N \times p)$ matrix of unit-specific auxiliary variables , β the $(p \times 1)$ vector of regression coefficients, Z a $(N \times M)$ matrix (defined below), v the $(M \times 1)$ vector of random small area effect, and e the $(N \times 1)$ vector of random sampling errors. Z is usually specified as (Petrucci and Salvati 2004b):

$$Z = \begin{bmatrix} 1_{N_1} & 0 & \cdots & 0 \\ 0 & 1_{N_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1_{N_M} \end{bmatrix}.$$

The mean of Y in (3) is $X\beta$, and its covariance matrix is $V = \sigma_v^2 ZZ^T + \sigma_e^2 I_N$. The covariance matrix V has dimensions (NxN) .

In this paper, we assume that each small area in population U is sampled, although in some applications, it could happen that some of small areas are not sampled. We also assume that there is no bias in sampling the small areas so that population model (1) holds for the sampled small areas:

$$y_{ij} = x_{ij}^T\beta + z_{ij}v_i + e_{ij}; \quad i = 1, 2 \ldots M; \quad j = 1, 2 \ldots n_i. \tag{4}$$

Suppose the parameter of interest is the i^{th} small area mean defined as:

$$\mu_i = \frac{1}{N_i}\sum_{i=1}^{N_i} y_{ij} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} y_{ij}\right]; \quad i = 1, 2 \ldots M \tag{5}$$

where s_i denotes the units sampled in small area i and r_i the non-sampled units. There are n_i units in s_i and $(N_i - n_i)$ units in r_i. Sample size n_i is assumed unequal to zero but too small to produce a direct estimator of μ_i with adequate precision. Note that $\sum_{j\in r_i} y_{ij}$ in (5) is the total of non-sampled y_{ij}. Under model (4), the non-sampled values of y_{ij} will be estimated by $\hat{y}_{ij} = E\left(y_{ij}|v_i\right) = x_{ij}^T\hat{\beta} + z_{ij}v_i$.

Under model (4), the best linear unbiased predictor (BLUP) of μ_i is:

$$\hat{\mu}_i^{\text{BLUP}} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\text{BLUP}}\right]; \quad i = 1, 2 \ldots M \tag{6}$$

where $\hat{y}_{ij}^{\text{BLUP}} = E\left(y_{ij}|v_i\right) = x_{ij}^T\hat{\beta} + z_{ij}v_i = x_{ij}^T\hat{\beta} + z_{ij}\gamma_i\left(\bar{y}_{is} - \bar{x}_{is}^T\hat{\beta}\right)$ with $\gamma_i = \frac{(z_{ij}^2\sigma_v^2)}{(z_{ij}^2\sigma_v^2+\sigma_e^2/n_i)}$ is the shrinkage factor (the ratio between the model variance relative to the total variance); $\hat{\beta} = \left[\sum_{i=1}^{M} x_i x_i^T / \left(\frac{\sigma_e^2}{n_i} + \sigma_v^2 z_{ij}^2\right)\right]^{-1} \left[\sum_{i=1}^{M} x_i y_i^T / \left(\left(\frac{\sigma_e^2}{n_i} + \sigma_v^2 z_{ij}^2\right)\right)\right] = \left(\sum_{i=1}^{M} X_i^T V_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{M} X_i^T V_i^{-1} y_i\right) = \left(X^T V^{-1} X\right)^{-1} \left(X^T V^{-1} Y\right)$ is the weighted least squares estimator of β; and $\bar{y}_{is}$ and $\bar{x}_{is}$ are the sample averages of y_{ij} and x_{ij}, respectively.

In practice, the variances components σ_v^2 and σ_e^2 are usually unknown. They can be estimated from the sample data using restricted maximum likelihood (REML) or maximum likelihood (ML). See Rao and Molina (2015) for details. By replacing $\left(\sigma_v^2, \sigma_e^2\right)$ in (6) by their estimates $\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right)$, the empirical best linear unbiased predictor (EBLUP) of μ_i is obtained as:

$$\hat{\mu}_i^{\text{EBLUP}} = \frac{1}{N_i}\left[\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij}^{\text{EBLUP}}\right]; \quad i = 1, 2 \ldots M \tag{7}$$

where $\hat{y}_{ij}^{\text{EBLUP}} = x_{ij}^T \hat{\beta} + z_{ij} \hat{v}_i = x_{ij}^T \hat{\beta} + z_{ij} \hat{\gamma}_i \left(\bar{y}_{is} - \bar{x}_{is}^T \hat{\beta}\right)$; $\hat{\gamma}_i = \frac{(z_{ij}^2 \hat{\sigma}_v^2)}{(z_{ij}^2 \hat{\sigma}_v^2 + \hat{\sigma}_e^2 / n_i)}$; $\hat{\beta} = \left[\sum_{i=1}^{M} x_i x_i^T / \left(\frac{\hat{\sigma}_e^2}{n_i} + \hat{\sigma}_v^2 z_{ij}^2\right)\right]^{-1} \left[\sum_{i=1}^{M} x_i y_i^T / \left(\frac{\hat{\sigma}_e^2}{n_i} + \hat{\sigma}_v^2 z_{ij}^2\right)\right]$.

An approximate estimator of the MSE $\left(\hat{\mu}_i^{\text{EBLUP}}\right)$ subject to approximations of order $O\left(m^{-1}\right)$ is given by (Prasad and Rao 1990):

$$\widehat{\text{MSE}}\left(\hat{\mu}_i^{\text{EBLUP}}\right) = \text{mse}\left(\hat{\theta}_i^{\text{EBLUP}}\right) = g_{1i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) + g_{2i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) + 2 g_{3i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) \tag{8}$$

where $g_{1i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right)$ is due to the prediction of the small area effect v_i and is $O(1)$, $g_{2i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right)$ is due to estimating β and is $O\left(m^{-1}\right)$ for large m, and $g_{3i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right)$ is due to estimating σ_v^2 and is $O\left(m^{-1}\right)$, where m is the number of small areas that are selected in the sample. For details about the derivation of $\widehat{\text{MSE}}\left(\hat{\mu}_i^{\text{EBLUP}}\right)$, we refer to Prasad and Rao (1990) and Rao and Molina (2015).

2.2 The empirical Bayes predictor (EBP) of the lognormal response

In many socioeconomic analyses, especially for income or expenditure problems, the variable of interest is positively skewed. In this case, EBLUP is an inaccurate estimator of small area parameters because of violation of the normality assumption. Berg and Chandra (2014), Kurnia and Chambers (2011), Slud and Maiti (2006) developed the empirical Bayes predictor (EBP) which assumes that the variable of interest follows a lognormal distribution, $y_{ij} \sim LN\left(\mu_i, \sigma^2\right)$ and the log transform of y_{ij} a normal distribution. The log transform of y_{ij} is linked to some auxiliary variables under the nested error regression model as follows:

$$\log y_{ij} = y_{ij}^* = x_{ij}^T \beta + z_{ij} v_i + e_{ij}; \quad i = 1, 2 \ldots M, \quad j = 1, 2 \ldots N_i \tag{9}$$

where v_i and e_{ij} are mutually independent error terms with $v_i \sim N\left(0, \sigma_v^2\right)$ and $e_{ij} \sim N\left(0, \sigma_e^2\right)$. Then, y_{ij}^* follows a normal distribution with mean $x_{ij}^T\beta$ and variance $z_{ij}^2\sigma_v^2 + \sigma_e^2$; $y_{ij}^* \sim N\left(x_{ij}^T\beta, z_{ij}^2\sigma_v^2 + \sigma_e^2\right)$.

Because $y_{ij} \sim LN\left(\mu_i, \sigma^2\right)$ and based on the properties of the lognormal distribution, the mean of y_{ij} is:

$$E\left(y_{ij}\right) = \mu_i = e^{x_{ij}^T\beta + \frac{1}{2}\left(z_{ij}^2\sigma_v^2 + \sigma_e^2\right)} \tag{10}$$

and its variance is:

$$\mathrm{Var}\left(y_{ij}\right) = \sigma^2 = e^{2\left[x_{ij}^T\beta + \frac{1}{2}\left(z_{ij}^2\sigma_v^2 + \sigma_e^2\right)\right]}\left[e^{\left(e^{z_{ij}^2\sigma_v^2 + \sigma_e^2}\right)} - 1\right] \tag{11}$$

Suppose the parameter of interest is the ith small area mean $\mu_i = \frac{1}{N_i}\sum_{i=1}^{N_i} y_{ij} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} y_{ij}\right]$. To obtain $\sum_{j\in r_i} y_{ij}$, Kurnia and Chambers (2011) estimated the non-sampled values of y_{ij} as $\hat{y}_{ij} = E\left(y_{ij}\right)$. Then, the Bayes predictor (BP) for μ_i is:

$$\hat{\mu}_i^{\mathrm{BP}} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\mathrm{BP}}\right] \tag{12}$$

where $\hat{y}_{ij}^{\mathrm{BP}} = e^{x_{ij}^T\hat{\beta} + \frac{1}{2}\left(z_{ij}^2\sigma_v^2 + \sigma_e^2\right)}$.

In practice, σ_v^2 and σ_e^2 are unknown and they are estimated by $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$, respectively, using REML or ML. The empirical Bayes predictor (EBP) of μ_i is obtained by replacing the unknown parameter σ_v^2 and σ_e^2 by their estimates:

$$\hat{\mu}_i^{\mathrm{EBP}} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\mathrm{EBP}}\right] \tag{13}$$

where $\hat{y}_{ij}^{\mathrm{EBP}} = e^{x_{ij}^T\hat{\beta} + \frac{1}{2}\left(z_{ij}^2\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right)}$.

If y_{ij} is not strictly lognormally distributed, then $\hat{\mu}_i^{\mathrm{EBP}}$ tends to be biased. Following Karlberg (2000), Kurnia and Chambers (2011) proposed the following bias correction for (13):

$$c_{ij} = 1 + \frac{1}{2}x_{ij}\mathrm{Var}\left(\hat{\beta}\right)x_{ij}^T + \frac{1}{8}\nabla\left(z_{ij}^2\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right) \tag{14}$$

where $\nabla(.)$ is the asymptotic variance–covariance matrix of total variability of response.

As a result, the EBP of μ_i with bias correction (14) is:

$$\hat{\mu}_i^{\text{EBP}} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\text{EBP}}\right] \tag{15}$$

where $\hat{y}_{ij}^{\text{EBP}c} = \left(c_{ij}\right)^{-1} e^{x_{ij}^T\hat{\beta}+\frac{1}{2}\left(z_{ij}^2\hat{\sigma}_v^2+\hat{\sigma}_e^2\right)}; c_{ij} = 1+\frac{1}{2}x_{ij}V\left(\hat{\beta}\right)x_{ij}^T+\frac{1}{8}\nabla\left(z_{ij}^2\hat{\sigma}_v^2+\hat{\sigma}_e^2\right)$ with $V\left(\hat{\beta}\right) = \left[\left(X^T V^{-1} X\right)^T\right]^{-1}$ the variance of $\hat{\beta}$

3 The spatial empirical Bayes predictor (SEBP)

The EBP (15) is derived under model (9) with the assumption that the random small area effects v_i are independent. However, in many applications, the response of unit in a given area toward certain characteristics often could influence the response from unit in other area. For example, house prices in neighboring small areas tend to mutually influence each other. In other words, there is spatial dependence among small areas.

Spatial dependence among small areas can be taken into account by spatially correlated random small area effects. Model (9) for unit i in spatially correlated small area j thus becomes:

$$y_{ij}^{\#} = \log y_{ij} = x_{ij}^T\beta + z_{ij}u_i + e_{ij};\quad i = 1, 2\ldots M,\quad j = 1, 2\ldots N_i \tag{16}$$

where u_i is the random small area effect which is assumed to follow a spatial error autoregressive (SAR) process with spatial correlation coefficient ρ and spatial weights matrix W.

Model (16) for the population can be written in matrix notation as follows:

$$Y^{\#} = X\beta + Zu + e \tag{17}$$

where $Y^{\#} = \left(y_{11}^{\#}\ldots y_{1N_1}^{\#}\ldots y_{M1}^{\#}\ldots y_{1N_M}^{\#}\right)^T = (\log y_{11}, \ldots \log y_{1N_1}, \ldots \log y_{M1}, \ldots \log y_{MN_M})^T$, $u = \rho Wu + v \Rightarrow u = (I-\rho W)^{-1}v; v \sim N\left(0, \sigma_v^2 I\right)$, $e \sim N\left(0, \sigma_e^2 I\right)$. Furthermore, the random small area effect u is assumed to follow a normal distribution: $u \sim N(0, D); D = \sigma_v^2\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}$. Note that the dimension of $Y^{\#}$ is $(N\times 1)$, $N = N_1 + N_2 + \cdots + N_M$, X is $(N\times p)$, Z is $(N\times M)$,u is $(M\times 1)$, v is $(M\times 1)$, and the weights matrix W is $(M\times M)$. Although there are many types of weights matrix W, we apply the symmetric binary weights matrix with elements $\{w_{ik}\} = 1$, if small area i is adjacent to small area k and zero otherwise $(i \neq k = 1, 2\ldots M)$. The matrix W is row-standardized such that the row elements sum to one. Thus, $y_{ij}^{\#}$ follows a normal distribution: $y_{ij}^{\#} = \log y_{ij} \sim N\left(x_{ij}^T\beta, z_{ij}^2\tau_i^2 + \sigma_e^2\right); \tau_i^2 = b_i^T D b_i$ with b_i^T a $(1xM)$ vector with 1 in the position of small area i.

We assume that there is no sampling bias for small areas so that population model (17) holds for the sampled small areas:

$$y_{ij}^{\#} = \log y_{ij} = x_{ij}^{T}\beta + z_{ij}u_i + e_{ij}; i = 1, 2 \ldots M, j = 1, 2 \ldots n_i. \tag{18}$$

Suppose the parameter of interest is ith small area mean, $\mu_i = \frac{1}{N_i}\sum_{i=1}^{N_i} y_{ij} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} y_{ij}\right]$. Under model (18) and following Kurnia and Chambers (2011), the spatial Bayes predictor (SBP) for μ_i is given by:

$$\hat{\mu}_i^{\text{SBP}} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{\mu}_{ij}^{\text{SBP}}\right]; \quad i = 1, 2 \ldots M, j = 1, 2 \ldots n_i \tag{19}$$

where $\hat{y}_{ij}^{\text{SBP}} = E\left(y_{ij}\right) = e^{x_{ij}^{T}\hat{\beta}+\frac{1}{2}\left(z_{ij}^{2}\tau_i^{2}+\sigma_e^{2}\right)}$, $\tau_i^2 = b_i^T D b_i$. The $\hat{y}_{ij}^{\text{SBP}}$ is obtained by using the property of the lognormal distribution (note that if $y_{ij} \sim LN\left(\mu_i, \sigma^2\right)$, then $\mu_i = E\left(y_{ij}\right) = e^{x_{ij}^{T}\beta+\frac{1}{2}\left(z_{ij}^{2}\sigma_i^{2}+\sigma_e^{2}\right)}$).

The spatial empirical Bayes predictor (SEBP) of μ_i, $\hat{\mu}_i^{\text{SEBP}}$, is derived by replacing $\left(\sigma_v^2, \sigma_e^2, \rho\right)$ in (19) by the estimates $\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2, \hat{\rho}\right)$. The estimates $\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2, \hat{\rho}\right)$ can be obtained by ML or REML. As in the case of the EBP proposed by Kurnia and Chambers (2011), if the log-transformed variable of interest does not strictly follow a normal distribution, then $\hat{\mu}_i^{\text{SEBP}}$ tends to be biased for μ_i. In that case, the modified Karlberg bias correction factor (see "Appendix 1") can be applied to $\hat{\mu}_i^{\text{SEBP}}$ as follows:

$$\hat{\mu}_i^{\text{SEBP\_}c} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\text{SEBP\_}c}\right]; \quad i = 1, 2 \ldots M \tag{20}$$

where

$$\hat{y}_{ij}^{\text{SEBP\_}c} = \left(c_{ij}^{\text{SEBP}}\right)^{-1} e^{x_{ij}^{T}\hat{\beta}+\frac{1}{2}\left(z_{ij}^{2}\hat{\tau}_i^{2}+\hat{\sigma}_e^{2}\right)}$$

$$c_{ij}^{\text{SEBP}} = 1 + \frac{1}{2}\left[D_1 V\left(\hat{\beta}\right) + \mathrm{D}_2 V\left(\hat{\rho}\right) + D_3 V\left(\hat{\sigma}_v^2\right) + D_4 V\left(\hat{\sigma}_e^2\right)\right]$$

$$D_1 = x_{ij}x_{ij}^T; D_2 = \left[A_1A_2A_1A_2 + A_3A_2 + A_1A_4\right]; D_3 = A_5A_5; D_4 = \frac{1}{4}; D_5 = \frac{1}{4}A_5$$

$$A_1 = \frac{1}{2}z_{ij}^2 b_i^T \sigma_v^2 \left\{\left(I - \rho W^T\right)(I - \rho W)\right\}^{-1}$$

$$A_2 = \left[W^T + W - 2\rho W^T W\right]\left\{\left(I - \rho W^T\right)(I - \rho W)\right\}^{-1} b_i$$

$$A_3 = \frac{1}{2}z_{ij}^2 b_i^T \sigma_v^2 \left\{\left(I - \rho W^T\right)(I - \rho W)\right\}^{-1}\left[W^T + W - 2\rho W^T W\right]\left\{\left(I - \rho W^T\right)(I - \rho W)\right\}^{-1}$$

$$A_4 = \left[-2W^T W\right]\left\{\left(I - \rho W^T\right)(I - \rho W)\right\}^{-1} b_i + \left[W^T + W - 2\rho W^T W\right]$$
$$\left\{\left\{\left(I - \rho W^T\right)(I - \rho W)\right\}^{-1}\left[W^T + W - 2\rho W^T W\right]\left\{\left(I - \rho W^T\right)(I - \rho W)\right\}^{-1} b_i\right\}$$

$$A_5 = \frac{1}{2} z_{ij}^2 b_i^T \left\{ \left(I - \rho W^T \right) (I - \rho W) \right\}^{-1} b_i.$$

The MSE of $\hat{\mu}_i^{\text{SEBP\_}c}$ can be approximated as:

$$\text{MSE}\left(\hat{\mu}_i^{\text{SEBP\_}c}\right) \approx \text{Var}\left\{ N_i^{-1} \sum_{j \in s_i} y_{ij} \right\} + N_i^{-2} \left\{ \sum_{j=1}^{(N_i - n_i)} e^{2x_{ij}^T \beta + \left(z_{ij}^2 \tau_i^2 + \sigma_e^2\right)} \right.$$
$$\left[e^{\left(z_{ij}^2 \tau_i^2 + \sigma_e^2\right)} - 1 \right] + 2 \sum_{1 \le j < k \le (N_i - n_i)} E\left[\left(x_{ij}^T \beta + z_{ij} u_i \right) \left(x_{ik}^T \beta + z_{ik} u_i \right) \right]$$
$$\left. - e^{x_{ij}^T \beta + \frac{1}{2}\left(z_{ij}^2 \tau_i^2 + \sigma_e^2\right)} e^{x_{ik}^T \beta + \frac{1}{2}\left(z_{ik}^2 \tau_i^2 + \sigma_e^2\right)} \right\} + N_i^{-2} \left\{ A^2 V(\hat{\rho}) + \nabla\left(B^2 \hat{\sigma}_v^2 + C \hat{\sigma}_e^2 \right) \right\};$$

where ∇ (.) is the asymptotic variance–covariance matrix total variability of response. Its derivation is presented in "Appendix 2." Since we have not yet derived the estimator of the MSE of $\hat{\mu}_i^{\text{SEBP\_}c}$, we do not use it any further. The relative performance of $\hat{\mu}_i^{\text{SEBP\_}c}$ is analyzed by means of simulation in the next section.

4 The performance of the SEBP compared to the EBP, EBLUP, and the direct estimator: evidence from Monte Carlo simulation

This section presents the results of a simulation of the performance of the SEBP compared to the EBP, EBLUP, and the direct estimator. To this end, the values of log y_{ij} are generated based on model (18) with single auxiliary variable X. The number of small areas is set at 30. To get insight into the impacts of the small area sample size, we consider small, medium, and large sample size ranging between 5 and 35. As a rule of thumb, samples larger than 30 are considered "large." We set the total sample size at 600 which is approximately 3% of the total population. Consequently, the population will have 20,000 elements (N = 20,000).

We generate random effects v from N (0, 0.09) and random sampling errors e from N (0, 0.25). We generate two sets of Y values which follow lognormal distributions with the same mean but different variances. To obtain the two sets of Y values, we fix $\beta = (\beta_0, \beta_1) = (2, 1)$ and generate $X \sim N$ (8.5, 4) and $X \sim N$ (6, 9), respectively. The generated X values from $X \sim N$ (8.5, 4) will yield Y values with smaller variance than the Y values based on $X \sim N$ (6, 9). We fix the spatial correlation at $\rho = 0.25$, 0.5, 0.75 to represent small, moderate, and large spatial correlation, respectively.

Based on the above specifications, there are six synthetic spatial populations: Populations 1–3 with $X \sim N$ (8.5, 4), $\rho = 0.25$, $\rho = 0.5$, and $\rho = 0.75$, respectively, and Populations 4–6 with $X \sim N$ (6, 9), $\rho = 0.25$, $\rho = 0.5$, and $\rho = 0.75$, respectively. The proximity matrix W is a symmetric binary weights matrix with w_{ij} equal to 1 if small area i is adjacent to small area j, and 0 otherwise. Further explanation about optimum spatial weighted matrix in SAE can be found in Asfar et al (2016). Based on these specifications and under model (19), the values of $\log\left(y_{ij}\right)$ for the six synthetic populations are determined. The transformation $y_{ij} = \exp\left(\log\left(y_{ij}\right)\right)$ gives the y_{ij} on the basis of which the means $E\left(y_{ij}\right)$ of all six synthetic populations are calculated. These means are the "true" population means.

For each synthetic population, we draw a 3% sample. Then, we combine the sample values of the 30 small areas to estimate the model parameters (β, σ_v^2, σ_e^2, ρ). These parameters are needed to obtain the SEBP (21). By combining the sample values, we "borrow strength," i.e., enlarge the data set to estimate the model parameters. Note that "borrowing strength" is an essential feature of SAE. To obtain small area estimators with adequate precision, one often applies indirect estimators that "borrow strength" by using values of the variable of interest from related areas and/or time periods which thus increases the "effective" sample size. These values are used in the estimation process through a statistical model.

Based on the combined sample values and under model (18), σ_v^2, σ_e^2, ρ are estimated using REML. The advantage of using REML instead of ML to estimate $\left(\sigma_v^2, \sigma_e^2, \rho\right)$ is that it takes into account the loss of degrees of freedom due to estimating β (Rao and Molina 2015). Furthermore, Jiang (1996) showed that REML is asymptotically consistent when normality does not hold. Based on the estimates $\hat{\sigma}_v^2$, $\hat{\sigma}_e^2$, $\hat{\rho}$, we obtain $\hat{\beta}$ and the estimates of the random area effects u_i. Next, we calculate the conditional mean $E\left(y_{ij}\right)$ under model (18) which is used as an estimate of the non-sampled y_{ij}. Next, the predictions of the non-sampled y_{ij} are combined with the sampled y_{ij}. The combined values are used to estimate the small area means (model 18). The procedure is applied to each of the six synthetic population.

For each synthetic population, $T = 1000$ replicates are generated. For each sample, the population mean is estimated by the SEBP, EBP, EBLUP, and direct. We evaluate these estimators by means of the average relative bias ($\overline{\text{ARB}}$) and the average relative root-mean-squared error ($\overline{\text{ARRMSE}}$) which are defined as:

$$\overline{\text{ARB}} = \frac{1}{M}\sum_{i}^{M}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{\mu}_{it}}{\mu_i} - 1\right) \times 100\% \tag{21}$$

$$\overline{\text{ARRMSE}} = \frac{1}{M}\sum_{i}^{M}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{\mu}_{it}}{\mu_i} - 1\right)^2} \times 100\% \tag{22}$$

The results of the simulations are reported in Tables 1 and 2 as well as in Figs. 1 and 2. The results in Table 1 and Fig. 1 are for moderately skewed data ($X \sim N$ (8, 5, 4)) and those in Table 2 and Fig. 2 for heavily skewed data ($X \sim N$ (6, 9)).

Table 1 and Fig. 1 show that for moderately skewed data, direct is best in terms of $\overline{\text{ARB}}$, except for $\rho = 0.5$, but worst in terms of $\overline{\text{ARRMSE}}$ everywhere. Overall, the EBLUP has the largest $\overline{\text{ARB}}$, while its $\overline{\text{ARRMSE}}$ is smaller than that of the direct but larger than that of the EBP and the SEBP. In terms of $\overline{\text{ARB}}$, the SEBP outperforms the EBLUP and the EBP everywhere and the direct for $\rho = 0.5$. In terms of $\overline{\text{ARRMSE}}$, the EBP and the SEBP perform approximately equally well and outperform their alternatives.

Table 2 and Fig. 2 show that for heavily skewed data, direct outperforms its alternatives in terms of $\overline{\text{ARB}}$, but its $\overline{\text{ARRMSE}}$ is worst for $\rho = 0.25$ and the next worst for $\rho = 0.5$ after EBLUP. The SEBP outperforms the EBP in terms of $\overline{\text{ARB}}$, except for

Table 1 $\overline{\text{ARB}}$ and $\overline{\text{ARRMSE}}$ of direct, EBP, EBLUP, and SEBP for moderately skewed data

ρ	Measure	Direct	EBLUP	EBP	SEBP
$\rho = 0.25$	$\overline{\text{ARB}}$ (%)	0.70	15.77	− 5.21	− 1.22
	$\overline{\text{ARRMSE}}$ (%)	144.38	42.20	12.57	13.40
$\rho = 0.5$	$\overline{\text{ARB}}$ (%)	0.72	11.31	− 4.91	0.48
	$\overline{\text{ARRMSE}}$ (%)	133.67	20.79	8.63	9.46
$\rho = 0.75$	$\overline{\text{ARB}}$ (%)	0.88	12.93	− 7.88	− 4.99
	$\overline{\text{ARRMSE}}$ (%)	123.95	23.93	8.37	8.46

Table 2 $\overline{ARB}$ and $\overline{ARRMSE}$ of direct, EBP, EBLUP, and SEBP for heavily skewed data

ρ	Measure	Direct	EBLUP	EBP	SEBP
$\rho = 0.25$	$\overline{\text{ARB}}$ (%)	− 1.77	64.88	− 5.36	− 2.78
	$\overline{\text{ARRMSE}}$ (%)	512.17	486.50	13.84	14.36
$\rho = 0.5$	$\overline{\text{ARB}}$ (%)	0.14	120.19	6.46	9.37
	$\overline{\text{ARRMSE}}$ (%)	548.17	1689.75	25.97	27.83
$\rho = 0.75$	$\overline{\text{ARB}}$ (%)	0.31	133.89	9.98	6.93
	$\overline{\text{ARRMSE}}$ (%)	442.28	3799.33	67.74	63.70

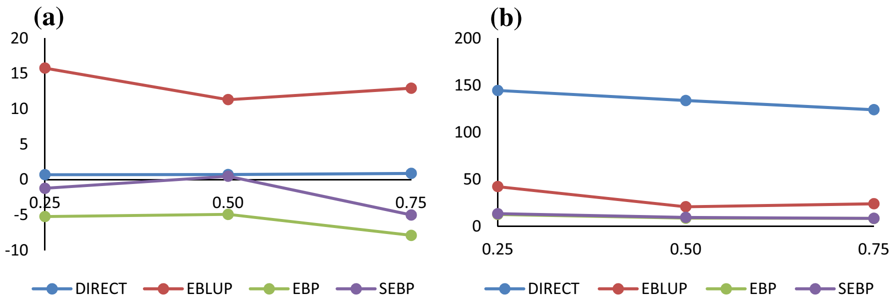

Fig. 1 The $\overline{\text{ARB}}$ (**a**) and $\overline{\text{ARRMSE}}$ (**b**) of direct, EBP, EBLUP, and SEBP for moderately skewed data

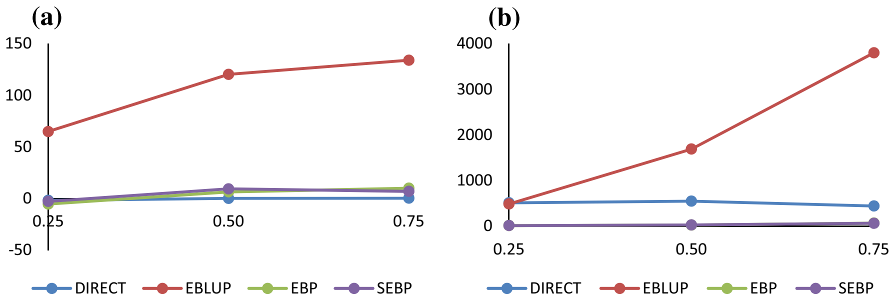

Fig. 2 The $\overline{\text{ARB}}$ (**a**) and $\overline{\text{ARRMSE}}$ (**b**) of direct, EBP, EBLUP, and the SEBP for heavily skewed data

$\rho = 0.5$. In terms of $\overline{\text{ARRMSE}}$, the EBP and the SEBP perform approximately equally well. Particularly, the EBP has slightly smaller $\overline{\text{ARRMSE}}$ for small and medium spatial correlation. For strong spatial correlation, the SEBP clearly outperforms EBP.

The following overall conclusions can be drawn from the simulations. First, although direct has the smallest average relative bias, its average relative root-mean-squared error is so large that its applicability is very limited indeed. Secondly, the results for the EBLUP show that failure to control for non-normality leads to substantial bias and root-mean-squared error. This applies to both moderately and heavily skewed data. Thirdly, taking into account spatial dependence among the random area effects tends to improve the bias and, to a less extent, the root-mean-squared error, though not uniformly, as shown by the comparison between the EBP and the SEBP.

5 Concluding remarks

Surveys usually only allow parameter estimation with an acceptable level of precision for large areas (for instance, the national, state, or provincial level). In recent years, the need for parameter estimates at lower administrative levels (subpopulations or small areas) has increased. For example, to allocate funds to local governments, information about local parameters, such as the poverty rate, the unemployment rate, or the mortality rate, is needed. Unfortunately, data sets from surveys are often not large enough to yield direct estimators of small area parameters with acceptable precision. Small area estimation (SAE) has been developed as a subfield of statistics to deal with estimation of parameters for small areas.

Standard SAE methods based on a linear mixed model assume normality and independence among random small area effects as well as normality and independence among the random sampling errors. In practice, however, these assumptions are frequently violated. In socioeconomic surveys, for example, the variable of interest is typically skewed. Moreover, there is often spatial dependence among small areas.

In this paper, we propose the spatial empirical Bayes predictor (SEBP) of the small area mean of a positively skewed variable of interest in the presence of spatial dependence among the random small area effects. The SEBP is derived under a log-transformed nested error regression model. By means of simulation, we analyzed the performance of the SEBP relative to the direct estimator (direct), the empirical best linear predictor (EBLUP) which does not take into account spatial dependence and skewness, and to the empirical Bayes predictor (EBP) which takes into account skewness but not spatial dependence among the small areas. The main finding is that in terms of average relative bias and average root-mean-squared error, the SEBP performs well, in the case of both moderately and heavily skewed data. Specifically, the SEBP has the smallest average relative bias and average relative root-mean-squared error for various combinations, though not all, of skewness and spatial correlation.

The simulations indicate that SEBP performs relatively well. It has the smallest bias compared to EBP, EBLUP, and direct estimator. Based on the results, the SEBP could potentially have smaller MSE. We have derived the theoretical mean-squared error of the SEPB $MSE\left(\hat{\mu}_i^{SEBP}\right)$ in this paper but not its estimator. To obtain the estimate of the MSE, we will investigate the expectation of the MSE. Derivation of the latter is a next step for our works.

Acknowledgements The authors thank the reviewers for helpful comments and suggestions which led to considerable improvements in the paper.

Appendix 1: The bias correction for $\hat{y}_{ij}^{\mathrm{SEBP}}$ if the log-transformed variable of interest does not strictly follow a normal distribution

The spatial empirical Bayes predictor (SEBP) of μ_i under model (18) is:

$$\hat{\mu}_i^{\mathrm{SEBP}} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\mathrm{SEBP}}\right] \tag{A1.1}$$

where $\hat{y}_{ij}^{\mathrm{SEBP}} = E\left(y_{ij}|\left(\boldsymbol{y}_{is}, \boldsymbol{x}_{is}\right)\right) = e^{x_{ij}^T\hat{\beta}+\frac{1}{2}\left(z_{ij}^2\hat{\tau}_i^2+\hat{\sigma}_e^2\right)}$; $\hat{\tau}_i^2 = b_i^T\hat{D}b_i = b_i^T\hat{\sigma}_v^2\left\{\left(I-\hat{\rho}W^T\right)\left(I-\hat{\rho}W\right)\right\}^{-1}b_i$; b_i^T is a $(1\times M)$ vector (0,0,…0,1,0,0…) with 1 for the ith area,

$$\hat{\beta} = \left[\sum_{i=1}^{M} x_i x_i^T \Big/ \left(\frac{\sigma_e^2}{n_i} + \sigma_v^2 z_{ij}^2\right)\right]^{-1}\left[\sum_{i=1}^{M} x_i y_i^T \Big/ \left(\left(\frac{\sigma_e^2}{n_i} + \sigma_v^2 z_{ij}^2\right)\right)\right]$$

Suppose: $\psi_{ij}(\eta) = e^{x_{ij}^T\beta+\frac{1}{2}\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)}$; $\tau_i^2 = b_i^T\sigma_v^2\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i$; $\eta = \left(\beta, \rho, \sigma_v^2, \sigma_e^2\right)$

$$\begin{aligned}
\frac{\partial}{\partial\beta}\psi_{ij}(\eta) &= \psi_{ij}(\eta)\, x_{ij}^T \\
\frac{\partial^2}{\partial\beta^2}\psi_{ij}(\eta) &= \psi_{ij}(\eta)\, x_{ij}x_{ij}^T \\
\frac{\partial}{\partial\rho}\psi_{ij}(\eta) &= \psi_{ij}(\eta)\left[\frac{1}{2}z_{ij}^2\frac{\partial}{\partial\rho}\tau_i^2\right] \\
&= \psi_{ij}(\eta)\left[\frac{1}{2}z_{ij}^2\frac{\partial}{\partial\rho}b_i^T\sigma_v^2\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i\right] \\
&= \psi_{ij}(\eta)\frac{1}{2}z_{ij}^2 b_i^T\sigma_v^2\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1} \\
&\quad\left[W^T + W - 2\rho W^T W\right]\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i \\
&= \psi_{ij}(\eta)\, A_1 A_2;
\end{aligned}$$

$$\begin{aligned}
A_1 &= \frac{1}{2}z_{ij}^2 b_i^T\sigma_v^2\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1} \\
A_2 &= \left[W^T + W - 2\rho W^T W\right]\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i
\end{aligned}$$

$$\frac{\partial^2}{\partial \rho^2}\psi_{ij}(\eta) = \frac{\partial}{\partial \rho}\left[\frac{\partial}{\partial \rho}\psi_{ij}(\eta)\right] = \frac{\partial}{\partial \rho}\left[\psi_{ij}(\eta)A_1A_2\right] = \frac{\partial}{\partial \rho}[BA_2]; B = \psi_{ij}(\eta)A_1$$

$$\frac{\partial^2}{\partial \rho^2}\psi_{ij}(\eta) = \left[\frac{\partial}{\partial \rho}B\right]A_2 + B\left[\frac{\partial}{\partial \rho}A_2\right]$$

$$\begin{aligned}\frac{\partial}{\partial \rho}B &= \left[\frac{\partial}{\partial \rho}\psi_{ij}(\eta)\right]A_1 + \psi_{ij}(\eta)\left[\frac{\partial}{\partial \rho}A_1\right] = \psi_{ij}(\eta)A_1A_2A_1 + \psi_{ij}(\eta)A_3\\ &= \psi_{ij}(\eta)[A_1A_2A_1 + A_3]\end{aligned}$$

$$\frac{\partial}{\partial \rho}A_1 = \frac{1}{2}z_{ij}^2 b_i^T\sigma_v^2\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}\left[W^T + W - 2\rho W^T W\right]$$

$$\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1} = A_3$$

$$\begin{aligned}\frac{\partial}{\partial \rho}A_2 &= \left\{\frac{\partial}{\partial \rho}\left[W^T + W - 2\rho W^T W\right]\right\}\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i\\ &\quad+\left[W^T + W - 2\rho W^T W\right]\left\{\frac{\partial}{\partial \rho}\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i\right\}\\ &= \left[-2W^TW\right]\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i + \left[W^T + W - 2\rho W^T W\right]\\ &\quad\left\{\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}\left[W^T + W - 2\rho W^T W\right]\right.\\ &\quad\left.\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i\right\} = A_4\end{aligned}$$

$$\frac{\partial^2}{\partial \rho^2}\psi_{ij}(\eta) = \frac{\partial}{\partial \rho}[BA_2] = \left[\frac{\partial}{\partial \rho}B\right]A_2 + B\left[\frac{\partial}{\partial \rho}A_2\right]$$

$$\begin{aligned}\frac{\partial^2}{\partial \rho^2}\psi_{ij}(\eta) &= \psi_{ij}(\eta)[A_1A_2A_1 + A_3]A_2 + \psi_{ij}(\eta)A_1A_4\\ &= \psi_{ij}(\eta)[A_1A_2A_1A_2 + A_3A_2 + A_1A_4]\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial \sigma_v^2}\psi_{ij}(\eta) &= \psi_{ij}(\eta)\frac{\partial}{\partial \sigma_v^2}\left(\frac{1}{2}z_{ij}^2\tau_i^2\right)\\ &= \psi_{ij}(\eta)\frac{1}{2}z_{ij}^2\frac{\partial}{\partial \sigma_v^2}\left(\tau_i^2\right)\\ &= \psi_{ij}(\eta)\frac{1}{2}z_{ij}^2 b_i^T\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i\\ &= \psi_{ij}(\eta)A_5; A_5 = \frac{1}{2}z_{ij}^2 b_i^T\\ &\quad\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}b_i\end{aligned}$$

$$\frac{\partial}{\partial \sigma_e^2}\psi_{ij}(\eta) = \psi_{ij}(\eta)\frac{\partial}{\partial \sigma_e^2}\left(\frac{1}{2}\sigma_e^2\right) = \frac{1}{2}\psi_{ij}(\eta)$$

$$\frac{\partial^2}{\partial^2 \sigma_e^2}\psi_{ij}(\eta) = \frac{\partial}{\partial \sigma_e^2}\left(\frac{\partial}{\partial \sigma_e^2}\psi_{ij}(\eta)\right) = \frac{\partial}{\partial \sigma_e^2}\left(\frac{1}{2}\psi_{ij}(\eta)\right)$$

$$= \frac{1}{2}\left(\frac{\partial}{\partial \sigma_e^2}\psi_{ij}(\eta)\right) = \frac{1}{4}\psi_{ij}(\eta)$$

$$\frac{\partial^2}{\partial \sigma_v^2 \partial \sigma_e^2}\psi_{ij}(\eta) = \frac{\partial}{\partial \sigma_v^2}\left(\frac{\partial}{\partial \sigma_e^2}\psi_{ij}(\eta)\right) = \frac{\partial}{\partial \sigma_v^2}\left(\frac{1}{2}\psi_{ij}(\eta)\right)$$

$$= \frac{1}{2}\left(\frac{\partial}{\partial \sigma_v^2}\psi_{ij}(\eta)\right) = \frac{1}{2}\psi_{ij}(\eta) A_5$$

The second-order Taylor approximation of $\psi_{ij}\left(\hat{\eta}\right)$ is given by:

$$\psi_{ij}\left(\hat{\eta}\right) \approx \psi_{ij}(\eta) + \left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(1)}(\eta) + \frac{1}{2}\left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(2)}(\eta)\left(\hat{\eta} - \eta\right). \quad \text{(A1.2)}$$

By taking the expectation on the left-hand side and the right-hand side of A1.2, we obtain:

$$E\left\{\psi_{ij}\left(\hat{\eta}\right)\right\} \approx E\left[\psi_{ij}(\eta)\right] + E\left[\left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(1)}(\eta)\right] + E\left\{\frac{1}{2}\left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(2)}(\eta)\left(\hat{\eta} - \eta\right)\right\}$$

$$= \psi_{ij}(\eta) + E\left\{\frac{1}{2}\left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(2)}(\eta)\left(\hat{\eta} - \eta\right)\right\}$$

(Note: $E\left[\left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(1)}(\eta)\right] = 0$ because $\hat{\eta}$ is assumed to be unbiased estimator for η)

$$E\left\{\psi_{ij}\left(\hat{\eta}\right)\right\} = \psi_{ij}(\eta) + \frac{1}{2}E\left\{\left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(2)}(\eta)\left(\hat{\eta} - \eta\right)\right\}$$

$$= \psi_{ij}(\eta) + \frac{1}{2}tr\left\{\psi_{ij}^{(2)}(\eta) E\left[\left(\hat{\eta} - \eta\right)\left(\hat{\eta} - \eta\right)^T\right]\right\}$$

$$= \psi_{ij}(\eta) + \frac{1}{2}tr$$

$$\left\{\begin{bmatrix} C_1 & 0 & 0 & 0 \\ 0 & C_2 & 0 & 0 \\ 0 & 0 & C_3 & C_5 \\ 0 & 0 & C_5 & C_4 \end{bmatrix}\begin{bmatrix} V(\hat{\beta}) & 0 & 0 & 0 \\ 0 & V(\hat{\rho}) & 0 & 0 \\ 0 & 0 & V(\hat{\sigma}_v^2) & Cov(\hat{\sigma}_v^2, \hat{\sigma}_\varepsilon^2) \\ 0 & 0 & Cov(\hat{\sigma}_v^2, \hat{\sigma}_\varepsilon^2) & V(\hat{\sigma}_\varepsilon^2) \end{bmatrix}\right\}$$

where

$$C_1 = \frac{\partial^2}{\partial \beta^2}\psi_{ij}(\eta) = \psi_{ij}(\eta) x_{ij} x_{ij}^T$$

$$C_2 = \frac{\partial^2}{\partial \rho^2}\psi_{ij}(\eta) = \psi_{ij}(\eta)\left[A_1 A_2 A_1 A_2 + A_3 A_2 + A_1 A_4\right]$$

$$C_3 = \frac{\partial^2}{\partial^2 \sigma_v^2}\psi_{ij}(\eta) = \psi_{ij}(\eta) A_5 A_5$$

$$C_4 = \frac{\partial^2}{\partial^2 \sigma_e^2}\psi_{ij}(\eta) = \frac{1}{4}\psi_{ij}(\eta)$$

$$C_5 = \frac{\partial^2}{\partial \sigma_v^2 \partial \sigma_e^2} \psi_{ij}(\eta) = \frac{1}{2} \psi_{ij}(\eta) A_5$$

Alternatively, $E\left\{\psi_{ij}\left(\hat{\eta}\right)\right\}$ also can be written by:

$$E\left\{\psi_{ij}\left(\hat{\eta}\right)\right\} = \psi_{ij}(\eta) + \frac{1}{2} tr$$
$$\left\{\psi_{ij}(\eta) \begin{bmatrix} D_1 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 \\ 0 & 0 & D_3 & D_5 \\ 0 & 0 & D_5 & D_4 \end{bmatrix} \begin{bmatrix} V(\hat{\beta}) & 0 & 0 & 0 \\ 0 & V(\hat{\rho}) & 0 & 0 \\ 0 & 0 & V(\hat{\sigma}_v^2) & Cov(\hat{\sigma}_v^2, \hat{\sigma}_\varepsilon^2) \\ 0 & 0 & Cov(\hat{\sigma}_v^2, \hat{\sigma}_\varepsilon^2) & V(\hat{\sigma}_\varepsilon^2) \end{bmatrix}\right\}$$

where

$$\begin{aligned} D_1 &= x_{ij} x_{ij}^T \\ D_2 &= [A_1 A_2 A_1 A_2 + A_3 A_2 + A_1 A_4] \\ D_3 &= A_5 A_5 \\ D_4 &= \frac{1}{4} \\ D_5 &= \frac{1}{4} A_5 \end{aligned}$$

Then,

$$\begin{aligned} E\left\{\psi_{ij}\left(\hat{\eta}\right)\right\} &\approx \psi_{ij}(\eta) + E\left\{\frac{1}{2}\left(\hat{\eta} - \eta\right)^T \psi_{ij}^{(2)}(\eta)\left(\hat{\eta} - \eta\right)\right\} \\ &= \psi_{ij}(\eta) + \frac{1}{2} tr\left\{\psi_{ij}^{(2)}(\eta)\, E\left[\left(\hat{\eta} - \eta\right)^T\left(\hat{\eta} - \eta\right)\right]\right\} \\ &= \psi_{ij}(\eta) + \frac{1}{2}\left\{\psi_{ij}(\eta)\left[D_1 V\left(\hat{\beta}\right) + D_2 V\left(\hat{\rho}\right) + D_3 V\left(\hat{\sigma}_v^2\right) + D_4 V\left(\hat{\sigma}_e^2\right)\right]\right\} \\ &= \psi_{ij}(\eta)\left\{1 + \frac{1}{2}\left[D_1 V\left(\hat{\beta}\right) + D_2 V\left(\hat{\rho}\right) + D_3 V\left(\hat{\sigma}_v^2\right) + D_4 V\left(\hat{\sigma}_e^2\right)\right]\right\} \end{aligned} \quad \text{(A1.3)}$$

Based on the result (A1.3), it can be seen that $\psi_{ij}\left(\hat{\eta}\right)$ is not unbiased estimate of $\psi_{ij}(\eta)$ .. The bias correction factor for $\hat{y}_{ij}^{\mathrm{SEBP}}$ is given by:

$$c_{ij}^{\mathrm{SEBP}} = 1 + \frac{1}{2}\left[D_1 V\left(\hat{\beta}\right) + \mathrm{D}_2 V\left(\hat{\rho}\right) + D_3 V\left(\hat{\sigma}_v^2\right) + D_4 V\left(\hat{\sigma}_e^2\right)\right] \quad \text{(A1.4)}$$

As a result, the spatial empirical Bayes predictor (SEBP) of μ_i with bias correction is given by:

$$\hat{\mu}_i^{\mathrm{SEBP\_}c} = \frac{1}{N_i}\left[\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij}^{\mathrm{SEBP\_}c}\right] \quad \text{(A1.5)}$$

where $\hat{y}_{ij}^{\mathrm{SEBP\_}c} = \left(c_{ij}^{\mathrm{SEBP}}\right)^{-1} \hat{y}_{ij}^{\mathrm{SEBP}} = \left(c_{ij}^{\mathrm{SEBP}}\right)^{-1} e^{x_{ij}^T\hat{\beta}+\frac{1}{2}\left(z_{ij}^2\hat{\tau}_i^2+\hat{\sigma}_e^2\right)}$

Appendix 2: The MSE of $\hat{\mu}_i^{SEBP\_c}$

The MSE of $\hat{\mu}_i^{SEBP\_c}$ is given by:

$$\begin{aligned}
\mathrm{MSE}(\hat{\mu}_i^{\mathrm{SEBP_}c}) &= E\left(\hat{\mu}_i^{\mathrm{SEBP_}c} - \mu_i\right)^2 \\
&= E\left(\hat{\mu}_i^{\mathrm{SEBP_}c} - \hat{\mu}_i^{\mathrm{SBP_}c} + \hat{\mu}_i^{\mathrm{SBP_}c} - \mu_i\right)^2 \\
&= E\left[\left(\hat{\mu}_i^{\mathrm{SEBP_}c} - \hat{\mu}_i^{\mathrm{SBP_}c}) + (\hat{\mu}_i^{\mathrm{SBP_}c} - \mu_i\right)\right]^2 \\
&= E\left(\hat{\mu}_i^{\mathrm{SBP_}c} - \mu_i\right)^2 + E\left(\hat{\mu}_i^{\mathrm{SEBP_}c} - \hat{\mu}_i^{\mathrm{SBP_}c}\right)^2 \\
&\quad + 2E\left(\hat{\mu}_i^{\mathrm{SEBP_}c} - \hat{\mu}_i^{\mathrm{SBP_}c}\right)\left(\hat{\mu}_i^{\mathrm{SBP_}c} - \mu_i\right)
\end{aligned} \tag{A2.1}$$

where

$$\hat{\mu}_i^{\mathrm{SEBP\_}c} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\mathrm{SEBP\_}c}\right]; \hat{y}_{ij}^{\mathrm{SEBP\_}c} = \left(c_{ij}^{\mathrm{SEBP}}\right)^{-1} e^{x_{ij}^T\hat{\beta}+\frac{1}{2}\left(z_{ij}^2\hat{\tau}_i^2+\hat{\sigma}_e^2\right)}$$

$$\hat{\mu}_i^{\mathrm{SBP\_}c} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\mathrm{SBP\_}c}\right]; \hat{y}_{ij}^{\mathrm{SBP\_}c} = \left(c_{ij}^{\mathrm{SBP}}\right)^{-1} e^{x_{ij}^T\hat{\beta}+\frac{1}{2}\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)}$$

Note: c_{ij}^{SEBP} has been derived in (A1.4) whereas c_{ij}^{SBP} is similar to c_{ij}^{SEBP} but $\eta = \left(\rho, \sigma_v^2, \sigma_e^2\right)$ is assumed to be known.

The first term of (A2.1), $E\left(\hat{\mu}_i^{\mathrm{SBP\_}c} - \mu_i\right)^2$, can be written as:

$$\begin{aligned}
E\left(\hat{\mu}_i^{\mathrm{SBP_}c} - \mu_i\right)^2 &= \mathrm{Var}\left(\hat{\mu}_i^{\mathrm{SBP_}c}\right) \\
&= \mathrm{Var}\left\{N_i^{-1}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{\mathrm{SBP_}c}\right]\right\}
\end{aligned} \tag{A2.2}$$

Note: the first part, $\sum_{j\in s_i} y_{ij}$, which is calculated using sample data values, is uncorrelated with the second part $\sum_{j\in r_i} \hat{y}_{ij}^{SBP\_c}$ that is based on non-sampled values. Therefore,

$$
\begin{aligned}
E\left(\hat{\mu}_i^{\mathrm{SBP\_}c} - \mu_i\right)^2 &= \operatorname{Var}\left\{N_i^{-1}\sum_{j\in s_i} y_{ij}\right\} + \operatorname{Var}\left\{N_i^{-1}\sum_{j\in r_i}\hat{y}_{ij}^{\mathrm{SBP\_}c}\right\} \\
&= \operatorname{Var}\left\{N_i^{-1}\sum_{j\in s_i} y_{ij}\right\} + N_i^{-2}\left\{\operatorname{Var}\sum_{j\in r_i}\hat{y}_{ij}^{\mathrm{SBP\_}c}\right\}
\end{aligned}
\tag{A2.3}
$$

The first term of (A2.3) $\operatorname{Var}\left\{N_i^{-1}\sum_{j\in s_i} y_{ij}\right\}$ would be obtained using sample data values.

The moment generating function for $Y^* = \log Y$ which follows normal distribution with mean μ and variance σ^2 is given by:

$$
M(t) = E\left(e^{tY^*}\right) = e^{\mu t + \frac{1}{2}t^2\sigma^2}
$$

or

$$
M(t) = E\left(e^{tY^*}\right) = e^{t x_{ij}^T \beta + \frac{1}{2}t^2\left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)}
$$

$E\left(y_{ij}^* y_{ik}^*\right) = E\left[\left(x_{ij}^T\beta + z_{ij}u_i\right)\left(x_{ik}^T\beta + z_{ik}u_i\right)\right]$, (note: It is the second moment of lognormal distribution)

$$
\begin{aligned}
&E\left(y_{ij}^*\right) = e^{x_{ij}^T\beta + \frac{1}{2}\left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)} \\
&E\left(y_{ik}^*\right) = e^{x_{ik}^T\beta + \frac{1}{2}\left(z_{ik}^2\tau_i^2 + \sigma_e^2\right)} \\
&\operatorname{cov}\left(y_{ij}^*, y_{ik}^*\right) = E\left(y_{ij}^* y_{ik}^*\right) - E\left(y_{ij}^*\right)E\left(y_{ik}^*\right) \\
&\quad = E\left[\left(x_{ij}^T\beta + z_{ij}u_i\right)\left(x_{ik}^T\beta + z_{ik}u_i\right)\right] - e^{x_{ij}^T\beta + \frac{1}{2}\left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)} e^{x_{ik}^T\beta + \frac{1}{2}\left(z_{ik}^2\tau_i^2 + \sigma_e^2\right)} \\
&E\left(y_{ij}^{*2}\right) = e^{2x_{ij}^T\beta + 2\left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)} \\
&\operatorname{var}\left(y_{ij}^*\right) = E\left(y_{ij}^{*2}\right) - \left[E\left(y_{ij}^*\right)\right]^2 = e^{2x_{ij}^T\beta + 2\left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)} - e^{2x_{ij}^T\beta + \left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)}
\end{aligned}
$$

The second term of (A2.3) is given by:

$$
\begin{aligned}
N_i^{-2}\left\{\operatorname{Var}\sum_{j\in r_i}\hat{y}_{ij}^{SBP\_c}\right\} &= N_i^{-2}\left\{\sum_{j\in r_i}\operatorname{Var}\left(\hat{y}_{ij}^{SBP\_c}\right) + 2\sum_{j<k\in r_i}\sum\operatorname{cov}\left(\hat{y}_{ij}^{SBP\_c}, \hat{y}_{ik}^{SBP\_c}\right)\right\} \\
&= N_i^{-2}\left\{\sum_{j\in r_i}\operatorname{Var}\left(\left(c_{ij}^{SBP}\right)^{-1}\hat{y}_{ij}^{SBP}\right) + 2\sum_{j<k\in r_i}\sum\operatorname{cov}\left(\left(c_{ij}^{SBP}\right)^{-1}\hat{y}_{ij}^{SBP}, \left(c_{ij}^{SBP}\right)^{-1}\hat{y}_{ik}^{SBP}\right)\right\} \\
&= N_i^{-2}\left\{\left(c_{ij}^{SBP}\right)^{-2}\sum_{j=1}^{(N_i-n_i)} e^{2x_{ij}^T\hat{\beta} + \left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)}\left[e^{\left(z_{ij}^2\tau_i^2 + \sigma_e^2\right)} - 1\right]\right. \\
&\quad \left. + 2\left(c_{ij}^{SBP}\right)^{-2}\sum_{1\le j<k\le(N_i-n_i)}\operatorname{cov}\left(\hat{y}_{ij}^{SBP}, \hat{y}_{ik}^{SBP}\right)\right\}, \; i = 1, 2 \ldots M
\end{aligned}
$$

$$= N_i^{-2}\left\{\left(c_{ij}^{SBP}\right)^{-2}\sum_{j=1}^{(N_i-n_i)} e^{2x_{ij}^T\beta+\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)}\left[e^{\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)}-1\right]\right.$$

$$\left.+2\left(c_{ij}^{SBP}\right)^{-2}\sum_{1\le j<k\le(N_i-n_i)} E\left[\left(x_{ij}^T\hat{\beta}+z_{ij}u_i\right)\left(x_{ik}^T\hat{\beta}+z_{ik}u_i\right)\right]-e^{x_{ij}^T\hat{\beta}+\frac{1}{2}\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)}e^{x_{ik}^T\hat{\beta}+\frac{1}{2}\left(z_{ik}^2\tau_i^2+\sigma_e^2\right)}\right\},$$

Note that $\hat{\mu}_i^{\text{SBP\_}c}$ depends on σ_v^2, σ_e^2 and ρ which are usually unknown. The $\hat{\mu}_i^{\text{SEBP\_}c}$ is obtained from $\hat{\mu}_i^{\text{SBP\_}c}$ by replacing σ_v^2, σ_e^2 and ρ by their estimates. The cross-product term in (A2.1) $E\left(\hat{\mu}_i^{\text{SEBP\_}c}-\hat{\mu}_i^{\text{SBP\_}c}\right)\left(\hat{\mu}_i^{\text{SBP\_}c}-\mu_i\right)$ is zero. See Rao and Molina (2015) for the proof that $E\left(\hat{\mu}_i^{\text{SBP\_}c}-\hat{\mu}_i^{\text{SBP\_}c}\right)\left(\hat{\mu}_i^{\text{SBP\_}c}-\hat{\mu}_i^{\text{SBP\_}c}\mu_i\right)=0$.

Using a Taylor approximation, $\hat{\mu}_i^{\text{SEBP\_}c}$ can be written by:

$$\begin{aligned}\hat{\mu}_i^{\text{SEBP\_}c}-\hat{\mu}_i^{\text{SBP}_c} &= (\hat{\eta}-\eta)\frac{\partial}{\partial\eta}\left[\hat{\mu}_i^{\text{SBP}_c}\right];\quad \eta\\ &= \left(\beta,\rho,\sigma_v^2,\sigma_e^2\right)\quad\text{and}\quad\hat{\eta}\\ &= \left(\hat{\beta},\hat{\rho},\hat{\sigma}_v^2,\hat{\sigma}_e^2\right)\end{aligned}\tag{A2.4}$$

By squaring on the left-hand side and the right-hand side of A2.4, and then taking the expectation on the both side, we will obtain:

$$\begin{aligned}&E\left(\hat{\mu}_i^{\text{SEBP_}c}-\hat{\mu}_i^{\text{SBP_}c}\right)^2\\
&\approx E\left\{(\hat{\eta}-\eta)\frac{\partial}{\partial\eta}\left[\hat{\mu}_i^{\text{SBP_}c}\right]\right\}^2\\
&= E\left\{(\hat{\eta}-\eta)\frac{\partial}{\partial\eta}\left(N_i^{-1}\sum_{j\in s_i}y_{ij}+N_i^{-1}\sum_{j\in r_i}\hat{y}_{ij}^{\text{SBP_}c}\right)\right\}^2\\
&= E\left\{(\hat{\eta}-\eta)\left\{\frac{\partial}{\partial\eta}\left(N_i^{-1}\sum_{j\in r_i}\hat{y}_{ij}^{\text{SBP_}c}\right)\right\}\right\}^2,\quad \text{Note:}\frac{\partial}{\partial\eta}\left(N_i^{-1}\sum_{j\in s_i}y_{ij}\right)=0\\
&= E\left\{(\hat{\eta}-\eta)\left\{N_i^{-1}\left(\sum_{j\in r_i}\frac{\partial}{\partial\eta}\hat{y}_{ij}^{\text{SBP_}c}\right)\right\}\right\}\left\{(\hat{\eta}-\eta)\left\{N_i^{-1}\left(\sum_{j\in r_i}\frac{\partial}{\partial\eta}\hat{y}_{ij}^{\text{SBP_}c}\right)\right\}\right\}\\
&= E\left\{(\hat{\eta}-\eta)(\hat{\eta}-\eta)^T\left\{N_i^{-1}\left(\sum_{j\in r_i}\frac{\partial}{\partial\eta}\hat{y}_{ij}^{\text{SBP_}c}\right)\right\}\left\{N_i^{-1}\left(\sum_{j\in r_i}\frac{\partial}{\partial\eta}\hat{y}_{ij}^{\text{SBP_}c}\right)\right\}^T\right\}\\
&= tr\left\{\begin{bmatrix}V(\hat{\rho}) & 0 & 0\\ 0 & V\left(\hat{\sigma}_v^2\right) & \text{Cov}\left(\hat{\sigma}_v^2,\hat{\sigma}_e^2\right)\\ 0 & \text{Cov}\left(\hat{\sigma}_v^2,\hat{\sigma}_e^2\right) & V\left(\hat{\sigma}_e^2\right)\end{bmatrix}\begin{bmatrix}N_i^{-1}A\\ N_i^{-1}B\\ N_i^{-1}C\end{bmatrix}\begin{bmatrix}N_i^{-1}A\\ N_i^{-1}B\\ N_i^{-1}C\end{bmatrix}^T\right\}\end{aligned}$$

$$\text{Note: } \frac{\partial}{\partial\eta}\hat{y}_{ij}^{\text{SBP\_}c} = \begin{bmatrix} \frac{\partial}{\partial\rho}\hat{y}_{ij}^{\text{SBP\_}c} \\ \frac{\partial}{\partial\sigma_v^2}\hat{y}_{ij}^{\text{SBP\_}c} \\ \frac{\partial}{\partial\sigma_e^2}\hat{y}_{ij}^{\text{SBP\_}c} \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

$$\begin{aligned} E\left(\hat{\mu}_i^{\text{SEBP\_}c} - \hat{\mu}_i^{\text{SBP\_}c}\right)^2 &\approx N_i^{-2} tr \left\{ \begin{bmatrix} V\left(\hat{\rho}\right) & 0 & 0 \\ 0 & V\left(\hat{\sigma}_v^2\right) & \text{Cov}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) \\ 0 & \text{Cov}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) & V\left(\hat{\sigma}_e^2\right) \end{bmatrix} \begin{bmatrix} A^2 & AB & AC \\ AB & B^2 & BC \\ AC & BC & C^2 \end{bmatrix} \right\} \\ &\approx N_i^{-2}\left\{A^2 V\left(\hat{\rho}\right) + B^2 V\left(\hat{\sigma}_v^2\right) + 2BCCov\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) + C^2 V\left(\hat{\sigma}_e^2\right)\right\} \\ &\approx N_i^{-2}\left\{A^2 V\left(\hat{\rho}\right) + \nabla\left(B^2\hat{\sigma}_v^2 + C\hat{\sigma}_e^2\right)\right\} \end{aligned}$$

$$\begin{aligned} A &= \frac{\partial}{\partial\rho}\sum_{j\in r_i}\hat{y}_{ij}^{\text{SBP\_}c} = \sum_{j\in r_i}\frac{\partial}{\partial\rho}\hat{y}_{ij}^{\text{SBP\_}c} \\ &= \sum_{j\in r_i}\left[\hat{y}_{ij}^{\text{SBP\_}c}\frac{1}{2}z_{ij}^2\left[b_i^T\sigma_v^2\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1}\right.\right. \\ &\quad \left.\left.\times\left[W^T + W - 2\rho W^T W\right]\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1} b_i\right]\right] \\ B &= \frac{\partial}{\partial\sigma_v^2}\sum_{j\in r_i}\hat{y}_{ij}^{\text{SBP\_}c} = \sum_{j\in r_i}\frac{\partial}{\partial\sigma_v^2}\hat{y}_{ij}^{\text{SBP\_}c} = \sum_{j\in r_i}\hat{y}_{ij}^{\text{SBP\_}c}\frac{1}{2}z_{ij}^2\left(b_i^T\left\{\left(I-\rho W^T\right)(I-\rho W)\right\}^{-1} b_i\right) \\ C &= \frac{\partial}{\partial\sigma_e^2}\sum_{j\in r_i}\hat{y}_{ij}^{\text{SBP\_}c} = \sum_{j\in r_i}\frac{\partial}{\partial\sigma_e^2}\hat{y}_{ij}^{\text{SBP\_}c} = \sum_{j\in r_i}\frac{1}{2}\hat{y}_{ij}^{\text{SBP\_}c} \end{aligned}$$

Finally, the MSE of $\hat{\mu}_i^{SEBP\_c}$ is:

$$\begin{aligned} &\text{MSE}\left(\hat{\mu}_i^{\text{SEBP\_}c}\right) \\ &\approx \text{Var}\left\{N_i^{-1}\sum_{j\in s_i} y_{ij}\right\} + N_i^{-2}\left\{\sum_{j=1}^{(N_i-n_i)} e^{2x_{ij}^T\beta+\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)}\left[e^{\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)} - 1\right]\right. \\ &\quad \left. + 2\sum_{1\le j<k\le(N_i-n_i)} E\left[\left(x_{ij}^T\beta + z_{ij}u_i\right)\left(x_{ik}^T\beta + z_{ik}u_i\right)\right] - e^{x_{ij}^T\beta+\frac{1}{2}\left(z_{ij}^2\tau_i^2+\sigma_e^2\right)} e^{x_{ik}^T\beta+\frac{1}{2}\left(z_{ik}^2\tau_i^2+\sigma_e^2\right)}\right\} \\ &\quad + N_i^{-2}\left\{A^2 V\left(\hat{\rho}\right) + \nabla\left(B^2\hat{\sigma}_v^2 + C\hat{\sigma}_e^2\right)\right\}; \end{aligned}$$

where ∇ (.) is asymptotic variance–covariance matrix of total variability of response.

References

Asfar, Kurnia A, Sadik K (2016) Optimum spatial weighted in small area estimation. Glob J Pure Appl Math 12(5):3977–3989

Bellow ME, Lahiri PS (2011) An empirical best linear unbiased prediction approach to small area estimation of crop parameters. In *Section on survey research methods*, pp 3976–3986

Berg E, Chandra H (2014) Small area prediction for a unit-level lognormal. Comput Stat Data Anal 78:159–175

Chandra H, Chambers R (2011) Small area estimation under transformation to linearity. Surv Methodol 37:39–51

Jiang J (1996) REML estimation: asymptotic behaviour and related topics. Ann Stat 24:256–286
Karlberg F (2000) Population total prediction under a lognormal superpopulation model. Metron, LVIII, pp 53–80
Kurnia A, Chambers R (2011) Small area inference for positively skewed distributions. In: The proceeding of the 6-th SEAMS-GMU international conference on mathematics and its applications, July 12–15, 2011, Yogyakarta
McCulloch CE, Searle SR (2001) Generalized, linear and mixed models. Wiley, New York
Molina I, Salvati N, Pratesi M (2009) Bootstrap for estimating the MSE of the spatial EBLUP. Comput Stat 24:441–458
Petrucci A, Salvati N (2004a) Small area estimation using spatial information, The Rathbun lake watershed case study. Working Paper no 2004/02, "G. Parenti" Department of Statistics, University of Florence
Petrucci A, Salvati N (2004b) Small area estimation considering spatially correlated errors: the unit level random effects model. Working Paper no 2004/10, "G. Parenti" Department of Statistics, University of Florence
Petrucci A, Salvati N (2006) Small area estimation for spatial correlation in watershed erosion assesment. J Agric Biol Environ Stat 11:169–182
Prasad NGN, Rao JNK (1990) The estimation of mean squared errors of small area estimators. J Am Stat Assoc 85:163–171
Pratesi M, Salvati N (2008) Small area estimation: the EBLUP estimator based on spatially correlated random effects. Stat. Methods Appl 17:113–141
Rao JNK, Molina I (2015) Small area estimation. Wiley, New York
Salvati N (2004) Small area estimation by spatial models: the spatial empirical best linear unbiased prediction (spatial EBLUP). Working Paper no 2004/03, "G. Parenti" Department of Statistics, University of Florence
Slud EV, Maiti T (2006) Mean-squared estimation in transformed Fay–Herriot models. J R Stat Soc B 68:67–72
Wang J, Fuller WA (2003) The mean squared error of small area predictor constructed with estimated area variances. J Am Stat Assoc 98:716–745