Department of Mathematics

Small Area Estimation under a Multivariate Linear Model for Repeated Measures Data

Innocent Ngaruye, Joseph Nzabanita, Dietrich von Rosen and Martin Singull

LiTH-MAT-R--2015/09--SE
Department of Mathematics
Linköping University
S-581 83 Linköping, Sweden.
Small Area Estimation under a Multivariate linear model for repeated measures data

Innocent Ngaruye $^{1,2}$, Joseph Nzabanita$^{1,2}$, Dietrich von Rosen$^{1,3}$ and Martin Singull$^{1}$

$^1$Department of Mathematics, Linköping University, SE–581 83 Linköping, Sweden
E-mail: {innocent.ngaruye, joseph.nzabanita, 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, j.nzabanita}@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 paper, we consider small area estimation under a multivariate linear regression model for repeated measures data. The aim of the proposed model is to get a model which borrows strength across small areas and over time, by incorporating simultaneously the area effects and time correlation. The model accounts for repeated surveys, group individuals and random effects variations. Estimation of model parameters is discussed within a restricted maximum likelihood based approach. Prediction of random effects and the prediction of small area means across time points and per group units for all time points are derived. The results are supported by a simulation study.

Keywords: Maximum likelihood; Multivariate linear model; Prediction of random effects; Repeated measures data; Small Area Estimation
1 Introduction

Population surveys are carried out via sampling designs and data collection of individual units with intention of making statistical inferences about a larger population of which these units are members. One commonly used design is simple random sampling which assigns equal selection probabilities to all elements in the population. These surveys are usually designed to provide efficient estimates of parameters of interest for large populations. In most cases, surveys are not originally designed to produce estimates for small domains and hence these domains are poorly represented in the sample. Thus, the surveys often provide very little information on a small area level and direct survey estimates on a target small area are not reliable due to a small sample size connected to this area.

In recent years, Small Area Estimation (SAE) methods have received much attention due to their usefulness in both public and private sectors and their demand has greatly increased worldwide. Several approaches and new developments in small area estimation have been investigated by different authors for example, Pfeffermann (2002, 2013); Rao (2003); Chambers and Clark (2012). The demand for small area statistics has increased due to their use in formulating social and economic policies, allocation of government funds, regional planning, business decision making etc. SAE has been used in a wide range of applications such as unemployment rates, poverty mapping, disease mapping, demography etc. One may refer to Ghosh and Rao (1994) and Rao (2003) for some examples and case studies in SAE.

Repeated measures data which refer to response outcomes taken on the same experimental unit at different time points have been widely used in research. The analysis of repeated measures data allows us to study trends over time. The demand for small area statistics is for both cross-sectional and for repeated measures data. For instance, small area estimates for repeated measures data may be used by public policy makers for different purposes such as funds allocation, new educational or health programs and in some cases, they might be interested in a given group of population.

It has been shown that the multivariate approach for model-based methods in small area estimation may achieve substantial improvement over the usual univariate approach (Datta et al., 1999). Some studies dealing with SAE problems for longitudinal surveys have been discussed by various authors, for example Consortium (2004); Nissinen (2009); Singh and Sisodia (2011). The latter has developed direct, synthetic and composite estimators for small area means at a given time point when the population units contain non-overlapping groups.

In the same framework, in order to gain considerable efficiency of esti-
mators, we propose a model which borrows strength across both small areas and over time by incorporating simultaneously the effects of areas and time correlation. The model allows for finding the small area means at each time point, per group units and particularly the pattern of changes or mean growth profiles over time. This model accounts for repeated surveys, group of individuals and random effects variations.

The paper is organized as follows. After the Introduction, the second section is devoted to the model formulation and notation used and Section 3 deals with estimation of model parameters. Sections 4 and 5 discuss the prediction of random effects and prediction of target small area means, respectively. We end up by a simulation study is Section 6 and general concluding remarks in Section 7.

2 Model formulation

We consider repeated measurements on the variable of interest $y$ taken at $p$ time points $t_1, \ldots, t_p$ from a finite population $U$ of size $N$ which is partitioned into $m$ disjoint subpopulations $U_1, \ldots, U_m$ called small areas of sizes $N_i, i = 1, \ldots, m$ such that $\sum_{i=1}^{m} N_i = N$. We also assume that in every area, there are $k$ different groups of units of size $N_g$ for group $g$ such that $\sum_{g=1}^{k} N_g = N_i$ and $\sum_{i=1}^{m} \sum_{g=1}^{k} N_{ig} = N$, where $N_{ig}$ is the group size within the $i$-th area. Suppose that a sample $s = s_1, \ldots, s_m$ is selected from the population using simple random sampling without replacement, where $s_i$ is the sample of size $n_i$ observed from area $i$. The sample remains the same over time.

