

Department of Mathematics

Small area estimation under a  
multivariate linear model for  
incomplete repeated measures data

Innocent Ngaruye, Dietrich von Rosen and Martin Singull

LiTH-MAT-R--2017/06--SE

Department of Mathematics  
Linköping University  
S-581 83 Linköping

# Small area estimation under a multivariate linear model for incomplete repeated measures data

Innocent Ngaruye<sup>1,2</sup>, Dietrich von Rosen<sup>1,3</sup>  
and Martin Singull<sup>1</sup>

<sup>1</sup>Department of Mathematics, Linköping University,  
SE-581 83 Linköping, Sweden  
E-mail: {innocent.ngaruye, martin.singull}@liu.se

<sup>2</sup>Department of Mathematics,  
College of Science and Technology, University of Rwanda,  
P.O. Box 3900 Kigali, Rwanda  
E-mail: i.ngaruye@ur.ac.rw

<sup>3</sup>Department of Energy and Technology,  
Swedish University of Agricultural Sciences,  
SE-750 07 Uppsala, Sweden  
E-mail: dietrich.von.rosen@slu.se

## Abstract

In this paper, the issue of analysis of multivariate repeated measures data that follow a monotonic sample pattern for small area estimation is addressed. Random effects growth curve models with covariates for both complete and incomplete data are formulated. A conditional likelihood based approach is proposed for estimation of the mean parameters and covariances. Further, the prediction of random effects and predicted small area means are also discussed. The proposed techniques may be useful for small area estimation under longitudinal surveys with grouped response units and drop outs.

**Keywords:** Conditional likelihood, Multivariate linear model, Monotone sample, Repeated measures data.

## 1 Introduction

The estimation of the characteristics of interest for subpopulations (also called domains or small areas) for which there are small sample sizes taken from these subpopulations leads to

unreliable design-based estimators. The problem of producing reliable estimates for small areas whose domain specific sample sizes are not large enough to provide direct estimates with high precision and the assessment of the estimation error is known as the Small Area Estimation (SAE) problem (Pfeffermann, 2002). Rao (2003) has given a comprehensive overview of theory and methods of model-based SAE. Most surveys are conducted continuously in time based on cross-sectional repeated measures data. In Ngaruye et al. (2016a), the authors have proposed a multivariate linear model for repeated measures data in a SAE context. This model accounts for longitudinal surveys, grouped response units and time correlated random effects. However, it is natural to have incomplete repeated measures data in longitudinal surveys.

It is natural to have incomplete repeated measures data (missing observations) for different reasons. The missing data may result on the limitations of the researcher due to several reasons such as time or budget constraints, but also it may happen that units (e.g., individuals) for which the measurements are taken over time drop out from the study or are not available at a given time point. Several authors have dealt with this problem of missing data and we can refer to Carriere (1999); Srivastava (2002); Kim and Timm (2006); Longford (2006), for example. The latter has investigated the methods of multiple imputation for missing data and multivariate composition for small area estimation.

The estimation from missing or incomplete data in the classical growth curve models and in random effects growth curve model has been considered, for example, by Kleinbaum (1973); Srivastava (1985); Liski (1985); Liski and Nummi (1990); Nummi (1997). The expectation maximization algorithm is one of the method widely used to deal with missing values under the assumption of multivariate normal distribution of the data. However, we have found that this approach is not suitable for the proposed random effects growth curve model with covariates in small area estimation. In this paper, we address the issue of missing repeated measures data for SAE and propose a model which aims to borrow strength over time and space by including specific random area effects and time effects in a multivariate linear model for incomplete repeated measures data.

After the introduction, in Section 2 we give a motivating example for our considerations. In Section 3, we present the formulation of a multivariate linear model for repeated measures data which is considered for the rest of the paper. The formulation of the model for incomplete measures data its canonical decomposition are discussed in Section 4. In Section 5, the estimation of parameters and prediction of random effects and small area means are derived.

## 2 Motivating example

As an example, we consider the seasonal agricultural survey (SAS) conducted by the National Institute of Statistics of Rwanda (NISR). This survey covers every year three agricultural Seasons A, B and C in Rwanda with the intention of providing annual up-to-date information about agricultural statistics which can be used for monitoring agriculture programs and policies. The SAS uses a two-stage sampling scheme where at the first stage, primary sampling units are selected using probability proportional to sample size and a simple random sampling is used for selecting the secondary sampling units called segments or farms. The country which is divided into 30 administrative districts is divided into strata according to land-use characteristics. Since the survey is designed to provide national level statistics, the sample sizes taken from districts are too small to produce direct stable estimates of good precision at district level.

This SAS has been used in an empirical study by Ngaruye et al. (2016b) that aimed on producing reliable beans yield estimates at district level for agricultural seasons 2014 in Rwanda. Ngaruye et al. (2016b) focused on producing district level estimates for bush beans and climbing beans. For each cultivated farmland (or segment considered in this study), only one variety of beans was observed. It is not applicable for some pieces of farmland to be cultivated for consecutive seasons, but also it is common in Rwanda to practice fallow system in order to restore fertility of some pieces of farmland so that they are not cultivated in consecutive seasons. For this reason, only 152 segments remained in the sample during season C out of 540 which were selected during season A and B. Therefore, the multivariate linear model proposed by Ngaruye et al. (2016a) was not suitable to account for these incomplete data, and that study was limited to season A and B.

## 3 Multivariate linear model for repeated measures data

Consider the multivariate linear regression model for repeated measurements with covariates at  $p$  time points defined in Ngaruye et al. (2016a) when the data are complete at all time points. We suppose that the target population of size  $N$  whose characteristic of interest  $y$  is divided into  $m$  subpopulations called small areas of sizes  $N_i, i = 1, \dots, m$  and the units in all small areas are grouped in  $k$  different categories. Further, we assume the mean growth of the  $j$ -th unit in area  $i$  for each one of the  $k$  group groups to be a polynomial in time with degree  $q - 1$  and suppose that we have covariate variables related to the characteristic of interest whose values  $x_{ij}$  are available for all units in the population. The model at small

area level is written as

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{A}\mathbf{B}\mathbf{C}_i + \mathbf{1}_p\boldsymbol{\gamma}'\mathbf{X}_i + \mathbf{u}_i\mathbf{z}'_i + \mathbf{E}_i, \\ \mathbf{u}_i &\sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}^u), \\ \mathbf{E}_i &\sim \mathcal{N}_{p, N_i}(\mathbf{0}, \boldsymbol{\Sigma}_e, \mathbf{I}_{N_i}), \end{aligned}$$

and the corresponding model combining all disjoint  $m$  small areas and all  $N$  units divided into  $k$  non-overlapping group units given by

$$\begin{aligned} \mathbf{Y} &= \mathbf{A}\mathbf{B}\mathbf{H}\mathbf{C} + \mathbf{1}_p\boldsymbol{\gamma}'\mathbf{X} + \mathbf{U}\mathbf{Z} + \mathbf{E}, \\ \mathbf{E} &\sim \mathcal{N}_{p, N}(\mathbf{0}, \boldsymbol{\Sigma}_e, \mathbf{I}_N), \\ \mathbf{U} &\sim \mathcal{N}_{p, m}(\mathbf{0}, \boldsymbol{\Sigma}^u, \mathbf{I}_m), \quad p \leq m, \end{aligned} \tag{1}$$

where  $\boldsymbol{\Sigma}^u$  is an unknown arbitrary positive definite matrix and  $\boldsymbol{\Sigma}_e = \sigma_e^2\mathbf{I}_p$  is assumed to be known. In model (1),  $\mathbf{Y} : p \times N$  is the data matrix,  $\mathbf{A} : p \times q$ ,  $q \leq p$  is the within individual design matrix indicating the time dependency within individuals,  $\mathbf{B} : q \times k$  is unknown parameter matrix,  $\mathbf{C} : mk \times N$  with  $\text{rank}(\mathbf{C}) + p \leq N$  and  $p \leq m$  is the between individuals design matrix accounting for group effects, the matrix  $\mathbf{U} : p \times m$  is a matrix of random effect whose columns are assumed to be independently distributed as a multivariate normal distribution with mean zero and a positive dispersion matrix  $\boldsymbol{\Sigma}^u$ , i.e.,  $\mathbf{U} \sim N_{p, m}(\mathbf{0}, \boldsymbol{\Sigma}^u, \mathbf{I}_m)$ ,  $\mathbf{Z} : m \times N$  is a design matrix for random effect and the columns of the error matrix  $\mathbf{E}$  are assumed to be independently distributed as  $p$ -variate normal distribution with mean zero and and known covariance matrix  $\boldsymbol{\Sigma}_e$ , i.e.,  $\mathbf{E} \sim N_{p, N}(\mathbf{0}, \boldsymbol{\Sigma}_e, \mathbf{I}_N)$ . More details about model formulation and estimation of model parameters can be found in Ngaruye et al. (2016a).

In what follows, we denote by  $\mathbf{A}^\circ$  any matrix of full rank spanning  $\mathcal{C}(\mathbf{A})^\perp$ , i.e.,  $\mathcal{C}(\mathbf{A}^\circ) = \mathcal{C}(\mathbf{A})^\perp$ , where  $\mathcal{C}(\mathbf{A})$  stands for the column vector space generated by the columns of the matrix  $\mathbf{A}$  and  $\mathcal{C}(\mathbf{A})^\perp$  its orthogonal complement. Moreover,  $\mathbf{P}_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$  denotes the orthogonal projection on  $\mathcal{C}(\mathbf{A})$ .

## 4 Incomplete repeated measures data

We now consider the corresponding model for sample data to the model (1) and suppose that there are missing values in such a way that the measurements taken at time  $t$ , (for  $t = 1, \dots, p$ ), on unit  $j$  are not all complete and the number of observations for the different  $p$  time points are  $n_1, \dots, n_p$  with  $n_1 \geq n_2 \geq \dots \geq n_p > p$ . Such a pattern of missing observations follows a so called *monotone sample*. Let the sample observations be composed of mutually disjoint  $h$  sets according to the monotonic pattern of missing data, where the  $l$ -th set, ( $l = 1, \dots, h$ ), is the sample data matrix  $\mathbf{Y}_l : p_l \times n_l$  whose units in the sample have completed  $l - 1$  periods and failed to complete the  $l$ th period with  $p_l \leq p$  and  $\sum_{l=1}^h p_l = p$ . For simplicity, in this paper we only study the case of two-step monotone missing data with

complete sample data for a given number of time points and incomplete sample data for an other given number of time points. Note that  $n_i$  is the size of the sample drawn from the  $i$ -th area of population size  $N_i$  with  $n = \sum_{i=1}^m n_i$  and  $N = \sum_{i=1}^m N_i$ .

#### 4.1 Model formulation

In this article we will only present details for a two-step monotonic pattern. We assume that the model defined in (1) holds for both sampled and non-sampled population units and with this type of monotonic missing structure, the model (1) for sample data can be presented by two equations:

$$\mathbf{Y} = \mathbf{A}_1 \mathbf{B} \mathbf{H} \mathbf{C} + \mathbf{1}_{p_1} \boldsymbol{\gamma}' \mathbf{X} + \mathbf{U}_1 \mathbf{Z} + \mathbf{E}_1, \quad (2)$$

$$\tilde{\mathbf{Y}} = \mathbf{A}_2 \mathbf{B} \mathbf{H} \tilde{\mathbf{C}} + \mathbf{1}_{p_2} \boldsymbol{\gamma}' \tilde{\mathbf{X}} + \mathbf{U}_2 \tilde{\mathbf{Z}} + \mathbf{E}_2, \quad (3)$$

where

$$\begin{aligned} \mathbf{U}_1 &\sim \mathcal{N}_{p_1, m}(\mathbf{0}, \boldsymbol{\Sigma}_{11}^u, \mathbf{I}_m), & \mathbf{E}_1 &\sim \mathcal{N}_{p_1, n}(\mathbf{0}, \mathbf{I}_{p_1}, \sigma_{E_1}^2 \mathbf{I}_n), \\ \mathbf{U}_2 &\sim \mathcal{N}_{p_2, m}(\mathbf{0}, \boldsymbol{\Sigma}_{22}^u, \mathbf{I}_m), & \mathbf{E}_2 &\sim \mathcal{N}_{p_2, n_1}(\mathbf{0}, \mathbf{I}_{p_2}, \sigma_{E_2}^2 \mathbf{I}_{n_1}). \end{aligned}$$

Note that (2) models the first  $p_1$  time points where we have complete observations for all units, whereas (3) models the last  $p_2$  time points ( $p = p_1 + p_2$ ) for those units where complete observations were obtained. This means that  $\mathbf{A}' = (\mathbf{A}'_1 : \mathbf{A}'_2)$ ,  $\tilde{\mathbf{C}}$  is included in  $\mathbf{C}$  and  $\tilde{\mathbf{X}}$  is included in  $\mathbf{X}$ . The matrices  $\mathbf{Z}$  and  $\tilde{\mathbf{Z}}$  are given by

$$\mathbf{Z}_{m \times n} = \begin{pmatrix} \mathbf{z}'_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{z}'_m \end{pmatrix}, \quad \tilde{\mathbf{Z}}_{m \times n_1} = \begin{pmatrix} \tilde{\mathbf{z}}'_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \tilde{\mathbf{z}}'_m \end{pmatrix},$$

where  $\mathbf{z}_i = \frac{1}{\sqrt{n_i}} \mathbf{1}_{n_i}$  and  $\tilde{\mathbf{z}}_i = \frac{1}{\sqrt{n_{1i}}} \mathbf{1}_{n_{1i}}$ . Moreover, it will be used that  $\mathcal{C}(\mathbf{Z}') \subseteq \mathcal{C}(\mathbf{C}')$ ,  $\mathcal{C}(\tilde{\mathbf{Z}}') \subseteq \mathcal{C}(\tilde{\mathbf{C}}')$  and  $\mathbf{Z} \mathbf{Z}' = \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}' = \mathbf{I}_m$ .

The main idea when finding estimators is to first derive a canonical form of the model presented in (2) and (3). Initially the mathematics follows the transformation given by Ngaruye et al. (2016a) and the main ideas of that article are adopted. Let

$$\begin{aligned} \mathbf{K}_1 &= \mathbf{H}(\mathbf{C} \mathbf{C}')^{1/2} \boldsymbol{\Gamma}_1, & \mathbf{K}_2 &= \mathbf{H}(\mathbf{C} \mathbf{C}')^{1/2} \boldsymbol{\Gamma}_2, \\ \mathbf{R}_1 &= \mathbf{C}'(\mathbf{C} \mathbf{C}')^{-1/2} \boldsymbol{\Gamma}_1, & \mathbf{R}_2 &= \mathbf{C}'(\mathbf{C} \mathbf{C}')^{-1/2} \boldsymbol{\Gamma}_2, \end{aligned}$$

and

$$\begin{aligned} \tilde{\mathbf{K}}_1 &= \tilde{\mathbf{H}}(\tilde{\mathbf{C}} \tilde{\mathbf{C}}')^{1/2} \tilde{\boldsymbol{\Gamma}}_1, & \tilde{\mathbf{K}}_2 &= \tilde{\mathbf{H}}(\tilde{\mathbf{C}} \tilde{\mathbf{C}}')^{1/2} \tilde{\boldsymbol{\Gamma}}_2, \\ \tilde{\mathbf{R}}_1 &= \tilde{\mathbf{C}}'(\tilde{\mathbf{C}} \tilde{\mathbf{C}}')^{-1/2} \tilde{\boldsymbol{\Gamma}}_1, & \tilde{\mathbf{R}}_2 &= \tilde{\mathbf{C}}'(\tilde{\mathbf{C}} \tilde{\mathbf{C}}')^{-1/2} \tilde{\boldsymbol{\Gamma}}_2. \end{aligned}$$

Now the model is one-to-one transformed into six different models:

$$\begin{aligned}
\mathbf{V}_0 &= \mathbf{Y}\mathbf{C}'^o = \mathbf{1}_{p_1}\boldsymbol{\gamma}'\mathbf{X}\mathbf{C}'^o + \mathbf{E}_1\mathbf{C}'^o, \\
\mathbf{V}_1 &= \mathbf{Y}\mathbf{R}_1 = \mathbf{A}_1\mathbf{B}\mathbf{K}_1 + \mathbf{1}_{p_1}\boldsymbol{\gamma}'\mathbf{X}\mathbf{R}_1 + (\mathbf{U}_1\mathbf{Z} + \mathbf{E}_1)\mathbf{R}_1, \\
\mathbf{V}_2 &= \mathbf{Y}\mathbf{R}_2 = \mathbf{A}_1\mathbf{B}\mathbf{K}_2 + \mathbf{1}_{p_1}\boldsymbol{\gamma}'\mathbf{X}\mathbf{R}_2 + \mathbf{E}_1\mathbf{R}_2, \\
\tilde{\mathbf{V}}_0 &= \tilde{\mathbf{Y}}\tilde{\mathbf{C}}'^o = \mathbf{1}_{p_2}\boldsymbol{\gamma}'\tilde{\mathbf{X}}\tilde{\mathbf{C}}'^o + \mathbf{E}_2\tilde{\mathbf{C}}'^o, \\
\tilde{\mathbf{V}}_1 &= \tilde{\mathbf{Y}}\tilde{\mathbf{R}}_1 = \mathbf{A}_2\mathbf{B}\tilde{\mathbf{K}}_1 + \mathbf{1}_{p_2}\boldsymbol{\gamma}'\tilde{\mathbf{X}}\tilde{\mathbf{R}}_1 + (\mathbf{U}_2\tilde{\mathbf{Z}} + \mathbf{E}_2)\tilde{\mathbf{R}}_1, \\
\tilde{\mathbf{V}}_2 &= \tilde{\mathbf{Y}}\tilde{\mathbf{R}}_2 = \mathbf{A}_2\mathbf{B}\tilde{\mathbf{K}}_2 + \mathbf{1}_{p_2}\boldsymbol{\gamma}'\tilde{\mathbf{X}}\tilde{\mathbf{R}}_2 + \mathbf{E}_2\tilde{\mathbf{R}}_2.
\end{aligned}$$

Note that the random error terms  $\mathbf{E}_1$  and  $\mathbf{E}_2$  are connected to sampling variances assumed to be known since the errors stem from surveys and are connected only to the sampling sizes not to model assumptions of the means. Since the surveys are independent, then the models  $\mathbf{V}_0, \tilde{\mathbf{V}}_0, \mathbf{V}_2$  and  $\tilde{\mathbf{V}}_2$  do not account for area and time effects.

Because  $\mathbf{E}_1$  is independent of  $\mathbf{E}_2$ ,  $\mathbf{E}_1\mathbf{C}'^o$  is independent of  $\mathbf{E}_1\mathbf{R}_1$  and  $\mathbf{E}_1\mathbf{R}_2$ , and  $\mathbf{E}_1\mathbf{R}_1$  is independent of  $\mathbf{E}_1\mathbf{R}_2$  and because  $\mathbf{E}_2\tilde{\mathbf{C}}'^o$  is independent of  $\mathbf{E}_2\tilde{\mathbf{R}}_1$  and  $\mathbf{E}_2\tilde{\mathbf{R}}_2$ , and  $\mathbf{E}_2\tilde{\mathbf{R}}_1$  is independent of  $\mathbf{E}_2\tilde{\mathbf{R}}_2$  it follows that  $\mathbf{V}_0, \mathbf{V}_2, \tilde{\mathbf{V}}_0, \tilde{\mathbf{V}}_2$  are mutually independent, and independent of  $\mathbf{V}_1, \tilde{\mathbf{V}}_1$ , but  $\mathbf{V}_1$  and  $\tilde{\mathbf{V}}_1$  are not independently distributed.

Thus, the likelihood based on  $\mathbf{V}_0, \mathbf{V}_2, \tilde{\mathbf{V}}_0, \tilde{\mathbf{V}}_2, \mathbf{V}_1, \tilde{\mathbf{V}}_1$  can easily be factored and what remains to be discussed is how the likelihood for  $\mathbf{V}_1, \tilde{\mathbf{V}}_1$  can be expressed. The natural approach is to factor the joint distribution by conditioning  $\tilde{\mathbf{V}}_1$  given  $\mathbf{V}_1$  and then also use the marginal distribution for  $\mathbf{V}_1$ . Note that it is only  $\mathbf{V}_1$  and  $\tilde{\mathbf{V}}_1$  which bear information about  $\boldsymbol{\Sigma}^u$  and we have the following lemma about the conditional distribution.

**Lemma 4.1.** *The conditional distribution of  $\tilde{\mathbf{V}}_1$  given  $\mathbf{V}_1$  is expressed by*

$$\tilde{\mathbf{V}}_1 | \mathbf{V}_1 \sim \mathcal{N}_{p_2, m}(\mathbf{M}_{2\bullet 1}, \boldsymbol{\Psi}_{2\bullet 1}, \mathbf{I}_m),$$

where

$$\begin{aligned}
\mathbf{M}_{2\bullet 1} &= \mathbf{A}_2\mathbf{B}\tilde{\mathbf{K}}_1 + \mathbf{1}_{p_2}\boldsymbol{\gamma}'\tilde{\mathbf{X}}\tilde{\mathbf{R}}_1 \\
&\quad + \boldsymbol{\Sigma}_{21}^u (\boldsymbol{\Sigma}_{11}^u + \sigma_{E_1}^2 \mathbf{I}_{p_1})^{-1} (\mathbf{Y}\mathbf{R}_1 - \mathbf{A}_1\mathbf{B}\mathbf{K}_1 - \mathbf{1}_{p_1}\boldsymbol{\gamma}'\mathbf{X}\mathbf{R}_1) \mathbf{R}_1' \mathbf{Z}' \tilde{\mathbf{Z}} \tilde{\mathbf{R}}_1, \\
\boldsymbol{\Psi}_{2\bullet 1} &= \boldsymbol{\Sigma}_{22}^u + \sigma_{E_2}^2 \mathbf{I}_{p_2} - \boldsymbol{\Sigma}_{21}^u (\boldsymbol{\Sigma}_{11}^u + \sigma_{E_1}^2 \mathbf{I}_{p_1})^{-1} \boldsymbol{\Sigma}_{12}^u.
\end{aligned}$$

*Proof.* In order to fully determine the conditional distribution, since it is normal, we will study  $E[\text{vec}\tilde{\mathbf{V}}_1 | \text{vec}\mathbf{V}_1]$  and  $D[\tilde{\mathbf{V}}_1 | \mathbf{V}_1]$ .

$$\begin{aligned}
E[\text{vec}\tilde{\mathbf{V}}_1 | \text{vec}\mathbf{V}_1] &= \text{vec}(\mathbf{A}_2\mathbf{B}\tilde{\mathbf{K}}_1 + \mathbf{1}_{p_2}\boldsymbol{\gamma}'\tilde{\mathbf{X}}\tilde{\mathbf{R}}_1) \\
&\quad + C[\tilde{\mathbf{V}}_1, \mathbf{V}_1] D[\mathbf{V}_1]^{-1} \text{vec}(\mathbf{V}_1 - \mathbf{A}_1\mathbf{B}\mathbf{K}_1 + \mathbf{1}_{p_1}\boldsymbol{\gamma}'\tilde{\mathbf{X}}\mathbf{R}_1).
\end{aligned}$$

Moreover, since  $\mathbf{R}_1' \mathbf{Z}' \mathbf{Z} \mathbf{R}_1 = \mathbf{I}_m$  and  $\mathbf{R}_1' \mathbf{R}_1 = \mathbf{I}_m$ , we have

$$D[\mathbf{V}_1] = \mathbf{I}_m \otimes (\boldsymbol{\Sigma}_{11}^u + \sigma_{E_1}^2 \mathbf{I}_{p_1}),$$



and similarly,

$$D[\tilde{\mathbf{V}}_1] = \mathbf{I}_m \otimes (\boldsymbol{\Sigma}_{22}^u + \sigma_{E_2}^2 \mathbf{I}_{p_2}).$$

For the covariance we have

$$C[\tilde{\mathbf{V}}_1, \mathbf{V}_1] = \tilde{\mathbf{R}}_1' \tilde{\mathbf{Z}}' \mathbf{Z} \mathbf{R}_1 \otimes \boldsymbol{\Sigma}_{21}^u$$

and

$$C[\tilde{\mathbf{V}}_1, \mathbf{V}_1] D[\mathbf{V}_1]^{-1} = \tilde{\mathbf{R}}_1' \tilde{\mathbf{Z}}' \mathbf{Z} \mathbf{R}_1 \otimes \boldsymbol{\Sigma}_{21}^u (\boldsymbol{\Sigma}_{11}^u + \sigma_{E_1}^2 \mathbf{I}_{p_1})^{-1}.$$

The conditional variance can then be written as

$$\begin{aligned} D[\tilde{\mathbf{V}}_1 | \mathbf{V}_1] &= D[\tilde{\mathbf{V}}_1] - C[\tilde{\mathbf{V}}_1, \mathbf{V}_1] D[\mathbf{V}_1]^{-1} C[\tilde{\mathbf{V}}_1, \mathbf{V}_1]' \\ &= \mathbf{I}_m \otimes (\boldsymbol{\Sigma}_{22}^u + \sigma_{E_2}^2 \mathbf{I}_{p_2}) - \tilde{\mathbf{R}}_1' \tilde{\mathbf{Z}}' \mathbf{Z} \mathbf{R}_1 \mathbf{R}_1' \mathbf{Z}' \tilde{\mathbf{Z}} \tilde{\mathbf{R}}_1 \otimes \boldsymbol{\Sigma}_{21}^u (\boldsymbol{\Sigma}_{11}^u + \sigma_{E_1}^2 \mathbf{I}_{p_1})^{-1} \boldsymbol{\Sigma}_{12}^u \\ &= \mathbf{I}_m \otimes \left( \boldsymbol{\Sigma}_{22}^u + \sigma_{E_2}^2 \mathbf{I}_{p_2} - \boldsymbol{\Sigma}_{21}^u (\boldsymbol{\Sigma}_{11}^u + \sigma_{E_1}^2 \mathbf{I}_{p_1})^{-1} \boldsymbol{\Sigma}_{12}^u \right), \end{aligned}$$

since  $\tilde{\mathbf{R}}_1' \tilde{\mathbf{Z}}' \mathbf{Z} \mathbf{R}_1 \mathbf{R}_1' \mathbf{Z}' \tilde{\mathbf{Z}} \tilde{\mathbf{R}}_1 = \mathbf{I}_m$  which follows from the next two relations:

$$\begin{aligned} \mathbf{Z} \mathbf{R}_1 \mathbf{R}_1' \mathbf{Z}' &= \mathbf{Z} \mathbf{C}' (\mathbf{C} \mathbf{C}')^{-1/2} \boldsymbol{\Gamma}_1 \boldsymbol{\Gamma}_1' (\mathbf{C} \mathbf{C}')^{-1/2} \mathbf{C} \mathbf{Z}' \\ &= \mathbf{Z} \mathbf{P}_{\mathbf{C}'} \mathbf{Z}' \mathbf{Z} \mathbf{P}_{\mathbf{C}'} \mathbf{Z}' = \mathbf{Z} \mathbf{Z}' \mathbf{Z} \mathbf{Z}' = \mathbf{I}_m \end{aligned}$$

because  $\mathcal{C}(\mathbf{Z}') \subseteq \mathcal{C}(\mathbf{C}')$  and

$$\begin{aligned} \tilde{\mathbf{R}}_1' \tilde{\mathbf{Z}}' \tilde{\mathbf{Z}} \tilde{\mathbf{R}}_1 &= \tilde{\boldsymbol{\Gamma}}_1' (\tilde{\mathbf{C}} \tilde{\mathbf{C}}')^{-1/2} \tilde{\mathbf{C}} \tilde{\mathbf{Z}}' \tilde{\mathbf{Z}} \tilde{\mathbf{C}}' (\tilde{\mathbf{C}} \tilde{\mathbf{C}}')^{-1/2} \tilde{\boldsymbol{\Gamma}}_1 \\ &= \tilde{\boldsymbol{\Gamma}}_1' \tilde{\boldsymbol{\Gamma}}_1 \tilde{\boldsymbol{\Gamma}}_1' \tilde{\boldsymbol{\Gamma}}_1 = \mathbf{I}_m, \end{aligned}$$

which completes the proof of the lemma.  $\square$

Denote the likelihood for the observations  $\mathbf{Y}$  with parameters  $\boldsymbol{\Theta}$  by  $L(\mathbf{Y}; \boldsymbol{\Theta})$  and using this notation we have the following theorem.

**Theorem 4.1.** *For the model given by (2) and (3) let  $\boldsymbol{\Psi}_{22} = \boldsymbol{\Sigma}_{22}^u + \sigma_{E_2}^2 \mathbf{I}_{p_2}$ ,  $\boldsymbol{\Psi}_{11} = \boldsymbol{\Sigma}_{11}^u + \sigma_{E_1}^2 \mathbf{I}_{p_1}$ ,  $\boldsymbol{\Theta} = \boldsymbol{\Sigma}_{21}^u \boldsymbol{\Psi}_{11}^{-1}$  and  $\boldsymbol{\Psi}_{2\bullet 1} = \boldsymbol{\Psi}_{22} - \boldsymbol{\Theta} \boldsymbol{\Psi}_{11} \boldsymbol{\Theta}'$ . Then*

$$\begin{aligned} L(\mathbf{Y}, \tilde{\mathbf{Y}}; \mathbf{B}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}^u) &= L(\mathbf{V}_0; \boldsymbol{\gamma}) L(\tilde{\mathbf{V}}_0; \boldsymbol{\gamma}) L(\mathbf{V}_2; \mathbf{B}, \boldsymbol{\gamma}) L(\tilde{\mathbf{V}}_0; \mathbf{B}, \boldsymbol{\gamma}) \\ &\quad \times L(\tilde{\mathbf{V}}_1 | \mathbf{V}_1; \mathbf{B}, \boldsymbol{\gamma}, \boldsymbol{\Psi}_{2\bullet 1}, \boldsymbol{\Theta}) L(\mathbf{V}_1; \mathbf{B}, \boldsymbol{\gamma}, \boldsymbol{\Psi}_{11}), \end{aligned}$$

where

$$\begin{aligned} L(\mathbf{V}_0; \boldsymbol{\gamma}) &= c_1 \exp \left\{ \frac{\sigma_{E_1}^{-2}}{2} \text{tr} \left\{ \left( \mathbf{Y} \mathbf{C}'^o - \mathbf{1}_{p_1} \boldsymbol{\gamma}' \mathbf{X} \mathbf{C}'^o \right) \left( \right)' \right\} \right\}, \\ L(\tilde{\mathbf{V}}_0; \boldsymbol{\gamma}) &= c_2 \exp \left\{ \frac{\sigma_{E_1}^{-2}}{2} \text{tr} \left\{ \left( \tilde{\mathbf{Y}} \tilde{\mathbf{C}}'^o - \mathbf{1}_{p_2} \boldsymbol{\gamma}' \tilde{\mathbf{X}} \tilde{\mathbf{C}}'^o \right) \left( \right)' \right\} \right\}, \end{aligned}$$

$$\begin{aligned}
L(\mathbf{V}_2; \mathbf{B}, \gamma) &= c_3 \exp \left\{ \frac{\sigma_{E_1}^{-2}}{2} \text{tr} \left\{ \left( \mathbf{Y} \mathbf{R}_2 - \mathbf{A}_1 \mathbf{B} \mathbf{K}_2 - \mathbf{1}_{p_1} \gamma' \mathbf{X} \mathbf{R}_2 \right) \left( \right)' \right\} \right\}, \\
L(\tilde{\mathbf{V}}_2; \mathbf{B}, \gamma) &= c_4 \exp \left\{ \frac{\sigma_{E_1}^{-2}}{2} \text{tr} \left\{ \left( \tilde{\mathbf{Y}} \tilde{\mathbf{R}}_2 - \mathbf{A}_2 \mathbf{B} \tilde{\mathbf{K}}_2 - \mathbf{1}_{p_2} \gamma' \tilde{\mathbf{X}} \tilde{\mathbf{R}}_2 \right) \left( \right)' \right\} \right\}, \\
L(\tilde{\mathbf{V}}_1 | \mathbf{V}_1; \mathbf{B}, \gamma, \Psi_{2 \bullet 1}, \Theta) &= c_5 |\Psi_{2 \bullet 1}|^{m/2} \exp \left\{ \frac{1}{2} \text{tr} \left\{ \Psi_{2 \bullet 1}^{-1} \left( \tilde{\mathbf{Y}} \tilde{\mathbf{R}}_1 - \mathbf{A}_2 \mathbf{B} \tilde{\mathbf{K}}_1 - \mathbf{1}_{p_2} \gamma' \tilde{\mathbf{X}} \tilde{\mathbf{R}}_1 \right. \right. \right. \\
&\quad \left. \left. \left. - \Theta \left( \mathbf{Y} \mathbf{R}_1 - \mathbf{A}_1 \mathbf{B} \mathbf{K}_1 - \mathbf{1}_{p_1} \gamma' \mathbf{X} \mathbf{R}_1 \right) \mathbf{R}'_1 \mathbf{Z}' \tilde{\mathbf{Z}} \tilde{\mathbf{R}}_1 \right) \left( \right)' \right\} \right\}, \\
L(\mathbf{V}_1; \mathbf{B}, \gamma, \Psi_{11}) &= c_6 |\Psi_{11}|^{m/2} \exp \left\{ \frac{1}{2} \text{tr} \left\{ \Psi_{11}^{-1} \left( \mathbf{Y} \mathbf{R}_1 - \mathbf{A}_1 \mathbf{B} \mathbf{K}_1 - \mathbf{1}_{p_1} \gamma' \mathbf{X} \mathbf{R}_1 \right) \left( \right)' \right\} \right\},
\end{aligned}$$

where  $c_1, \dots, c_6$  are known constants.

## 5 Estimation of parameters and prediction of small area means

Since the monotone missing value problem described and treated in the previous sections was possible to present in a canonical form, which fortunately seems easy to utilize, the remaining part of the report consists of a relatively straight forward standard approach for predicting small areas.

### 5.1 Estimation

In order to estimate the parameters in the model given by (2) and (3) we propose a restricted likelihood approach described in the next proposition.

**Proposition 5.1** For the likelihood given in Theorem 4.1  $\mathbf{B}$  and  $\gamma$  are estimated by maximizing

$$L(\mathbf{V}_0; \gamma) \times L(\tilde{\mathbf{V}}_0; \gamma) \times L(\mathbf{V}_2; \mathbf{B}, \gamma) \times L(\tilde{\mathbf{V}}_2; \mathbf{B}, \gamma) \quad (4)$$

and  $\Sigma^u$  is estimated by maximizing

$$L(\tilde{\mathbf{V}}_1 | \mathbf{V}_1; \mathbf{B}, \gamma, \Psi_{2 \bullet 1}, \Theta) \times L(\mathbf{V}_1; \mathbf{B}, \gamma, \Psi_{11}),$$

when inserting the estimated  $\mathbf{B}$  and  $\gamma$ , and using  $\Sigma_{11}^u = \Psi_{11} - \sigma_{E_1}^2 \mathbf{I}_{p_1}$ ,  $\Sigma_{22}^u = \Psi_{2 \bullet 1} - \sigma_{E_2}^2 \mathbf{I}_{p_2} + \Theta \Psi_{11} \Theta'$  and  $\Sigma_{21}^u = \Theta \Psi_{11}$ .

Note that by applying Proposition 5.1 it is fairly easy to obtain explicit estimators for all parameters in the model (2) and (3). In the below given presentation estimators are indicated by  $\hat{\cdot}$ , i.e.,  $\hat{\mathbf{B}}$ ,  $\hat{\gamma}$  and  $\hat{\Sigma}^u$ .

Differentiating the joint likelihood (4) with respect to  $\gamma$  and  $\mathbf{B}$  leads to the likelihood equations

$$\begin{aligned} & \sigma_{E_1}^{-2} \left( (\mathbf{X}\mathbf{C}'^o(\mathbf{C}'^o)' + \mathbf{X}\mathbf{R}_2\mathbf{R}_2')\mathbf{Y}'\mathbf{1}_{p_1} - \mathbf{X}\mathbf{R}_2\mathbf{K}'_2\mathbf{B}'\mathbf{A}'_1\mathbf{1}_{p_1} \right) \\ & + \sigma_{E_2}^{-2} \left( (\tilde{\mathbf{X}}\tilde{\mathbf{C}}'^o(\tilde{\mathbf{C}}'^o)' + \tilde{\mathbf{X}}\tilde{\mathbf{R}}_2\tilde{\mathbf{R}}_2')\tilde{\mathbf{Y}}'\mathbf{1}_{p_2} - \tilde{\mathbf{X}}\tilde{\mathbf{R}}_2\tilde{\mathbf{K}}'_1\mathbf{B}'\mathbf{A}'_2\mathbf{1}_{p_2} \right) \\ & - \sigma_{E_1}^{-2} p_1 \mathbf{X} \left( \mathbf{C}'^o(\mathbf{C}'^o)' + \mathbf{R}_2\mathbf{R}_2' \right) \mathbf{X}'\gamma - \sigma_{E_2}^{-2} p_2 \tilde{\mathbf{X}} \left( \tilde{\mathbf{C}}'^o(\tilde{\mathbf{C}}'^o)' + \tilde{\mathbf{R}}_2\tilde{\mathbf{R}}_2' \right) \tilde{\mathbf{X}}'\gamma = \mathbf{0}, \end{aligned}$$

$$\begin{aligned} & \sigma_{E_1}^{-2} \mathbf{A}'_1 \left( \mathbf{Y}\mathbf{R}_2 - \mathbf{A}_1\mathbf{B}\mathbf{K}_2 - \mathbf{1}_{p_1}\gamma'\mathbf{X}\mathbf{R}_2 \right) \mathbf{K}'_2 \\ & + \sigma_{E_2}^{-2} \mathbf{A}'_2 \left( \tilde{\mathbf{Y}}\tilde{\mathbf{R}}_2 - \mathbf{A}_2\mathbf{B}\tilde{\mathbf{K}}_2 - \mathbf{1}_{p_2}\gamma'\tilde{\mathbf{X}}\tilde{\mathbf{R}}_2 \right) \tilde{\mathbf{K}}'_2 = \mathbf{0}. \end{aligned}$$

Assume that

$$\mathbf{F} = \sigma_{E_1}^{-2} \mathbf{K}_2\mathbf{K}'_2 \otimes \mathbf{A}'_1\mathbf{A}_1 + \sigma_{E_2}^{-2} \tilde{\mathbf{K}}_2\tilde{\mathbf{K}}'_2 \otimes \mathbf{A}'_2\mathbf{A}_2$$

is of full rank. Then,  $\gamma$  and  $\text{vec}\mathbf{B}$  can be estimated by

$$\begin{aligned} \hat{\gamma} &= \left[ \mathbf{X}\mathbf{P}\mathbf{Y}'\mathbf{1}_{p_1} + \tilde{\mathbf{X}}\tilde{\mathbf{P}}\tilde{\mathbf{Y}}'\mathbf{1}_{p_2} - \left( \mathbf{X}\mathbf{R}_2\mathbf{K}'_2 \otimes \mathbf{1}'_{p_1}\mathbf{A}_1 + \tilde{\mathbf{X}}\tilde{\mathbf{R}}_2\tilde{\mathbf{K}}'_2 \otimes \mathbf{1}'_{p_2}\mathbf{A}_2 \right) \mathbf{F}^{-1} \right. \\ & \quad \times \left. \left( \mathbf{K}_2\mathbf{R}'_2 \otimes \mathbf{1}_{p_1} + \tilde{\mathbf{K}}_2\tilde{\mathbf{R}}'_2 \otimes \mathbf{1}_{p_2} \right) \right]^{-1} \left[ \begin{aligned} & p_1 \mathbf{X}\mathbf{P}\mathbf{X}' + p_2 \tilde{\mathbf{X}}\tilde{\mathbf{P}}\tilde{\mathbf{X}}' \\ & - \left( \mathbf{X}\mathbf{R}_2\mathbf{K}'_2 \otimes \mathbf{1}'_{p_1}\mathbf{A}_1 + \tilde{\mathbf{X}}\tilde{\mathbf{R}}_2\tilde{\mathbf{K}}'_2 \otimes \mathbf{1}'_{p_2}\mathbf{A}_2 \right) \mathbf{F}^{-1} \text{vec} \left( \sigma_{E_1}^{-2} \mathbf{A}'_1\mathbf{Y}\mathbf{R}_2\mathbf{K}'_2 + \sigma_{E_2}^{-2} \mathbf{A}'_2\tilde{\mathbf{Y}}\tilde{\mathbf{R}}_2\tilde{\mathbf{K}}'_2 \right) \end{aligned} \right] \\ \text{vec}\hat{\mathbf{B}} &= \mathbf{F}^{-1} \left( \text{vec} \left( \sigma_{E_1}^{-2} \mathbf{A}'_1\mathbf{Y}\mathbf{R}_2\mathbf{K}'_2 + \sigma_{E_2}^{-2} \mathbf{A}'_2\tilde{\mathbf{Y}}\tilde{\mathbf{R}}_2\tilde{\mathbf{K}}'_2 \right) - \left( \sigma_{E_1}^{-2} \mathbf{K}_2\mathbf{R}'_2 \otimes \mathbf{1}_{p_1} + \sigma_{E_2}^{-2} \tilde{\mathbf{K}}_2\tilde{\mathbf{R}}'_2 \otimes \mathbf{1}_{p_2} \right) \hat{\gamma} \right), \end{aligned}$$

where

$$\begin{aligned} \mathbf{P} &= \sigma_{E_1}^{-2} \left( \mathbf{C}'^o(\mathbf{C}'^o)' + \mathbf{R}_2\mathbf{R}_2' \right) \\ \tilde{\mathbf{P}} &= \sigma_{E_2}^{-2} \left( \tilde{\mathbf{C}}'^o(\tilde{\mathbf{C}}'^o)' + \tilde{\mathbf{R}}_2\tilde{\mathbf{R}}_2' \right). \end{aligned}$$

Furthermore, the estimator for  $\Sigma_{11}^u$  obtained by maximizing the likelihood density function  $L(\mathbf{V}_1; \mathbf{B}, \gamma, \Psi_{11})$  with respect to  $\Psi_{11}$  can be expressed by

$$\hat{\Sigma}_{11}^u = \frac{1}{m} \left( \mathbf{Y}\mathbf{R}_1 - \mathbf{A}_1\hat{\mathbf{B}}\mathbf{K}_1 - \mathbf{1}_{p_1}\hat{\gamma}'\mathbf{X}\mathbf{R}_1 \right) \left( \mathbf{Y}\mathbf{R}_1 - \mathbf{A}_1\hat{\mathbf{B}}\mathbf{K}_1 - \mathbf{1}_{p_1}\hat{\gamma}'\mathbf{X}\mathbf{R}_1 \right)' - \sigma_{e_1}^2 \mathbf{I}_{p_1}.$$

It follows that the estimators for  $\Theta$  and  $\Psi_{2\bullet 1}$  obtained by maximizing

$$L(\tilde{\mathbf{V}}_1 | \mathbf{V}_1; \mathbf{B}, \gamma, \Psi_{2\bullet 1}, \Theta) \times L(\mathbf{V}_1; \mathbf{B}, \gamma, \Psi_{11}),$$

are given by

$$\begin{aligned}\hat{\Theta} &= \frac{1}{m} \left( \tilde{Y}\tilde{R}_1 - A_2\hat{B}\tilde{K}_1 - \mathbf{1}_{p_2}\hat{\gamma}'\tilde{X}\tilde{R}_1 \right) \left( YR_1 - A_1\hat{B}K_1 - \mathbf{1}_{p_1}\hat{\gamma}'XR_1 \right)' \Psi_{11}^{-1}, \\ \hat{\Psi}_{2\bullet 1} &= \frac{1}{m} \left( \tilde{Y}\tilde{R}_1 - A_2\hat{B}\tilde{K}_1 - \mathbf{1}_{p_2}\hat{\gamma}'\tilde{X}\tilde{R}_1 - \hat{\Theta}(YR_1 - A_1\hat{B}K_1 - \mathbf{1}_{p_1}\hat{\gamma}'XR_1) \right), \\ &\quad \times \left( \tilde{Y}\tilde{R}_1 - A_2\hat{B}\tilde{K}_1 - \mathbf{1}_{p_2}\hat{\gamma}'\tilde{X}\tilde{R}_1 - \hat{\Theta}(YR_1 - A_1\hat{B}K_1 - \mathbf{1}_{p_1}\hat{\gamma}'XR_1) \right)'.\end{aligned}$$

Hence, the covariances  $\Sigma_{21}^u$  and  $\Sigma_{22}^u$  can be estimated by

$$\begin{aligned}\hat{\Sigma}_{21}^u &= \frac{1}{m} \left( \tilde{Y}\tilde{R}_1 - A_2\hat{B}\tilde{K}_1 - \mathbf{1}_{p_2}\hat{\gamma}'\tilde{X}\tilde{R}_1 \right) \left( YR_1 - A_1\hat{B}K_1 - \mathbf{1}_{p_1}\hat{\gamma}'XR_1 \right)' \\ \hat{\Sigma}_{22}^u &= \hat{\Psi}_{2\bullet 1} - \sigma_{E_2}^2 I_{p_2} + \hat{\Theta}\hat{\Psi}_{11}\hat{\Theta}'.\end{aligned}$$

## 5.2 Prediction

In order to perform predictions of small area means we first have to predict  $U_1$  and  $U_2$  in the model given by (2) and (3). Put

$$\mathbf{y} = \begin{pmatrix} \text{vec}Y \\ \text{vec}\tilde{Y} \end{pmatrix} \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} \text{vec}U_1 \\ \text{vec}U_2 \end{pmatrix}.$$

Following Henderson's prediction approach to linear mixed model (Henderson, 1975), the prediction of  $\mathbf{v}$  can be derived in a two stages, where in at the first stage  $\Sigma^u$  is supposed to be known. Thus the plan is to maximize the joint density of

$$\begin{aligned}f(\mathbf{y}, \mathbf{v}) &= f(\mathbf{y} | \mathbf{v})f(\mathbf{v}) \\ &= c \exp \left\{ \frac{1}{2} \text{tr} \left\{ (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) + \mathbf{v}' \Omega^{-1} \mathbf{v} \right\} \right\},\end{aligned}\quad (5)$$

with respect to  $\text{vec}B$ ,  $\gamma$  and  $\mathbf{v}$ , where  $c$  is a known constant and  $\Omega$  is given by

$$\Omega = \begin{pmatrix} I \otimes \Sigma_{11}^u & I \otimes \Sigma_{12}^u \\ I \otimes \Sigma_{21}^u & I \otimes \Sigma_{22}^u \end{pmatrix}.$$

The vector  $\boldsymbol{\mu}$  and the matrix  $\Sigma$  are the expectation and dispersion of  $\mathbf{y} | \mathbf{v}$  and are given by

$$E[\mathbf{y} | \mathbf{v}] = \boldsymbol{\mu} = H_1 \text{vec}B + H_2 \gamma + H_3 \mathbf{v},$$

where

$$H_1 = \begin{pmatrix} C'H' \otimes A_1 \\ \tilde{C}'H' \otimes A_2 \end{pmatrix}, \quad H_2 = \begin{pmatrix} X' \otimes \mathbf{1}_{p_1} \\ \tilde{X}' \otimes \mathbf{1}_{p_2} \end{pmatrix}, \quad H_3 = \begin{pmatrix} Z' \otimes I \\ \tilde{Z}' \otimes I \end{pmatrix},$$

and

$$D[\mathbf{y} | \mathbf{v}] = \Sigma = \begin{pmatrix} \sigma_{E_1}^2 I_{p_1 n} & \mathbf{0} \\ \mathbf{0} & \sigma_{E_2}^2 I_{p_2 n_1} \end{pmatrix}.$$

At the second stage, using (5) we find  $\widehat{\mathbf{v}}$  and hence  $\widehat{\mathbf{U}}$  via standard calculations and replacement of  $\boldsymbol{\Sigma}^u$  by its estimator which is obtained as described in Section 5.1.

The prediction of small area means is performed under the superpopulation model approach to finite population in the sense that estimating the small area means is equivalent to predicting small area means of non sampled values, given the sample data and auxiliary data. To this end, for each  $i$ -th area and each  $g$ -th group units, we consider the means for sample observations of the data matrices  $\mathbf{Y}$  and  $\widetilde{\mathbf{Y}}$  and predict the means of non sampled values.

The target small area means at each time point are elements of the vectors

$$\widehat{\boldsymbol{\mu}}_i = \frac{1}{N_i} \left( \widehat{\boldsymbol{\mu}}_i^{(s)} + \left( \mathbf{A}\widehat{\mathbf{B}}\mathbf{C}_i^{(r)} + \mathbf{1}_p\widehat{\boldsymbol{\gamma}}'\mathbf{X}_i^{(r)} + \widehat{\mathbf{u}}_i\mathbf{z}_i^{(r)'} \right) \mathbf{1}_{N_i-n_i} \right), \quad i = 1, \dots, m,$$

where

$$\widehat{\boldsymbol{\mu}}_i^{(s)} = \begin{pmatrix} \mathbf{Y}_i^{(s)} \mathbf{1}_{n_i} \\ \widetilde{\mathbf{Y}}_i^{(s)} \mathbf{1}_{n_i} \end{pmatrix}.$$

The small area means at each time point for each group units for complete and incomplete data sets and are given by

$$\widehat{\boldsymbol{\mu}}_{ig} = \frac{1}{N_{ig}} \left( \widehat{\boldsymbol{\mu}}_{ig}^{(s)} + \left( \mathbf{A}\widehat{\boldsymbol{\beta}}_g \mathbf{1}'_{N_{ig}-n_{ig}} + \mathbf{1}_p\widehat{\boldsymbol{\gamma}}'\mathbf{X}_{ig}^{(r)} + \widehat{\mathbf{u}}_i\mathbf{z}_{ig}^{(r)'} \right) \mathbf{1}_{N_{ig}-n_{ig}} \right),$$

$$i = 1, \dots, m, \quad g = 1, \dots, k,$$

where

$$\widehat{\boldsymbol{\mu}}_{ig}^{(s)} = \begin{pmatrix} \mathbf{Y}_{ig}^{(s)} \mathbf{1}_{n_{ig}} \\ \widetilde{\mathbf{Y}}_{ig}^{(s)} \mathbf{1}_{n_{ig}} \end{pmatrix}.$$

The superscripts  $s$  and  $r$  indicate the corresponding partitions for observed sample data and non observed sample data, respectively. Moreover  $\mathbf{X}_i^{(r)}$ ,  $\mathbf{C}_i^{(r)}$  and  $\mathbf{z}_i^{(r)}$  are the corresponding matrix of auxiliary information, design matrix and design vector for non sampled units, respectively. Note that the predicted vector  $\widehat{\mathbf{u}}_i$  is the  $i$ -th column of the predicted matrix  $\widehat{\mathbf{U}}$  and  $\widehat{\boldsymbol{\beta}}_g$  is the column of the estimated parameter matrix  $\widehat{\mathbf{B}}$  for the corresponding group  $g$ .

Depending on the type of the characteristics of interest, the target small area means for each group across all time points are obtained as a linear combination of  $\widehat{\boldsymbol{\mu}}_{ig}$ .

## References

- Carriere, K. (1999). Methods for repeated measures data analysis with missing values. *Journal of Statistical Planning and Inference*, 77(2):221–236.

- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2):423–447.
- Kim, K. and Timm, N. (2006). *Univariate and Multivariate general linear Models: Theory and Applications with SAS*. CRC Press.
- Kleinbaum, D. G. (1973). A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis*, 3(1):117–124.
- Liski, E. P. (1985). Estimation from incomplete data in growth curves models. *Communications in Statistics - Simulation and Computation*, 14(1):13–27.
- Liski, E. P. and Nummi, T. (1990). Prediction in growth curve models using the EM algorithm. *Computational Statistics and Data Analysis*, 10(2):99–108.
- Longford, N. T. (2006). *Missing data and small-area estimation: Modern analytical equipment for the survey statistician*. Springer Science & Business Media, New York.
- Ngaruye, I., Nzabanita, J., von Rosen, D., and Singull, M. (2016a). Small area estimation under a multivariate linear model for repeated measures data. *Communications in Statistics - Theory and Methods*. <http://dx.doi.org/10.1080/03610926.2016.1248784>.
- Ngaruye, I., von Rosen, D., and Singull, M. (2016b). Crop yield estimation at district level for agricultural seasons 2014 in Rwanda. *African Journal of Applied Statistics*, 3(1):69–90.
- Nummi, T. (1997). Estimation in a random effects growth curve model. *Journal of Applied Statistics*, 24(2):157–168.
- Pfeffermann, D. (2002). Small area estimation-new developments and directions. *International Statistical Review*, 70(1):125–143.
- Rao, J. N. K. (2003). *Small Area Estimation*. John Wiley and Sons, New York.
- Srivastava, M. (1985). Multivariate data with missing observations. *Communications in Statistics - Theory and Methods*, 14(4):775–792.
- Srivastava, M. S. (2002). *Methods of Multivariate Statistics*. Wiley-Interscience New York.