Department of Mathematics

Small area estimation with missing data using a multivariate linear random effects model

Innocent Ngaruye, Dietrich von Rosen and Martin Singull

LiTH-MAT-R--2017/07--SE

LINKÖPING UNIVERSITY

Department of Mathematics
Linköping University
S-581 83 Linköping

Small area estimation with missing data using a multivariate linear random effects model

Innocent Ngaruye [1,2], Dietrich von Rosen[1,3]
and Martin Singull[1]

[1]Department of Mathematics, Linköping University,
SE–581 83 Linköping, Sweden
E-mail: {innocent.ngaruye, martin.singull}@liu.se

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

[3]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 article small area estimation with multivariate data that follow a monotonic missing sample pattern is addressed. Random effects growth curve models with covariates are formulated. A likelihood based approach is proposed for estimation of the unknown parameters. Moreover, the prediction of random effects and predicted small area means are also discussed.

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

1 Introduction

In survey analysis estimation of characteristics of interest for subpopulations (also called domains or small areas) for which sample sizes are small is challenging. We adopt an approach were the survey estimates are improved via covariate information. To produce reliable estimates in surveys utilizing covariates for small areas 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. There are also some works related to time series and longitudinal surveys in small area estimation, for example, one can refer to Consortium (2004); Ferrante and Pacei (2004); Nissinen (2009); Singh and Sisodia (2011); Ngaruye et al. (2016). In Ngaruye et al. (2016), the authors have proposed a multivariate linear model for repeated measures data in a SAE context. The model is a combination of the classical growth curve model (Potthoff and Roy, 1964) with a random effects model. This model accounts for longitudinal surveys, i.e. units are sampled ones and then followed in time, grouped response units and time correlated random effects. Commonly incomplete repeated measures data are obtained. In this article we extend the above mentioned model and let the model include a monotonic missing observation structure. In particular drop-outs from the survey can be handled, i.e. when it is planned to follow units in time but before the end-point some units disappear.

Missing data may be due to a number of limitations such as unexpected budget constraints, but also it may happen that for various reasons units for which the measurements were expected to be sampled over time disappeared from the survey. The statistical analysis of data with missing values emerged early in 1970s with advancement of modern computer based technology (Little and Rubin, 1987). Since then, several methods of analysis of missing data have been developed following the missing data mechanism whether ignorable for inferences which includes missing data at random and missing data completely at random or nonignorable missing data. Many authors have dealt with the problem of missing data and we can refer to Little and Rubin (1987); Carriere (1999); Srivastava (2002); Kim and Timm (2006); Longford (2006), for example. In particular, incomplete data in the classical growth curve models and in random effects growth curve model has been considered, for example, by Kleinbaum (1973); Woolson and Leeper (1980); Srivastava (1985); Liski (1985); Liski and Nummi (1990); Nummi (1997) The missing values are assumed to be independently distributed of the observed values.

In Section 3, we present the formulation of a multivariate linear model for repeated measures data. Thereafter this model is extended to handle missing data. A "canonical" form of the model is considered in Section 4. In Section 5, the estimation of parameters and prediction of random effects and small area means are derived.

2 Multivariate linear model for repeated measures data

We will in this section consider the multivariate linear regression model for repeated measurements with covariates at p time points suitable for discussing the SAE problem, which was defined by Ngaruye et al. (2016), when data are complete. It is supposed that the target population of size N whose characteristic of interest y is divided into m subpopulations called small areas of sizes N_d, $d = 1, .., m$, and the units in all small areas are grouped in k different categories. Furthermore, we assume the mean growth of each unit in area d for each one of the k groups to be, for example, a polynomial in time with degree $q-1$ and also suppose that we have covariate variables related to the characteristic of interest whose values are available for all units in the population. Out of the whole population N and small areas N_d, n and n_d "units" are sampled according to some sampling scheme which however technically in the present work is of no interest. The model at small area level for the sampled units is written

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

and when combining all disjoint m small areas and all n sampled units divided into k non-overlapping group units yields

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

where $\boldsymbol{\Sigma}^u$ is an unknown arbitrary positive definite matrix and $\boldsymbol{\Sigma}_e = \sigma_e^2\boldsymbol{I}_p$ is assumed to be known. In practise σ_e^2 is estimated from the survey and only depends on how many units are sampled from the total population N. In model (1), $\boldsymbol{Y} : p \times n$ is the data matrix, $\boldsymbol{A} : p \times q$, $q \leq p$, is the within individual design matrix indicating the time dependency within individuals, $\boldsymbol{B} : q \times k$ is unknown parameter matrix, $\boldsymbol{C} : mk \times n$ with $\text{rank}(\boldsymbol{C}) + p \leq n$ and $p \leq m$ is the between individuals design matrix accounting for group effects, $\boldsymbol{\gamma}$ is an r-vector of fixed regression coefficients representing the effects of auxiliary variables, $\boldsymbol{X} : r \times n$ is a known matrix taking the values of the covariates, the matrix $\boldsymbol{U} : p \times m$ 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., $\boldsymbol{U} \sim \mathcal{N}_{p,m}(\mathbf{0}, \boldsymbol{\Sigma}^u, \boldsymbol{I}_m)$, $\boldsymbol{Z} : m \times n$ is a design matrix for random effect and the columns of the error matrix $\boldsymbol{E}$ are assumed to be independently distributed as p-variate normal distribution with mean zero and and known covariance matrix $\boldsymbol{\Sigma}_e$, i.e., $\boldsymbol{E} \sim N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}_e, \boldsymbol{I}_n)$. More details about model formulation and estimation of model parameters can be found in Ngaruye et al. (2016).

3 Incomplete data

Consider model (1) and suppose that there are missing values in such a way that the measurements taken at time t, (for $t = 1, ..., p$), on each unit are not all complete and the number of observations for the different p time points are $n_1, ..., n_p$, with $n_1 \geq n_2 \geq ... \geq n_p > p$. 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 i-th set, $(i = 1, ..., h)$, is the sample data matrix $\boldsymbol{Y}_i : p_i \times n_i$ whose units in the sample have completed $i - 1$ periods and failed to complete the ith period with $p_i \leq p$ and $\sum_{i=1}^{h} p_i = p$. For technical simplicity, in this paper we only study a three-step monotone missing structure with complete sample data for a given number of time points and incomplete sample data for the other time points.