Let $y_{ij}$ be the $p$-vector of measurements on the $j$-th unit, in the $i$-th area. That is $y_{ij} = (y_{ij1}, \ldots, y_{ijp})'$, $j = 1, \ldots, N_i$, $i = 1, \ldots, m$.

We assume the mean growth of the $j$th unit in area $i$ for each group to be a polynomial in time of degree $q - 1$. Furthermore, we suppose that we have auxiliary data $x_{ij}$ of $r$ concomitant variables (covariates) whose values are known for all units in all $m$ small areas. These auxiliary variables are included in the model to strengthen the limited sample size data from areas. The values $x_{ij}$ can be the values of the survey from the same area in the past and/or the values of the other variables that are related to the variable of interest. Moreover, they can be register based information or the data measured for characteristics of interest in other similar areas.

The relationship between $y_{ij}$ and $x_{ij}$ in each small area is not always considered as the same as the relationship between the variables in the population as whole (Chambers and Clark, 2012). So, we have to add an area specific term to allow them to better account for the between area variability in the distribution of $y_{ij}$. That is, consequently to assume that we have $u_{it}$
random area-effects which vary over time. Thus, for each one of the \( k \) groups, the unit level regression model for \( j \)-th unit coming from the small area \( i \) at time \( t \) can be expressed by

\[
y_{ijt} = \beta_0 + \beta_1 t + \cdots + \beta_q t^{q-1} + \gamma' x_{ij} + u_{it} + e_{ijt},
\]

where \( e_{ijt} \) are random sampling errors depending on the sampling scheme assumed to be i.i.d normal with mean zero and known sampling variance \( \sigma_e^2 \) independent of \( u_{it} \). The \( \gamma \) is a vector of fixed regression coefficients of auxiliary variables. The \( \beta_0, \ldots, \beta_q \) are unknown parameters assumed to be the same in all areas under the assumption that there is no area effect on polynomial growth in time.

For all time points, the model can be written in matrix form as

\[
y_{ij} = A\beta + 1_p \gamma' x_{ij} + u_i + e_{ij},
\]

where \( A = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{q-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{q-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_p & t_p^2 & \cdots & t_p^{q-1} \end{pmatrix}, \quad 1_p = (1, 1, \ldots, 1)' : p \times 1.
\]

The vector \( u_i \) is assumed to be multivariate normally distributed with zero mean and unknown covariance matrix \( \Sigma_u \). Collecting the vectors \( y_{ij} \) for all units in small area \( i \) coming from the group \( g \) gives

\[
Y_{ig} = AB_g + 1_p \gamma' X_{ig} + u_i z'_i + E_{ig}, \quad g = 1, \ldots, k,
\]

where \( Y_{ig} = (y_{i1}, \ldots, y_{iN_i})' \); \( B_g = (\beta_{g1}, \ldots, \beta_{gq}) : q \times N_i \); \( X_{ig} = (x_{i1}, \ldots, x_{iN_i}) \); \( z_i = (z_{i1}', \ldots, z_{ik}')' \); \( E_{ig} = (e_{i1}, \ldots, e_{iN_i}) \).

The model presented in (1) holds for each group unit in area \( i \). Since we are interested in every group unit, we need to include a known design matrix \( C_i : k \times N_i \) of group separation indicators when collecting all \( k \) groups in the \( i \)-th area, which are supposed to be the same in all areas.

The model at small area level for \( k \) groups is then written as

\[
Y_i = ABC_i + 1_p \gamma' X_i + u_i z'_i + E_i,
\]

where \( Y_i = (Y_{i1}, \ldots, Y_{ik})' \); \( B = (\beta_{11}, \ldots, \beta_{kk}) : q \times k \); \( X_i = (X_{i1}, \ldots, X_{ik}) \); \( z_i = (z'_{i1}, \ldots, z'_{ik})' \); \( E_{ig} = (e_{i1}, \ldots, e_{ik}) \);

\[
C_i = \begin{pmatrix} 1'_{N_{i1}} & 0 \\ \vdots & \ddots \\ 0 & 1'_{N_{ik}} \end{pmatrix},
\]
where $E_i \sim \mathcal{N}_{p,N_i}(0, \Sigma_e, I_{N_i})$ standing for matrix normal distribution with mean zero, positive definite covariance matrix between rows $\Sigma_e = \sigma_e^2 I_p$ and independent columns. $Y_i$ is a $p \times N_i$ data matrix; $A$ is a $p \times q, q \leq p$ known within individuals design matrix for fixed effects; $B$ is a $q \times k$ unknown parameter matrix; $C_i$ with rank($C_i$) + $p \leq N_i$ is a $k \times N_i$ known between individuals design matrix for fixed effects and $X_i$ is a $r \times N_i$ known matrix taking the values of the covariates.

Combining all small areas for $N = \sum_{i=1}^{m} N_i$ units, we get the model

$Y = A B H C + 1_p \gamma' X + U Z + E$, \hspace{1cm} (2)

where $Y = [Y_1, \cdots, Y_m]$; $H = (I_k : \cdots : I_k)$; $X = [X_1, \cdots, X_m]$;

$U = [u_1, \cdots, u_m]$; $E = [E_1, \cdots, E_m]$;

$C = \begin{pmatrix} C_1 & 0 \\ \vdots & \ddots \\ 0 & C_m \end{pmatrix}$ and $Z = \begin{pmatrix} z'_1 \\ \vdots \\ z'_m \end{pmatrix}$.

It is assumed that $E \sim \mathcal{N}_{p,N}(0, \Sigma_e, I_N)$, $U \sim \mathcal{N}_{p,m}(0, \Sigma_u, I_m)$, $p \leq m$. Furthermore, $\text{vec}(Y) \sim \mathcal{N}_{pN}(\text{vec}(ABHC + 1_p \gamma' X), \Sigma)$ for $\Sigma = Z'Z \otimes \Sigma_u + I_N \otimes \Sigma_e$ and

$\Sigma_u = \begin{pmatrix} \sigma_{u1}^2 & \sigma_{u12} & \cdots & \sigma_{u1p} \\ \sigma_{u12} & \sigma_{u2}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u1p} & \cdots & \cdots & \sigma_{up}^2 \end{pmatrix} : p \times p$,

where the covariance between random effects $u_i$’s at two distinct times $t, s$ is $\text{Cov}(u_t, u_s) = \sigma_{uts} = \rho_{t,s} \sigma_u \sigma_u$ with $\rho_{t,s}$ standing for the correlation.

3 Estimation for the mean and covariance

The model defined in (2) is a sum of two matrix normal distributions which is normally distributed but not in general matrix normally distributed. Therefore, in order to use the maximum likelihood estimation approach, we make some transformations to achieve a matrix normal distribution which is easier to handle. In the following, we use the notation $A^o$ for any matrix of full rank.
positive definite, and \( Q \) full row rank and the matrix spanning both relations imply that \((C C')\) also denote by \( P \) an arbitrary generalized inverse of the matrix \( A \) such that \( A A^{-} A = A \). We also denote by \( P_A = A(A' A)^{-} A' \), \( Q_A = I - P_A \), \( P_{A,B} = A(A' B A)^{-} A' B \) and \( Q_{A,B} = I - P_{A,B} \). \( P_A \) and \( Q_A \) are projector matrices and if \( B \) is positive definite, \( P_{A,B} \) and \( Q_{A,B} \) are also projectors.

First of all, we observe that the matrices \( Z, H, C \) from model (2) are of full row rank and the matrix \( A \) is of full column rank. Moreover,

\[
C(Z') \subseteq C(C') \quad \text{and} \quad ZZ' = I_m.
\]

Both relations imply that \((C C')^{-1/2} C Z' Z C'(C C')^{-1/2} \) is an idempotent matrix and thus can be diagonalized by an orthogonal matrix, say \( \Gamma \) with corresponding eigenvalues 1 and 0, i.e.,

\[
(CC')^{-1/2} C Z' Z C'(CC')^{-1/2} = \Gamma D \Gamma' = \Gamma \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \Gamma'.
\]

Since each column of \( \Gamma \) is an eigenvector associated to the corresponding eigenvalue in \( D \), we can partition \( \Gamma = [\Gamma_1 : \Gamma_2] \) such that \( \Gamma_1 \) corresponds to the block \( I_m \) and \( \Gamma_2 \) corresponds to the block \( 0 \), with \( \Gamma_1 : mk \times m \) and \( \Gamma_2 : mk \times (mk - m) \). It follows that \( \Gamma_1^T \Gamma_1 = I_m \) and \( \Gamma_2^T \Gamma_2 = I_{mk-m} \).

Choose \( C^{\gamma \circ} \) to be a matrix which columns are unit length eigenvectors corresponding to non zero eigenvalues of \( I - C'(CC')^{-1} C \). Then we have \((C^{\gamma \circ})' C^{\gamma \circ} = I_{N-mk} \) and we can make a one to one transformation on (2) to get \( [V : W] = Y[C'(CC')^{-1/2} : C^{\gamma \circ}] \), where

\[
V = Y C'(CC')^{-1/2} = ABH(CC')^{1/2} + 1_p \gamma' X C'(CC')^{-1/2}
+ (U Z + E) C'(CC')^{-1/2}, \quad (3)
\]

\[
W = Y C^{\gamma \circ} = 1_p \gamma' X C^{\gamma \circ} + E C^{\gamma \circ}, \quad (4)
\]

since \( CC^{\gamma \circ} = 0 \) and \( C(Z') \subseteq C(C') \) implies that \( Z C^{\gamma \circ} = 0 \). Make a further transformation on (3) so that \( [V_1 : V_2] = Y C'(CC')^{-1/2} [\Gamma_1 : \Gamma_2] \), where

\[
V_1 = