Estimation of Mean Population in Small Area with Spatial Best Linear Unbiased Prediction Method

Syahril Ramadhan¹, Titin Siswantining, and Saskya Mary Soemartojo

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia (UI)
Kampus Baru UI Depok West Java –16424, INDONESIA

E-mail: syahril.ramadhan@sci.ui.ac.id, titin@sci.ui.ac.id, saskya@sci.ui.ac.id

Abstract. Survey is aimed at estimating population parameters such as the total as well as the mean of an area with a large sample size. One approach in estimating population parameters are obtained through direct estimation methods. However, there is a problem when the direct estimation is used for a small area, which will cause a large standard error. This problem was addressed by developing a method of parameter estimation known as the Small Area Estimation (SAE) method. In this paper, we will describe the procedure to find the mean population estimate in a small area using Spatial Best Linear Unbiased Prediction (Spatial BLUP) method that follows Simultaneously Autoregressive (SAR) model. In general, this procedure begins with the definition of an area-level model. Then, the area-level model is expanded by the addition of spatial effect into the random effects of the area. In the end, the spatial model of the area level is used as the basis for estimating the mean population in small areas.

1. Introduction

Surveys are an important part of the data. Survey implementation is aimed at estimating population parameters. The results of this survey can be used to obtain a reliable estimate of the total as well as the mean population value of an object with a large sample size. The object used as a population is called a domain (area).

One approach to estimating population value parameters is based on the application of design-based models such as simple random sampling, stratified random sampling, etc. where the estimators generated through this approach are called direct estimation. Direct estimation is a method of estimation in an area based on information of sample data from that area. However, there are problems where such direct estimates are used for an area with a small sample size, which will lead to a large standard error (Ghosh & Rao, 1994). An area is regarded as “small” if the area-specific sample is not large enough to support direct estimates of adequate precision (Rao, 2003). Or in other words, a small area can be expressed as an area with a small sample size. We use the term small area to denote any area for which direct estimates of adequate precision cannot be produced.

In addition to causing a large standard error, Rao (2003) also adds that direct estimation methods cannot be performed on non-selected areas as samples, in the absence of data that can be used to make the estimation. These problems were addressed by developing a method of parameter estimation known as the Small Area Estimation (SAE) method. According to Rao (2003), the SAE method can be defined as a method for estimating population parameters in a relatively small area of the survey.
sampling by utilizing information from outside the area in the form of other information from a similar area, from within the area itself in the form of a previous survey in the same area, and from outside the survey in the form of supporting variables related to the variable to be expected. Chand & Alexander (1995) also mentioned that the SAE procedure essentially harnesses the power of the surrounding area information and data sources outside of the area whose statistical value is to be obtained through the establishment of appropriate models to improve the effectiveness of the sample size. This statement means that using a small sample size can get a good estimator. The estimation of the parameters and their inferences by using such additional information is called indirect estimation. Indirect estimation is a method of estimating population parameters in an area by involving additional information from within the area of concern, other areas, or other surveys (Rao, 2003).

There are two basic assumptions in the development of a model for the estimation of small areas, namely that the variability within the observed small area of the variable can be explained by the corresponding variability relationships in the supplemental information called fixed effects. Another assumption is that the small area-specific variability cannot be explained by additional information and is a small area random effect. The combined two assumptions form a linear mixed model.

Fay & Herriot (1979) is the first researcher to develop SAE based on a mixed linear model. This model came to be known as the Fay-Herriot model and became the basis in SAE which assumed that the random effects of the areas are independent. But in some cases, these assumptions are often violated. General, the variance of an area is affected by the surrounding area so that spatial effect can be incorporated into random effects. Spatial effect is common between one area and another, or it can be said that one area affects the other. Moreover, there is no reason to exclude the assumption that the random effect between the neighboring areas are correlated, and that correlation decays to zero as distance increases. Based on this, spatial information can be used in the SAE model. By incorporating the effect of spatial dependence into the random effects of the area, SAE will be obtained by considering the effect of spatial dependence across areas (Asfar, 2016).

One method in SAE is the best linear unbiased prediction method (BLUP). The BLUP estimator taking into account the random effects of spatially dependent areas is known as the spatial best linear unbiased prediction (Spatial BLUP). Researchers have previously mentioned that the use of spatial information can minimize variance.

The paper is organized as it follows. The spatial model of the area level is described in Section 2. In Section 2 discussed: small area model, area-level model, Fay-Herriot model, SAR model, and the spatial model of the area level. Section 3 discusses the BLUP method, and Section 4 shows the Spatial BLUP estimator. The conclusion can be found in Section 5.

2. The spatial model of the area level
In this section, we will first discuss the understanding of the area-level model that is the basis of the models to be discussed in this paper. Furthermore, the area-level model is expanded with the addition of spatial effects to the random effects of the area. Finally, the mean population parameter in a small area is presumably using the Spatial Best Linear Unbiased Prediction (Spatial BLUP) method.

2.1. Small area model
Small-area estimation models are classified into two types: the area-level model and the unit-level model (Rao, 2003). Area-level model is used if the auxiliary data corresponding to observed variable data is not available at the unit level of observation, whereas unit-level model is used if the auxiliary data corresponding to observed variable data is available up to the unit level of observation.

2.1.1. Area-level model. Area-level model is models based on the availability of the auxiliary data that exist only for a given area level. The auxiliary variable is a collection of variables that contain information from each unit of population and not a variable of concern. If the auxiliary variables are included in the model, it can increase the information in predicting the parameters in the SAE (Small Area Estimation) method, so that the resulting estimates will be better. A critical assumption for area-
level models, we assume the absence of informative sampling of the areas in situations where only some of the areas are selected to the sample, that is, the sample area value (the direct estimates) obey the assumed population model (Rao, 2003).

Suppose a population of U is divided into the M area, where there is only a m ≤ M area that allows surveying and obtains an auxiliary variables \( x_i = (x_{i1}, x_{i2}, ..., x_{ip})^T \) which is the auxiliary variables of the i-th area. If the parameter to be expected \( \theta_i \) (\( \theta_i \) represents the parameter of the i-th area) to know its value, then the area-level model assumes that \( \theta_i \) will be interconnected with the auxiliary variables \( x_i \) through the following model:

\[
\theta_i = x_i^T \beta + b_i v_i, \quad i = 1, 2, ..., m
\]

Where \( \beta \) is the \( p \times 1 \) vector of regression coefficients, \( b_i \) is the weight of the random effects for the i-th area, \( m \) is the number of small areas, and \( v_i \) is the random effects of the area in the i-th area with assumption \( v_i \sim i.i.d N(0, \sigma_v^2) \).

From now on, the model of equation (1) is referred to as the **linking model** between parameter \( \theta_i \) and the auxiliary variables \( x_i \). Availability of good auxiliary data and determination of suitable linking models are crucial to the formation of indirect estimators.

The assumption in making the conclusions about the population based on the model in equation (1) is the availability of direct estimator \( \hat{\theta}_i \) and can be written as follows:

\[
\hat{\theta}_i = \theta_i + e_i, \quad i = 1, 2, ..., m
\]

Where \( e_i \) is the independent sampling error in the i-th area with assumption \( e_i | \theta_i \sim \text{i.i.d } N(0, \psi_i) \). From now on, the model of equation (2) is referred to as the **sampling model** showing the relationship between direct observation of \( \theta_i \) and parameter \( \hat{\theta}_i \).

Equations (1) and (2) can be combined so that the model is obtained:

\[
\hat{\theta}_i = x_i^T \beta + b_i v_i + e_i, \quad i = 1, 2, ..., m
\]

With the additional assumption that \( v_i \) and \( e_i \) are independent. From now on, the model of equation (3) is referred to as the **area-level model**.

### 2.2 Fay-Herriot Model

Research on SAE was first performed by Fay and Herriot (1979) i.e., by estimating the per capita income of a small area based on US Census Bureau survey data. This model is then known as the Fay-Herriot model, where in this model the weight of random effect between areas \( b_i \) is assumed to have the same effect, that is value \( b_i = 1 \) in equation (3).

Similar to the area-level model in SAE, the Fay-Herriot model consists of two components, the linking model (1) and the sampling model (2) where the sampling error variance \( \psi_i \) in the sampling model is assumed to be known. Through this definition, the Fay-Herriot model is defined as follows:

\[
\hat{\theta}_i = x_i^T \beta + v_i + e_i, \quad i = 1, 2, ..., m
\]

Where assumptions on the linking model and sampling model are maintained in the Fay-Herriot model. In this paper, the model in equation (4) is used as the basis of the area-level model.

The effect of random area \( v_i \) on equation (4) is still assumed to be independent between areas. In fact, the spatial relationship between small areas becomes an interesting thing to observe. Therefore, in this paper will be discussed further about the expansion of \( v_i \) as a random effect area that is no longer independent but has a spatial dependence.

### 2.3. SAR Model

In fact, it is reasonable to say that there is a dependency between the adjacent area and the dependence will decrease as the distance between the areas increases. Inter-area spatial dependence needs to be explained through a model known as spatial modeling. In this paper, the spatial model that focuses on this paper is the **simultaneous autoregressive model** (SAR model).
The SAR model was introduced by Anselin in 1992 where the vector of random effects of area \( \mathbf{v} \) is defined as follows:

\[
\mathbf{v} = (\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{u}
\]

Where \( \mathbf{v} \) is the \( m \times 1 \) vector of dependent random variables according to an SAR error process with mean \( \mathbf{0} \) and variance-covariance matrix \( \mathbf{G} = \sigma_u^2 (\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^T)^{-1} \). In this matrix, \( \rho \) is the spatial autoregressive coefficient, \( \mathbf{W} \) is the \( m \times m \) proximity matrix, \( \mathbf{I} \) is a \( m \times m \) identity matrix, and \( \mathbf{u} \) is the \( m \times 1 \) vector of independent error term with zero mean and constant variance \( \sigma_u^2 \).

A variance-covariance matrix \( \mathbf{G} \) can be derived directly with some matrix manipulation:

\[
\mathbf{G} = E[\mathbf{v}\mathbf{v}^T] = E[(\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{u}\mathbf{u}^T((\mathbf{I} - \rho \mathbf{W})^{-1})^T]
\]

\[
= (\mathbf{I} - \rho \mathbf{W})^{-1}E[\mathbf{u}\mathbf{u}^T](\mathbf{I} - \rho \mathbf{W})^{-1}
\]

\[
= (\mathbf{I} - \rho \mathbf{W})^{-1}\sigma_u^2(\mathbf{I} - \rho \mathbf{W})^{-1}
\]

\[
= \sigma_u^2((\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^T))^{-1}
\]

The spatial autoregressive coefficient \( \rho \) in equation (6) defines the strength of the spatial relationship among the random effects associated with neighboring areas. The value of \( \rho \) ranges from -1 to 1. The value of \( \rho > 0 \) indicates that an area with a high parameter value tends to be surrounded by another area with a high parameter value and vice versa. On the other hand, the value of \( \rho < 0 \) indicates that an area with a high parameter value is surrounded by another area with a low parameter value or vice versa.

The proximity matrix \( \mathbf{W} \) is an significant component in spatial data modeling in which the data contain spatial dependence. This matrix declares neighboring structures of small areas (such as correlation, distance, threshold, etc.). A general proximity matrix spatial weight matrix can be defined by a symmetric binary contiguity matrix, which can be generated from topological information provided by the geographical information system (GIS) based on adjacency criteria: the element of the proximity matrix \( w_{ij} \) is one if area \( i \) is adjacent to area \( j \), and zero otherwise.

### 2.4. The spatial model of the area level

This section will explain the extension of the area-level model with the addition of spatial random effects.

The SAR model (5) is associated with the parameter you want to know \( \theta_i \) through the linking model, in the form of a matrix can be written as follows:

\[
\theta = \mathbf{X}\beta + \mathbf{Z}(\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{u}
\]

Where \( \mathbf{X} \) is the \( m \times p \) matrix of the area specific auxiliary covariates \( \mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})^T \), \( \beta \) is the \( p \times 1 \) vector of regression coefficients, \( \mathbf{Z} \) is a \( m \times m \) matrix of known positive constant, and \( \mathbf{v} = (\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{u} \) is the \( m \times 1 \) vector of dependent random variables according to an SAR error process with mean \( \mathbf{0} \) and variance-covariance matrix \( \mathbf{G} = \sigma_u^2(\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^T))^{-1} \). Moreover, it assumes that direct estimators \( \hat{\theta}_i \) are available and follow the following sampling model:

\[
\hat{\theta} = \theta + \mathbf{e}
\]

Where \( \mathbf{e} \) represents the \( m \times 1 \) matrix of independent sampling errors with mean \( \mathbf{0} \) and known \( m \times m \) variance matrix \( \text{diag}(\psi_i) \). Combining (7) and (8), with \( \mathbf{e} \) independent of \( \mathbf{v} \), the resulting model is:

\[
\hat{\theta} = \mathbf{X}\beta + \mathbf{Z}(\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{u} + \mathbf{e}
\]

Where \( m \times m \) variance-covariance matrix for \( \mathbf{v} \) and \( \mathbf{e} \), respectively, as follows:

\[
\mathbf{G} = \sigma_u^2((\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^T))^{-1}
\]
and

\[ R = \text{diag}(\psi_i). \]  

Then, the covariance matrix of studied variable is:

\[ V = R + ZGZ^T = \text{diag}(\psi_i) + Z\sigma^2_i [(1 - \rho W)(I - \rho W^T)]^{-1} Z^T. \]  

The model in equation (9) is then later used in this paper as the spatial Fay-Herriot model.

3. Best linear unbiased prediction (BLUP) method

In the previous section, we have introduced a small area model that can be viewed as a special case of a general mixed linear model involving both fixed and random effects. In fact, small area parameters can be written as linear combinations of fixed effects and random effects. One of the interesting properties of a mixed linear model is its ability to estimate linear combinations of fixed effects and random effects. One method for estimating parameters based on a linear mixed model has been developed by Henderson (1975), the Best Linear Unbiased Prediction (BLUP) method.

The BLUP estimator minimizes the Mean Square Error (MSE) among the class of linear unbiased estimators and do not depend on the normality of the random effects. But, they depend on the variances (and covariances) of random effects which can be estimated. Maximum likelihood method can be used to estimate the variance and covariance components, assuming normality. Using these estimated components in the BLUP estimator, we obtain a two-stage estimator which is referred to as empirical BLUP or EBLUP estimator (Harville, 1991).

The BLUP context is the following mixed linear model:

\[ \tilde{\theta} = X\beta + Zv + e \]  

where \( \tilde{\theta} \) is a vector of \( n \) observable random variables, \( \beta \) is the vector of \( p \) unknown parameters having fixed values (fixed effects), \( X \) and \( Z \) are known matrices, and \( v \) and \( e \) are vectors of \( q \) and \( n \), respectively, unobservable random variables (random effects) such that \( E(v) = 0, E(e) = 0 \), and

\[ \text{Var} \begin{bmatrix} v \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \sigma^2 \]  

where \( G \) and \( R \) are known positive definite matrices, and \( \sigma^2 \) is a positive constant.

In the BLUP estimator, \textit{best} states that the resulting estimates have a minimum MSE value in other unbiased estimators, \textit{linear} states that the resulting estimator is a linear function of the data \( \tilde{\theta} \), and \textit{unbiased} states that the expected value of the estimate is equal to the value of the parameter.

Mathematically, the BLUP estimates \( \hat{\beta} \) of \( \beta \) and \( \hat{\psi} \) of \( \psi \) are defined as solutions to the following simultaneous equations which were given by Henderson (1950) and is known as the Henderson linear mixed model. The Henderson linear mixed model is used to find \( \hat{\beta} \) and \( \hat{\psi} \). Here is the Henderson linear mixed model:

\[ X^T R^{-1} X \beta + X^T R^{-1} Z v = X^T R^{-1} \tilde{\theta} \]  

\[ Z^T R^{-1} X \beta + (Z^T R^{-1} Z + G^{-1}) v = Z^T R^{-1} \tilde{\theta} \]  

the details of the calculation are reported in Appendix A.

3.1. Estimator for \( \beta \)

First, there will be an elimination of \( v \) between equation (15) and equation (16). To do this, multiply equation (16) with \( X^T R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} \) and perform a subtraction operation. The following results are obtained:

\[ X^T R^{-1} \beta - X^T R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \beta = \]  

\[ X^T R^{-1} \tilde{\theta} - X^T R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \tilde{\theta}. \]  

Now using a matrix identity which is commonly used in this subject area:

\[ V^{-1} := (ZGZ^T + R)^{-1} = R^{-1} - R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \]
Based on the model of equation (9) and the predicted result in Section 3, it is substituted by the

\[ X^T R^{-1} X \beta - X^T (R^{-1} - V^{-1}) X \beta = X^T R^{-1} \bar{\theta} - X^T (R^{-1} - V^{-1}) \bar{\theta} \]

or can be written as:

\[ X^T V^{-1} X \beta = X^T V^{-1} \bar{\theta}. \]

So that the BLUP estimator for \( \beta \) is:

\[ \hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \bar{\theta}. \]  \hspace{1cm} (19)

3.2. Estimator for \( \nu \)

Based on equation (16), where there is no fixed effect, the estimator for random effects is given as follows:

\[ Z^T R^{-1} X \beta + (Z^T R^{-1} Z + G^{-1}) \nu = Z^T R^{-1} \bar{\theta} \]

or can be written as:

\[ \nu = (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (\bar{\theta} - X \beta) \]  \hspace{1cm} (20)

Now using a matrix identity which is commonly used in this subject area:

\[ G Z^T V^{-1} = (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \]

gives

\[ \hat{\nu} = G Z^T V^{-1} (\bar{\theta} - X \beta). \]  \hspace{1cm} (22)

If \( \beta \) in equation (22) is substituted by \( \hat{\beta} \), then the estimator for \( \nu \) is the BLUP estimator. So that the BLUP estimator for \( \nu \) is:

\[ \hat{\nu} = G Z^T V^{-1} (\bar{\theta} - X \hat{\beta}). \]  \hspace{1cm} (23)

Then, the results of these two BLUP estimators will be used on the Fay-Herriot spatial model that has been defined earlier in this paper.

4. Implementation of BLUP method in the spatial model of the area level

Based on the model of equation (9) and the predicted result in Section 3, it is substituted by the covariance matrix (12) on the resultant estimator. The Spatial Best Linear Unbiased Predictor (Spatial BLUP) estimator of \( \theta_i \) is:

\[ \hat{\theta}_i = (\sigma_u^2, \rho) = x_i \hat{\beta} \]

\[ + b_i^T \left[ \sigma_u^2 \left[ (1 - \rho W)(1 - \rho W^T) \right]^{-1} \right] Z^T \left[ \text{diag}(\psi_i) \right] \]

\[ + Z \sigma_u^2 \left[ (1 - \rho W)(1 - \rho W^T) \right]^{-1} Z^T \bar{\theta} - X \hat{\beta} \]  \hspace{1cm} (24)

where \( \hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \bar{\theta} \) and \( b_i^T \) is \( 1 \times m \) vector \((0,0,...,0,1,0,...,0)\) with 1 referred to \( i \)-th area. Spatial BLUP will be the same value as BLUP under the model of random effect area when \( \rho = 0 \) (Pratesi & Salvati, 2005).

In this paper, the estimator in equation (24) is the mean population parameter estimator in a small area of the spatial Fay-Herriot model based on the BLUP method.

5. Final Remark

Our result shows that the Spatial BLUP estimator obtained concerning spatial dependence across areas, where the use of the available spatial auxiliary information can reduce both the bias and the sampling error in small area estimation. Moreover, give additional help in borrowing strength from the related area in the estimation of small area parameters.

However, the SBLUP estimator is still dependent on the component of the variance of the random area error \( \sigma_u^2 \) and spatial autoregressive coefficient \( \rho \) whose value is predictable.

Next step is to explore what happens when random area effects follow a conditional autoregressive (CAR) process or when the Fay-Herriot model with correlated sampling errors.
Acknowledgments
The authors are grateful to our campus for support. We also thank the referees for the many suggestions which improved the article.

Appendix A
It is assumed that \( v \) and \( e \) are normally distributed, and will maximize the joint probability density function (pdf) of \( \bar{\theta} \) (which depends on \( \beta \) and \( v \)) and \( v \).

Noted that
\[
E(v) = 0 \\
\text{Var}(v) = G\sigma^2
\]
and
\[
E(\bar{\theta}|\beta, v) = E(X\beta + Zv + e|\beta, v) = X\beta + Zv \\
\text{Var}(\bar{\theta}|\beta, v) = \text{Var}(X\beta + Zv + e|\beta, v) = R\sigma^2.
\]

Based on the multivariate normal distribution, pdf for \( v \) and \( \bar{\theta}|(\beta, v) \), respectively is:
\[
f(v) = \frac{1}{\sqrt{(2\pi)^n|G\sigma^2|^2}} \exp\left[-\frac{1}{2} \left( v^T(G\sigma^2)^{-1}v \right) \right]
\]
\[
= \frac{1}{\sqrt{(2\pi)^q(\sigma^2)q^2|G|^2}} \exp\left[-\frac{1}{2\sigma^2} (v^T(G)^{-1}v) \right]
\]
\[
= \frac{1}{(2\pi\sigma^2)^{q^2}G^2} \exp\left[-\frac{1}{2\sigma^2} (v^T(G)^{-1}v) \right] \quad (A.1)
\]

and
\[
f(\bar{\theta}|\beta, v) = \frac{1}{(2\pi\sigma^2)^{n^2}R^2} \exp\left[-\frac{1}{2\sigma^2} (\bar{\theta} - X\beta - Zv)^T(R)^{-1}(\bar{\theta} - X\beta - Zv) \right] \quad (A.2)
\]

Furthermore, with the additional assumption that \( v \) and \( \bar{\theta}|(\beta, v) \) are independent, then joint pdf for \( v \) dan \( \bar{\theta}|(\beta, v) \) is
\[
(2\pi\sigma^2)^{-\frac{n-1}{2}} \left( \det\left( \begin{array}{cc} G & 0 \\ 0 & R \end{array} \right) \right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2} \left( \bar{\theta} - X\beta - Zv \right)^T \left( \begin{array}{cc} G & 0 \\ 0 & R \end{array} \right)^{-1} \left( \bar{\theta} - X\beta - Zv \right) \right] \quad (A.3)
\]

To maximize the joint pdf respect to \( \beta \) and \( v \), that is by minimizing
\[
\min_{\beta, v} v^T G^{-1}v + (\bar{\theta} - X\beta - Zv)^T R^{-1} (\bar{\theta} - X\beta - Zv) \quad (A.4)
\]

In the end, will be minimized \( v^T G^{-1}v + (\bar{\theta} - X\beta - Zv)^T R^{-1} (\bar{\theta} - X\beta - Zv) \) by doing the derivative of \( \beta \) and \( v \) then equating the derived result with zero in such a way as to obtain:
\[
-2\bar{\theta}^T R^{-1}X + 2v^T Z^T R^{-1}X + 2\beta^T X^T R^{-1}X = 0 \quad (A.5)
\]
\[
-2\bar{\theta}^T R^{-1}Z + 2\beta^T X^T R^{-1}Z + 2v^T (Z^T R^{-1}Z + G^{-1}) = 0 \quad (A.6)
\]
or can be written as follows:
\[
X^T R^{-1}X\beta + X^T R^{-1}Zv = X^T R^{-1} \bar{\theta} \quad (A.7)
\]
\[
Z^T R^{-1}X\beta + (Z^T R^{-1}Z + G^{-1})v = Z^T R^{-1} \bar{\theta} \quad (A.8)
\]
where the last two equations are referred to as Henderson's linear mixed model.
References

[1] Anselin L 1992 Spatial Econometrics: Method and Models (Boston: Kluwer Academic Publishers)
[2] Anton H 1993 Elementary Linear Algebra (New York: John Wiley & Sons)
[3] Asfar 2016 Study of Determining Optimum Spatial Weighting Matrix in Small Area Estimation (thesis) (Bogor: Bogor Agricultural University)
[4] Bingham N H and Fry J M 2010 Regression: Linear Models In Statistics (London: Springer-Verlag, Inc)
[5] Chand N and Alexander C H 1995 Using administrative record for small area estimation in the American community survey (US Bureau of the Census)
[6] Fay R E and Herriot R A 1979 Estimates of income for small places: an application of James Stein procedures to census data. J Am Stat Assoc. 74:269-277
[7] Ghosh M and Rao J N K 1994 Small Area Estimation: An Appraisal. Statistical Science. 9(1): 55-76
[8] Harville D A 1991 That That BLUP is a Good Thing: The Estimation of Random Effects. Statistical Science. 6, 1, 35-39.
[9] Harville D A 1997 Matrix Algebra From a Statistician’s Perspective (New York: Springer-Verlag, Inc)
[10] Henderson C R 1975 Best linear unbiased estimation and prediction under a selection model. Biometrics. 31:423-447
[11] Hogg R V, McKean J W, and Craig A T 2005 Introduction to Mathematical Statistics 6th ed (Boston: Pearson Education, Inc)
[12] Jennrich R I and Sampson P F 1976 Newton-Rapson and related algorithms for maximum likelihood variance component estimation. Technometrics. 18, 11-17
[13] Kackar R N and Harville D A 1984 Approximations for Standard Error of Estimator for Fixed and Random Effect in Mixed Models. Journal of the American Statistical Association. 79, 853-862.
[14] Kurnia A 2009 An Empirical Best Prediction Method for Logarithmic Transformation Model in Small Area Estimation with Particular Application to Susenas Data (dissertation) (Bogor: Bogor Agricultural University)
[15] Matualage D 2012 Spatial Empirical Best Linear Unbiased Prediction Methods for Small Areas to Estimate monthly Expenditure per capita (Case Study: The Province of East Java District Jember) (Bogor: Bogor Agricultural University)
[16] Montgomery C, Peck A, and Vining G 2001 Introduction to Linear Regression Analysis (New York: John Wiley & Sons)
[17] Petrucci A and Salvati N 2006 Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment. Journal of Agricultural, Biological, and Environmental Statistics. 11(2): 169-182
[18] Pratesi M and Salvati N 2005 Small area estimation: the EBLUP estimator with autoregressive random area effects. Report no 261 Departement of Statistics and Mathematics, University of Pisa
[19] Pratesi M and Salvati N 2008 Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Stat. Meth. & Appl. 17:113-141
[20] Rao J N K 2003 Small Area Estimation (New York: John Wiley and Sons)
[21] Rencher A C 2002 Method of Multivariate Analysis 2nd ed (New York: John Wiley & Sons, Inc)
[22] Robinson G K 1991 That BLUP is a Good Thing: The Estimation of Random Effects. Statistical Science. 6, 1, 15-51
[23] Salvati N 2004 Small area estimation by spatial models: the spatial empirical best linear unbiased prediction (spatial EBLUP) Working Paper 2004/03. Firenze: Università degli Studi di Firenze
[24] Searle S R 1982 Matrix algebra useful for statistics (John Wiley & Sons, Inc)
[25] Susvitasari K 2014 *Searching for Point Estimation for Mean of Fay-Herriot Model and CAR Fay-Herriot Model Using Hierarchical Bayesian (HB) Approach on Small Area Estimation (SAE) (mini thesis)* (Depok: Universitas Indonesia)

[26] Yudistira 2014 *Point Estimation of Population Parameter of Small Area with Spatial Empirical Bayes based on Area Level Model (mini thesis)* (Depok: Universitas Indonesia)