3.1 The model which handles missing data

In this article we will only present details for a three-step monotonic pattern. We assume that the model, defined in (1), holds together with a monotonic missing structure. This extended model can be presented by three equations:

$$\boldsymbol{Y_1} = \boldsymbol{A}_1\boldsymbol{BHC_1} + \mathbf{1}_{p_1}\boldsymbol{\gamma}'\boldsymbol{X_1} + \boldsymbol{U}_1\boldsymbol{Z_1} + \boldsymbol{E}_1, \tag{2}$$

$$\boldsymbol{Y_2} = \boldsymbol{A}_2\boldsymbol{BHC_2} + \mathbf{1}_{p_2}\boldsymbol{\gamma}'\boldsymbol{X_2} + \boldsymbol{U}_2\boldsymbol{Z_2} + \boldsymbol{E}_2, \tag{3}$$

$$\boldsymbol{Y_3} = \boldsymbol{A}_3\boldsymbol{BHC_3} + \mathbf{1}_{p_3}\boldsymbol{\gamma}'\boldsymbol{X_3} + \boldsymbol{U}_3\boldsymbol{Z_3} + \boldsymbol{E}_3, \tag{4}$$

where $\boldsymbol{A}' = (\boldsymbol{A}_1' : \boldsymbol{A}_2' : \boldsymbol{A}_3')$, $\boldsymbol{A}_i$: $p_i \times q$, $q < p$, $\sum_{i=1}^{3} p_i = p$, $\boldsymbol{H} = (\boldsymbol{I}_k : \boldsymbol{I}_k \ldots \boldsymbol{I}_k)$: $k \times km$,

$$\boldsymbol{C}_i = \begin{pmatrix} \boldsymbol{C}_{i1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{C}_{im} \end{pmatrix}, \qquad \boldsymbol{C}_{id} = \begin{pmatrix} \mathbf{1}'_{n_{id1}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{1}'_{n_{idk}} \end{pmatrix},$$

n_{idg} equals the number of observations for the response $\boldsymbol{Y}_i$, d-th small area and g-th group, $\boldsymbol{X}_i$ represents all covariates for the $\boldsymbol{Y}_i$ response,

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{z}'_{i1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{z}'_{im} \end{pmatrix}, \qquad \boldsymbol{z}_{id} = \frac{1}{\sqrt{n_{id}}}\mathbf{1}_{n_{id}}, \qquad i = 1,2,3, \quad d = 1,2,\dots,m,$$

$\boldsymbol{U}_1 = (\boldsymbol{I}_{p_1} : \mathbf{0} : \mathbf{0})\boldsymbol{U}$, $\boldsymbol{U}_2 = (\mathbf{0} : \boldsymbol{I}_{p_2} : \mathbf{0})\boldsymbol{U}$, $\boldsymbol{U}_3 = (\mathbf{0} : \mathbf{0} : \boldsymbol{I}_{p_3})\boldsymbol{U}$, $\boldsymbol{U} \sim \mathcal{N}_{p,m}\Big(\mathbf{0}, \boldsymbol{\Sigma}^u, \boldsymbol{I}_m\Big)$, $\boldsymbol{E}_i \sim \mathcal{N}_{p_i,n_i}\Big(\mathbf{0}, \boldsymbol{I}_{p_i}, \sigma_i^2\boldsymbol{I}_{n_i}\Big)$, $\{\boldsymbol{E}_i\}$ are mutually independent and $\boldsymbol{E}_i$ is independent of $\boldsymbol{U}_i$. In particular the construction of $\boldsymbol{Z}_i$ helps to derive a number of mathematical results including

$$\mathcal{C}(\boldsymbol{Z}'_i) \subseteq \mathcal{C}(\boldsymbol{C}'_i), \qquad \boldsymbol{Z}_i\boldsymbol{Z}'_i = \boldsymbol{I}_m, \tag{5}$$

where $\mathcal{C}(\boldsymbol{Q})$ stands for the column vector space generated by the columns of the matrix $\boldsymbol{Q}$.

3.2 A canonical version of the model

The model defined through (2), (3) and (4) will be transmitted to a simpler model which will be utilized when estimating the unknown parameters. A couple of definitions will be necessary to introduce but first it is noted that because $\mathcal{C}(\boldsymbol{Z}'_i) \subseteq \mathcal{C}(\boldsymbol{C}'_i)$

$$(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1/2}\boldsymbol{C}_i\boldsymbol{Z}'_i\boldsymbol{Z}_i\boldsymbol{C}'_i(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1/2}, \quad i = 1,2,3,$$

are idempotent. It is supposed that we have so many observations that the inverses exist. Therefore there exists an orthogonal matrix $\boldsymbol{\Gamma}_i = (\boldsymbol{\Gamma}_{i1} : \boldsymbol{\Gamma}_{i2})$, $km \times m$, $km \times (k-1)m$, such that

$$(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1/2}\boldsymbol{C}_i\boldsymbol{Z}'_i\boldsymbol{Z}_i\boldsymbol{C}'_i(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1/2} = \boldsymbol{\Gamma}_i \begin{pmatrix} \boldsymbol{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{\Gamma}'_i = \boldsymbol{\Gamma}_{i1}\boldsymbol{\Gamma}'_{i1}, \quad i = 1,2,3.$$

Moreover, $\boldsymbol{\Gamma}'_{i1}\boldsymbol{\Gamma}_{i1} = \boldsymbol{I}_m$. Put

$$\begin{aligned} \boldsymbol{K}_{ij} &= \boldsymbol{H}(\boldsymbol{C}_i\boldsymbol{C}'_i)^{1/2}\boldsymbol{\Gamma}_{ij}, \qquad i = 1,2,3, \quad j = 1,2, \\ \boldsymbol{R}_{ij} &= \boldsymbol{C}'_i(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1/2}\boldsymbol{\Gamma}_{ij}, \qquad i = 1,2,3, \quad j = 1,2, \end{aligned} \tag{6}$$

and let $\boldsymbol{Q}^o$ be any matrix of full rank spanning $\mathcal{C}(\boldsymbol{Q})^\perp$, the orthogonal complement to $\mathcal{C}(\boldsymbol{Q})$. The following transformations of $\boldsymbol{Y}_i$, $i = 1,2,3$, is made

$$\boldsymbol{V}_{i0} = \boldsymbol{Y}_i(\boldsymbol{C}'_i)^o = \mathbf{1}_{p_i}\boldsymbol{\gamma}'\boldsymbol{X}_i(\boldsymbol{C}'_i)^o + \boldsymbol{E}_i(\boldsymbol{C}'_i)^o, \quad i = 1,2,3, \tag{7}$$

$$\boldsymbol{V}_{i1} = \boldsymbol{Y}_i\boldsymbol{R}_{i1} = \boldsymbol{A}_i\boldsymbol{B}\boldsymbol{K}_{i1} + \mathbf{1}_{p_i}\boldsymbol{\gamma}'\boldsymbol{X}_i\boldsymbol{R}_{i1} + (\boldsymbol{U}_i\boldsymbol{Z}_i + \boldsymbol{E}_i)\boldsymbol{R}_{i1}, \quad i = 1,2,3, \tag{8}$$

$$\boldsymbol{V}_{i2} = \boldsymbol{Y}_i\boldsymbol{R}_{i2} = \boldsymbol{A}_i\boldsymbol{B}\boldsymbol{K}_{i2} + \mathbf{1}_{p_i}\boldsymbol{\gamma}'\boldsymbol{X}_i\boldsymbol{R}_{i2} + \boldsymbol{E}_i\boldsymbol{R}_{i2}, \quad i = 1,2,3. \tag{9}$$

3.3 The likelihood

The transformation which has taken place in the previous section is one-to-one. Based on $\{\boldsymbol{V}_{ij}\}$, $i = 1, 2, 3$, $j = 0, 1, 2$, we will set up the likelihood for all observations. However, firstly we present the marginal densities (likelihood function) for $\{\boldsymbol{V}_{ij}\}$, which of course are normally distributed. Thus, to determine the distributions it is enough to present means and dispersion matrices:

$$E[\boldsymbol{V}_{i0}] = \mathbf{1}_{p_i}\boldsymbol{\gamma}'\boldsymbol{X}_i(\boldsymbol{C}_i')^o, \qquad D[\boldsymbol{V}_{i0}] = \sigma_i^2(\boldsymbol{C}_i')^{o'}(\boldsymbol{C}_i')^o, \qquad i = 1, 2, 3, \tag{10}$$

$$E[\boldsymbol{V}_{i1}] = \boldsymbol{A}_i\boldsymbol{B}\boldsymbol{K}_{i1} + \mathbf{1}_{p_i}\boldsymbol{\gamma}'\boldsymbol{X}_i\boldsymbol{R}_{i1}, \qquad D[\boldsymbol{V}_{i1}] = \boldsymbol{R}_{i1}'\boldsymbol{Z}_1'\boldsymbol{Z}_1\boldsymbol{R}_{i1} \otimes \boldsymbol{\Sigma}_{ii}^u + \sigma_i^2\boldsymbol{R}_{i1}'\boldsymbol{R}_{i1} \otimes \boldsymbol{I},$$
$$i = 1, 2, 3, \tag{11}$$

$$E[\boldsymbol{V}_{i2}] = \boldsymbol{A}_i\boldsymbol{B}\boldsymbol{K}_{i2} + \mathbf{1}_{p_i}\boldsymbol{\gamma}'\boldsymbol{X}_i\boldsymbol{R}_{i2}, \qquad D[\boldsymbol{V}_{i2}] = \boldsymbol{R}_{i2}'\boldsymbol{R}_{i2} \otimes \boldsymbol{I}, \qquad i = 1, 2, 3, \tag{12}$$

Concerning the simultaneous distribution of $\{\boldsymbol{V}_{ij}\}$, $i = 1, 2, 3$, $j = 0, 1, 2$, $\boldsymbol{V}_{i0}$ and $\boldsymbol{V}_{i2}$, $i = 1, 2, 3$, are independently distributed and these variables are also independent of $\{\boldsymbol{V}_{i1}\}$. However, the elements in $\{\boldsymbol{V}_{i1}\}$, are not independently distributed. We have to pay attention to the likelihood of these variables and $\{\mathrm{vec}\boldsymbol{V}_{i1}\}$, $i = 1, 2, 3$, will be considered.

Let $L(\boldsymbol{V}; \boldsymbol{\Theta})$ denote the likelihood function for the random variable $\boldsymbol{V}$ with parameter $\boldsymbol{\Theta}$. We are going to discuss

$$L(\mathrm{vec}\boldsymbol{V}_{31}, \mathrm{vec}\boldsymbol{V}_{21}, \mathrm{vec}\boldsymbol{V}_{11}; \bullet)$$
$$= L(\mathrm{vec}\boldsymbol{V}_{31}|\mathrm{vec}\boldsymbol{V}_{21}, \mathrm{vec}\boldsymbol{V}_{11}; \bullet)L(\mathrm{vec}\boldsymbol{V}_{21}|\mathrm{vec}\boldsymbol{V}_{11}; \bullet)L(\mathrm{vec}\boldsymbol{V}_{11}; \bullet), \tag{13}$$

where in (13) $\bullet$ indicates that no parameters have been specified. Before obtaining some useful results we need a few technical relations concerning $\boldsymbol{Z}_i$, $i = 1, 2, 3$. To some extent the next lemma is our main contribution because without it the mathematics would become very difficult to carry out. Note that the result depends on the definition of $\boldsymbol{Z}_i$, $i = 1, 2, 3$.

Lemma 3.1. *Let $\boldsymbol{Z}_i$, $i = 1, 2, 3$, be as in (2), (3) and (4), and let $\boldsymbol{R}_{i1}$, $i = 1, 2, 3$, be defined in (6). Then*

(i) $\boldsymbol{Z}_i\boldsymbol{R}_{i1}\boldsymbol{R}_{i1}'\boldsymbol{Z}_i' = \boldsymbol{I}_m$;

(ii) $\boldsymbol{R}_{i1}'\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{R}_{i1} = \boldsymbol{I}_m$.

Proof. Using (6), (5) and the definition of $\mathbf{\Gamma}_{i1}$ it follows that

$$\begin{aligned}\boldsymbol{Z}_i\boldsymbol{R}_{i1}\boldsymbol{R}_{i1}'\boldsymbol{Z}_i' &= \boldsymbol{Z}_i\boldsymbol{C}_i'(\boldsymbol{C}_i\boldsymbol{C}_i')^{-1/2}\mathbf{\Gamma}_{i1}\mathbf{\Gamma}_{i1}'(\boldsymbol{C}_i\boldsymbol{C}_i')^{-1/2}\boldsymbol{C}_i\boldsymbol{Z}_i'\\ &= \boldsymbol{Z}_i\boldsymbol{P}_{C_i}\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{P}_{C_i}\boldsymbol{Z}_i' = \boldsymbol{Z}_i\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{Z}_i' = \boldsymbol{I}_m,\end{aligned}$$

where $\boldsymbol{P}_{C_i} = \boldsymbol{C}_i(\boldsymbol{C}_i'\boldsymbol{C}_i)^{-1}\boldsymbol{C}_i'$ is the unique orthogonal projection on $\mathcal{C}(\boldsymbol{C}_i)$, and thus statement (i) is established. Moreover, once again using (6) and the definition of $\mathbf{\Gamma}_{i1}$

$$\begin{aligned}\boldsymbol{R}_{i1}'\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{R}_{i1} &= \mathbf{\Gamma}_{i1}(\boldsymbol{C}_i\boldsymbol{C}_i')^{-1/2}\boldsymbol{C}_i\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{C}_i'(\boldsymbol{C}_i\boldsymbol{C}_i')^{-1/2}\mathbf{\Gamma}_{i1}\\ &= \mathbf{\Gamma}_{i1}'\mathbf{\Gamma}_{i1}\mathbf{\Gamma}_{i1}'\mathbf{\Gamma}_{i1} = \boldsymbol{I}_m,\end{aligned}$$

and statement (ii) is verified. □

The next result will be used in the forthcoming presentation:

$$D\begin{bmatrix}\mathrm{vec}\boldsymbol{V}_{11}\\ \mathrm{vec}\boldsymbol{V}_{21}\\ \mathrm{vec}\boldsymbol{V}_{31}\end{bmatrix} = \left(\boldsymbol{R}_{i1}'\boldsymbol{Z}_i'\boldsymbol{Z}_j\boldsymbol{R}_{j1}\otimes\mathbf{\Sigma}_{ij}^u\right)_{i=1,2,3;j=1,2,3} + \mathrm{diag}(\boldsymbol{R}_{i1}'\boldsymbol{R}_{i1}\otimes\sigma_i^2\boldsymbol{I}_m), \tag{14}$$

where $\left(\bullet\right)_{i=1,2,3;j=1,2,3}$ denotes a block partitioned matrix and diag($\bullet$) operates as follows:

$$\mathrm{diag}(\boldsymbol{Q}_{ii}) = \begin{pmatrix}\boldsymbol{Q}_{11} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{Q}_{22} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{Q}_{33}\end{pmatrix},$$

which is obtained by straight forward calculations. Note that Lemma 3.1 together with the fact $\boldsymbol{R}_{i1}'\boldsymbol{R}_{i1} = \boldsymbol{I}$ yield that the variances in (14) are of the form

$$\boldsymbol{I}\otimes(\mathbf{\Sigma}_{ii}^u + \sigma_i^2\boldsymbol{I}_{p_i}), \qquad i = 1,2,3,$$

which is an important result.

From the factorization of the likelihood in (13) it follows that we have to investigate

$$L(\mathrm{vec}\boldsymbol{V}_{31}|\mathrm{vec}\boldsymbol{V}_{21}, \mathrm{vec}\boldsymbol{V}_{11}; \bullet).$$

Thus we are interested in the conditional expectation and the conditional dispersion. The conditional mean equals

$$\begin{aligned}&E[\mathrm{vec}\boldsymbol{V}_{31}|\mathrm{vec}\boldsymbol{V}_{11}, \mathrm{vec}\boldsymbol{V}_{21}]\\ &\quad= E[\mathrm{vec}\boldsymbol{V}_{31}] + (C[\boldsymbol{V}_{31}, \boldsymbol{V}_{11}], C[\boldsymbol{V}_{31}, \boldsymbol{V}_{21}])D[(\mathrm{vec}'\boldsymbol{V}_{11}, \mathrm{vec}'\boldsymbol{V}_{21})']^{-1}\\ &\qquad\times((\mathrm{vec}'\boldsymbol{V}_{11}, \mathrm{vec}'\boldsymbol{V}_{21})' - (E[\mathrm{vec}'\boldsymbol{V}_{11}], E[\mathrm{vec}'\boldsymbol{V}_{21}])'),\end{aligned}$$

where the expectations for $\mathrm{vec}\boldsymbol{V}_{i1}$, $i = 1, 2, 3$ can be obtained from (11). Moreover, the conditional dispersion is given by

$$D[\mathrm{vec}\boldsymbol{V}_{31}|\mathrm{vec}\boldsymbol{V}_{11}, \mathrm{vec}\boldsymbol{V}_{21}] = D[\boldsymbol{V}_{31}]$$
$$-(C[\boldsymbol{V}_{31}, \boldsymbol{V}_{11}], C[\boldsymbol{V}_{31}, \boldsymbol{V}_{21}])D[(\mathrm{vec}'\boldsymbol{V}_{11}, \mathrm{vec}'\boldsymbol{V}_{21})']^{-1}(C[\boldsymbol{V}_{31}, \boldsymbol{V}_{11}], C[\boldsymbol{V}_{31}, \boldsymbol{V}_{21}])'.$$

The next lemma fills in the details of this relation and the conditional mean and indeed shows that relative complicated expressions can be dramatically simplified using Lemma 3.1.

Lemma 3.2. *Let $\boldsymbol{V}_{i1}$, $i = 1, 2, 3$, be defined in (8). Then*

(i) $D[\boldsymbol{V}_{31}] = \boldsymbol{I} \otimes (\boldsymbol{\Sigma}^u_{33} + \sigma^2_3 \boldsymbol{I}_{p_3})$;

(ii) $C[\boldsymbol{V}_{31}, \boldsymbol{V}_{11}] = \boldsymbol{R}'_{31}\boldsymbol{Z}'_3\boldsymbol{Z}_1\boldsymbol{R}_{11} \otimes \boldsymbol{\Sigma}^u_{31}$;

(iii) $C[\boldsymbol{V}_{31}, \boldsymbol{V}_{21}] = \boldsymbol{R}'_{31}\boldsymbol{Z}'_3\boldsymbol{Z}_2\boldsymbol{R}_{21} \otimes \boldsymbol{\Sigma}^u_{32}$;

(iv)

$$D\begin{bmatrix} \mathrm{vec}\boldsymbol{V}_{11} \\ \mathrm{vec}\boldsymbol{V}_{21} \end{bmatrix} = \begin{pmatrix} \boldsymbol{I} \otimes (\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1}) & \boldsymbol{R}'_{11}\boldsymbol{Z}'_1\boldsymbol{Z}_2\boldsymbol{R}_{21} \otimes \boldsymbol{\Sigma}^u_{12} \\ \boldsymbol{R}'_{21}\boldsymbol{Z}'_2\boldsymbol{Z}_1\boldsymbol{R}_{11} \otimes \boldsymbol{\Sigma}^u_{21} & \boldsymbol{I} \otimes (\boldsymbol{\Sigma}^u_{22} + \sigma^2_2 \boldsymbol{I}_{p_2}) \end{pmatrix};$$

(v)

$$D\begin{bmatrix} \mathrm{vec}\boldsymbol{V}_{11} \\ \mathrm{vec}\boldsymbol{V}_{21} \end{bmatrix}^{-1}$$
$$= \begin{pmatrix} \boldsymbol{Q}^{-1}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{Q}^{-1}_{11}\boldsymbol{Q}_{12} \\ \boldsymbol{I} \end{pmatrix} (\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}^{-1}_{11}\boldsymbol{Q}_{12})^{-1}\big(-\boldsymbol{Q}_{21}\boldsymbol{Q}^{-1}_{11} \quad \boldsymbol{I}\big),$$

where

$$\boldsymbol{Q}^{-1}_{11} = \boldsymbol{I} \otimes (\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1},$$
$$\boldsymbol{Q}^{-1}_{11}\boldsymbol{Q}_{12} = \boldsymbol{R}'_{11}\boldsymbol{Z}'_1\boldsymbol{Z}_2\boldsymbol{R}_{21} \otimes (\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1}\boldsymbol{\Sigma}^u_{12},$$
$$\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}^{-1}_{11}\boldsymbol{Q}_{12} = \boldsymbol{I} \otimes (\boldsymbol{\Sigma}^u_{22} + \sigma^2_2 \boldsymbol{I}_{p_2} - \boldsymbol{\Sigma}^u_{21}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1}\boldsymbol{\Sigma}^u_{12});$$

(vi)

$$(C[\boldsymbol{V}_{31}, \boldsymbol{V}_{11}], C[\boldsymbol{V}_{31}, \boldsymbol{V}_{21}])D[(\mathrm{vec}'\boldsymbol{V}_{11}, \mathrm{vec}'\boldsymbol{V}_{21})']^{-1}(C[\boldsymbol{V}_{31}, \boldsymbol{V}_{11}], C[\boldsymbol{V}_{31}, \boldsymbol{V}_{21}])'$$
$$= \boldsymbol{I} \otimes (\boldsymbol{\Sigma}^u_{31}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_m)^{-1}\boldsymbol{\Sigma}^u_{13} + \boldsymbol{\Psi}_{32}\boldsymbol{\Psi}^{-1}_{22}\boldsymbol{\Psi}_{23}),$$

where

$$\begin{aligned}\boldsymbol{\Psi}_{32} &= \boldsymbol{\Psi}'_{23} = \boldsymbol{\Sigma}^u_{32} - \boldsymbol{\Sigma}^u_{31}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1}\boldsymbol{\Sigma}^u_{12},\\ \boldsymbol{\Psi}_{22} &= \boldsymbol{\Sigma}^u_{22} + \sigma^2_2 \boldsymbol{I}_{p_2} - \boldsymbol{\Sigma}^u_{21}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1}\boldsymbol{\Sigma}^u_{12}.\end{aligned}$$

Proof. Statements (i), (ii), (iii) and (iv) follow directly from (14). In (v) the inverse of a partitioned matrix is utilized and (vi) is obtained by straight forward matrix manipulations and application of Lemma 3.1. □

Put

$$\boldsymbol{B}_1 = \boldsymbol{\Sigma}^u_{31}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1}, \tag{15}$$

$$\boldsymbol{B}_2 = \boldsymbol{\Sigma}^u_{32}\boldsymbol{\Psi}^{-1}_{22}, \tag{16}$$

$$\boldsymbol{\Psi}_{33} = \boldsymbol{\Sigma}^u_{33} - \boldsymbol{\Sigma}^u_{31}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1}\boldsymbol{\Sigma}^u_{13}, \tag{17}$$

where $\boldsymbol{\Psi}_{22}$ is given in Lemma 3.2 and then the next theorem is directly established using Lemma 3.2.

Theorem 3.1. *Let $\boldsymbol{V}_{i1}$, $i = 1, 2, 3$, be defined in (8) and $\boldsymbol{\Psi}_{ij}$, $i, j = 2, 3$, be defined in Lemma 3.2 and (17). Moreover, let $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ be given by (15) and (16), respectively. Then* $\mathrm{vec}\boldsymbol{V}_{31}|\mathrm{vec}\boldsymbol{V}_{11}, \mathrm{vec}\boldsymbol{V}_{21} \sim N_{p_3 m}(\boldsymbol{M}_{31}, \boldsymbol{D}_{31})$*, where*

$$\begin{aligned}\boldsymbol{M}_{31} = {} & E[\mathrm{vec}\boldsymbol{V}_{31}|\mathrm{vec}\boldsymbol{V}_{11}, \mathrm{vec}\boldsymbol{V}_{21}] = E[\mathrm{vec}\boldsymbol{V}_{31}]\\ & +(\boldsymbol{R}'_{31}\boldsymbol{Z}'_3\boldsymbol{Z}_1\boldsymbol{R}_{11} \otimes \boldsymbol{B}_1(\boldsymbol{I} + \boldsymbol{\Sigma}^u_{12}\boldsymbol{\Psi}^{-1}_{22}\boldsymbol{\Sigma}^u_{21}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1}))\mathrm{vec}(\boldsymbol{V}_{11} - E[\boldsymbol{V}_{11}])\\ & -(\boldsymbol{R}'_{31}\boldsymbol{Z}'_3\boldsymbol{Z}_2\boldsymbol{R}_{21} \otimes \boldsymbol{B}_1\boldsymbol{\Sigma}^u_{12})\mathrm{vec}(\boldsymbol{V}_{21} - E[\boldsymbol{V}_{21}])\\ & +(\boldsymbol{R}'_{31}\boldsymbol{Z}'_3\boldsymbol{Z}_2\boldsymbol{R}_{21} \otimes \boldsymbol{B}_2)\mathrm{vec}(\boldsymbol{V}_{21} - E[\boldsymbol{V}_{21}])\\ & -(\boldsymbol{R}'_{31}\boldsymbol{Z}'_3\boldsymbol{Z}_1\boldsymbol{R}_{11} \otimes \boldsymbol{B}_2\boldsymbol{\Sigma}^u_{21}(\boldsymbol{\Sigma}^u_{11} + \sigma^2_1 \boldsymbol{I}_{p_1})^{-1})\mathrm{vec}(\boldsymbol{V}_{11} - E[\boldsymbol{V}_{11}])\end{aligned}$$

and

$$\boldsymbol{D}_{31} = D[\mathrm{vec}\boldsymbol{V}_{31}|\mathrm{vec}\boldsymbol{V}_{11}, \mathrm{vec}\boldsymbol{V}_{21}] = \boldsymbol{I}_m \otimes \boldsymbol{\Psi}_{3\bullet 2},$$

where

$$\boldsymbol{\Psi}_{3\bullet 2} = \boldsymbol{\Psi}_{33} - \boldsymbol{\Psi}_{32}\boldsymbol{\Psi}^{-1}_{22}\boldsymbol{\Psi}_{23}.$$

The result of the theorem shows that $\mathrm{vec}\boldsymbol{V}_{31}$ given $\mathrm{vec}\boldsymbol{V}_{11}$ and $\mathrm{vec}\boldsymbol{V}_{21}$, and if $E[\mathrm{vec}\boldsymbol{V}_{11}]$, $E[\mathrm{vec}\boldsymbol{V}_{21}]$, $\boldsymbol{\Sigma}^u_{21}$, $\boldsymbol{\Sigma}^u_{11}$ and $\boldsymbol{\Psi}_{22}$ do not depend on unknown parameters the model with unknown mean parameters $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ and unknown dispersion $\boldsymbol{\Psi}_{3\bullet 2}$ is the same as a vectorized MANOVA model (e.g. see Srivastava, 2002, for information about MANOVA).

Moreover, it follows from (13) that

$$L(\text{vec}\boldsymbol{V}_{21}|\text{vec}\boldsymbol{V}_{11};\bullet)$$

is needed. However, the calculations are the same as above and we only present the final result.

Theorem 3.2. *Let $\boldsymbol{V}_{i1}$, $i = 1, 2$, be defined in (8) and $\boldsymbol{\Psi}_{22}$ in Lemma 3.2. Put $\boldsymbol{B}_0 = \boldsymbol{\Sigma}^u_{21}(\boldsymbol{\Sigma}^u_{11} + \sigma_1^2\boldsymbol{I}_{p_1})^{-1}$. Then $\text{vec}\boldsymbol{V}_{21}|\text{vec}\boldsymbol{V}_{11} \sim N_{p_2 m}(\boldsymbol{M}_{21}, \boldsymbol{I}_m \otimes \boldsymbol{\Psi}_{22})$, where*

$$\begin{aligned}\boldsymbol{M}_{21} &= E[\text{vec}\boldsymbol{V}_{21}|\text{vec}\boldsymbol{V}_{11}] = E[\text{vec}\boldsymbol{V}_{21}] \\ &\quad +(\boldsymbol{R}'_{21}\boldsymbol{Z}'_2\boldsymbol{Z}_1\boldsymbol{R}_{11} \otimes \boldsymbol{B}_0)\text{vec}(\boldsymbol{V}_{11} - E[\boldsymbol{V}_{11}]).\end{aligned}$$

Hence, it has been established that $\text{vec}\boldsymbol{V}_{21}|\text{vec}\boldsymbol{V}_{11}$ is a vectorized MANOVA model.

Theorem 3.3. *The likelihood for $\{\boldsymbol{V}_{ij}\}$, $i = 1, 2, 3$, $j = 0, 1, 2$, given in (7), (8) and (9) equals*

$$\begin{aligned}&L(\{\boldsymbol{V}_{ij}\}, i = 1, 2, 3,\ j = 0, 1, 2; \boldsymbol{\gamma}, \boldsymbol{B}, \boldsymbol{\Sigma}^u) = \prod_{i=1}^{3} L(\{\boldsymbol{V}_{i0}\}, i = 1, 2, 3; \boldsymbol{\gamma}) \\ &\times \prod_{i=1}^{3} L(\{\boldsymbol{V}_{i2}\}, i = 1, 2, 3; \boldsymbol{\gamma}, \boldsymbol{B}) \\ &\times L(\text{vec}\boldsymbol{V}_{31}|\text{vec}\boldsymbol{V}_{11}, \text{vec}\boldsymbol{V}_{21}; \boldsymbol{\gamma}, \boldsymbol{B}, \boldsymbol{\Sigma}^u_{33}, \boldsymbol{\Sigma}^u_{22}, \boldsymbol{\Sigma}^u_{12}, \boldsymbol{\Sigma}^u_{11}, \boldsymbol{B}_1, \boldsymbol{B}_2) \\ &\times L(\boldsymbol{V}_{21}|\boldsymbol{V}_{11}; \boldsymbol{\gamma}, \boldsymbol{B}, \boldsymbol{B}_0, \boldsymbol{\Sigma}^u_{22}, \boldsymbol{\Sigma}^u_{11}) L(\boldsymbol{V}_{11}\boldsymbol{\gamma}, \boldsymbol{B}, \boldsymbol{\Sigma}^u_{11}),\end{aligned}$$

where all parameters mentioned in the likelihoods have been defined earlier in Section 3.

4 Estimation of parameters and prediction of small area means

For the monotone missing value problem, treated in the previous sections, it was shown that it is possible to present a model which seems to be easy to utilize. The remaining part of the report consists of a relatively straight forward approach for predicting the small areas which is of concern in this article.

4.1 Estimation

In order to estimate the parameters a restricted likelihood approach is proposed which is described in the next proposition.

Proposition 5.1 For the likelihood given in Theorem 3.3 $\boldsymbol{B}$ and $\boldsymbol{\gamma}$ are estimated by maximizing

$$\prod_{i=1}^{3} L(\{\boldsymbol{V}_{i0}\},\, i=1,2,3;\boldsymbol{\gamma})\prod_{i=1}^{3} L(\{\boldsymbol{V}_{i2}\},\, i=1,2,3;\boldsymbol{\gamma},\boldsymbol{B}).$$

Inserting these estimators in

$$L(\boldsymbol{V}_{21}|\boldsymbol{V}_{11};\boldsymbol{\gamma},\boldsymbol{B},\boldsymbol{B}_0,\boldsymbol{\Sigma}^u_{22},\boldsymbol{\Sigma}^u_{11})L(\boldsymbol{V}_{11}\boldsymbol{\gamma},\boldsymbol{B},\boldsymbol{\Sigma}^u_{11}),$$

and thereafter maximizing the likelihoods with respect to the remaining unknown parameters produces estimators for $\boldsymbol{\Sigma}^u_{11}$, $\boldsymbol{\Sigma}^u_{12}$ and $\boldsymbol{\Sigma}^u_{22}$. Inserting all the obtained estimators in

$$L(\text{vec}\boldsymbol{V}_{31}|\text{vec}\boldsymbol{V}_{11},\text{vec}\boldsymbol{V}_{21};\boldsymbol{\gamma},\boldsymbol{B},\boldsymbol{\Sigma}^u_{33},\boldsymbol{\Sigma}^u_{22},\boldsymbol{\Sigma}^u_{12},\boldsymbol{\Sigma}^u_{11},\boldsymbol{B}_1,\boldsymbol{B}_2)$$

and then maximizing the likelihood with respect to $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ and $\boldsymbol{\Psi}_{33}-\boldsymbol{\Psi}_{32}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{23}$ yields estimators for $\boldsymbol{\Sigma}^u_{31}$, $\boldsymbol{\Sigma}^u_{32}$ and $\boldsymbol{\Sigma}^u_{33}$.

4.2 Prediction

In order to perform predictions of small area means we first have to predict $\boldsymbol{U}_1$, $\boldsymbol{U}_2$ and $\boldsymbol{U}_3$ in the model given by (2), (3) and (4). Put

$$\boldsymbol{y}=\begin{pmatrix}\text{vec}\boldsymbol{Y}_1\\ \text{vec}\boldsymbol{Y}_2\\ \text{vec}\boldsymbol{Y}_3\end{pmatrix}\quad\text{and }\boldsymbol{v}=\begin{pmatrix}\text{vec}\boldsymbol{U}_1\\ \text{vec}\boldsymbol{U}_2\\ \text{vec}\boldsymbol{U}_3\end{pmatrix}.$$

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

$$\begin{aligned} f(\boldsymbol{y},\boldsymbol{v}) =& f(\boldsymbol{y}\mid\boldsymbol{v})f(\boldsymbol{v})\\ =& c\exp\left\{-\frac{1}{2}\text{tr}\Big\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})+\boldsymbol{v}'\boldsymbol{\Omega}^{-1}\boldsymbol{v}\Big\}\right\}, \end{aligned} \tag{18}$$

with respect to vec$\boldsymbol{B}$, $\boldsymbol{\gamma}$, which are included in $\boldsymbol{\mu}$, and $\boldsymbol{v}$, which is included $\boldsymbol{\mu}$ but also appears in the term in $\boldsymbol{v}'\boldsymbol{\Omega}^{-1}\boldsymbol{v}$. Moreover, in (18) c is a known constant and $\boldsymbol{\Omega}$ is given by

$$\boldsymbol{\Omega}=\begin{pmatrix}\boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{11} & \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{12} & \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{13}\\ \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{21} & \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{22} & \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{23}\\ \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{31} & \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{32} & \boldsymbol{I}\otimes\boldsymbol{\Sigma}^u_{33}\end{pmatrix}.$$

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

$$E[\boldsymbol{y} \mid \boldsymbol{v}] = \boldsymbol{\mu} = \boldsymbol{H}_1 \text{vec}\boldsymbol{B} + \boldsymbol{H}_2\gamma + \boldsymbol{H}_3\boldsymbol{v},$$

where

$$\boldsymbol{H}_1 = \begin{pmatrix} \boldsymbol{C}_1'\boldsymbol{H}' \otimes \boldsymbol{A}_1 \\ \boldsymbol{C}_2'\boldsymbol{H}' \otimes \boldsymbol{A}_2 \\ \boldsymbol{C}_3'\boldsymbol{H}' \otimes \boldsymbol{A}_3 \end{pmatrix}, \quad \boldsymbol{H}_2 = \begin{pmatrix} \boldsymbol{X}_1' \otimes \mathbf{1}_{p_1} \\ \boldsymbol{X}_2' \otimes \mathbf{1}_{p_2} \\ \boldsymbol{X}_3' \otimes \mathbf{1}_{p_3} \end{pmatrix}, \quad \boldsymbol{H}_3 = \begin{pmatrix} \boldsymbol{Z}_1' \otimes \boldsymbol{I} \\ \boldsymbol{Z}_2' \otimes \boldsymbol{I} \\ \boldsymbol{Z}_3' \otimes \boldsymbol{I} \end{pmatrix},$$

and

$$D[\boldsymbol{y} \mid \boldsymbol{v}] = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 \boldsymbol{I}_{p_1 n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \boldsymbol{I}_{p_2 n_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_2^2 \boldsymbol{I}_{p_3 n_3} \end{pmatrix}.$$

Supposing $\boldsymbol{\Sigma}^u$ is known, and then using (18) together with standard results from linear models theory we find estimators of the unknown parameters and of $\boldsymbol{v}$ as a function of $\boldsymbol{\Sigma}^u$ and thereafter replacement of $\boldsymbol{\Sigma}^u$ by its estimator, which is obtained as described in Section 4.1, yields an estimator $\widehat{\boldsymbol{v}}$, among other estimators.

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 d-th area and each g-th group units, we consider the means for sample observations of the data matrices $\boldsymbol{Y}_1, \boldsymbol{Y}_2$ and $\boldsymbol{Y}_3$ and predict the means of non sampled values. Use the superscripts s and r to indicate the corresponding partitions for observed sample data and non observed sample data in the target population, respectively. Therefore, we denote by $\boldsymbol{X}_{id}^{(r)} : r \times (N_d - n_{id})$, $\boldsymbol{C}_{id}^{(r)} : mk \times (N_d - n_{id})$ and $\boldsymbol{z}_{id}^{(r)} : (N_d - n_{id}) \times 1$ the corresponding matrix of covariates, design matrix and design vector for non sampled units in the population, respectively. Then, the prediction of small area means at each time point and for different group units is presented in the next proposition

Proposition 4.1. *Consider repeated measures data with missing values on the variable of interest for three-steps monotone sample data described by models (2-4). Then, the target small* *area means at each time point are elements of the vectors*

$$\widehat{\boldsymbol{\mu}}_d = \frac{1}{N_d}\left(\widehat{\boldsymbol{\mu}}_d^{(s)} + \widehat{\boldsymbol{\mu}}_d^{(r)}\right),\ d = 1,\dots,m,$$

where

$$\widehat{\boldsymbol{\mu}}_d^{(s)} = \begin{pmatrix} \boldsymbol{Y}_{1d}^{(s)}\mathbf{1}_{n_{1d}} \\ \boldsymbol{Y}_{2d}^{(s)}\mathbf{1}_{n_{2d}} \\ \boldsymbol{Y}_{3d}^{(s)}\mathbf{1}_{n_{3d}} \end{pmatrix},$$

and

$$\widehat{\boldsymbol{\mu}}_d^{(r)} = \begin{pmatrix} \left(\boldsymbol{A}_1\widehat{\boldsymbol{B}}\boldsymbol{C}_{1d}^{(r)} + \mathbf{1}_{p_1}\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_{1d}^{(r)} + \widehat{\boldsymbol{u}}_{1d}\boldsymbol{z}_{1d}^{(r)'}\right)\mathbf{1}_{N_d-n_{1d}} \\ \left(\boldsymbol{A}_2\widehat{\boldsymbol{B}}\boldsymbol{C}_{2d}^{(r)} + \mathbf{1}_{p_2}\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_{2d}^{(r)} + \widehat{\boldsymbol{u}}_{2d}\boldsymbol{z}_{2d}^{(r)'}\right)\mathbf{1}_{N_d-n_{2d}} \\ \left(\boldsymbol{A}_3\widehat{\boldsymbol{B}}\boldsymbol{C}_{3d}^{(r)} + \mathbf{1}_{p_3}\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_{3d}^{(r)} + \widehat{\boldsymbol{u}}_{3d}\boldsymbol{z}_{3d}^{(r)'}\right)\mathbf{1}_{N_d-n_{1d}} \end{pmatrix},\ d = 1,\dots,m.$$

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

$$\widehat{\boldsymbol{\mu}}_{dg} = \frac{1}{N_{dg}}\left(\widehat{\boldsymbol{\mu}}_{dg}^{(s)} + \widehat{\boldsymbol{\mu}}_{dg}^{(r)}\right),\ d = 1,\dots,m,\ g = 1,\dots,k,$$

where

$$\widehat{\boldsymbol{\mu}}_{dg}^{(s)} = \begin{pmatrix} \boldsymbol{Y}_{1d}^{(s)}\mathbf{1}_{n_{1dg}} \\ \boldsymbol{Y}_{2d}^{(s)}\mathbf{1}_{n_{2dg}} \\ \boldsymbol{Y}_{3d}^{(s)}\mathbf{1}_{n_{3dg}} \end{pmatrix},$$

and

$$\widehat{\boldsymbol{\mu}}_{dg}^{(r)} = \begin{pmatrix} \left(\boldsymbol{A}_1\widehat{\boldsymbol{B}}\boldsymbol{C}_{1dg}^{(r)} + \mathbf{1}_{p_1}\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_{1dg}^{(r)} + \widehat{\boldsymbol{u}}_{1d}\boldsymbol{z}_{1dg}^{(r)'}\right)\mathbf{1}_{N_{dg}-n_{1dg}} \\ \left(\boldsymbol{A}_2\widehat{\boldsymbol{B}}\boldsymbol{C}_{2dg}^{(r)} + \mathbf{1}_{p_2}\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_{2dg}^{(r)} + \widehat{\boldsymbol{u}}_{2d}\boldsymbol{z}_{2dg}^{(r)'}\right)\mathbf{1}_{N_{dg}-n_{2dg}} \\ \left(\boldsymbol{A}_3\widehat{\boldsymbol{B}}\boldsymbol{C}_{3dg}^{(r)} + \mathbf{1}_{p_3}\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_{3dg}^{(r)} + \widehat{\boldsymbol{u}}_{3d}\boldsymbol{z}_{3dg}^{(r)'}\right)\mathbf{1}_{N_{dg}-n_{3dg}} \end{pmatrix},$$

$$d = 1,\dots,m,\ g = 1,\dots,k.$$

Note that the predicted vector $\widehat{\boldsymbol{u}}_{id}$ is the d-th column of the predicted matrix $\widehat{\boldsymbol{U}}_i,\ i = 1,\dots,3$ and $\widehat{\boldsymbol{\beta}}_g$ is the column of the estimated parameter matrix $\widehat{\boldsymbol{B}}$ for the corresponding group g.

A direct application of Proposition 4.1 is to find the target small area means for each group across all time points obtained as a linear combination of $\widehat{\boldsymbol{\mu}}_{dg}$ depending on the type of the characteristics of interest.

References

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

Consortium, T. E. (2004). Enhancing small area estimation techniques to meet european needs. Technical report, Office for National Statistics, London.

Ferrante, M. R. and Pacei, S. (2004). Small area estimation for longitudinal surveys. *Statistical Methods & Applications*, 13(3):327–340.

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.

Little, R. J. and Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley & Sons, New York.

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. (2016). 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`.

Nissinen, K. (2009). *Small Area Estimation with Linear Mixed Models from Unit-level panel and Rotating panel data*. PhD thesis, University of Jyväskylä.

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.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51(3/4):313–326.

Rao, J. N. K. (2003). *Small Area Estimation.* John Wiley and Sons, New York.

Singh, B. and Sisodia, B. S. V. (2011). Small area estimation in longitudinal surveys. *Journal of Reliability and Statistical Studies*, 4(2):83–91.

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.

Woolson, R. F. and Leeper, J. D. (1980). Growth curve analysis of complete and incomplete longitudinal data. *Communications in Statistics - Theory and Methods*, 9(14):1491–1513.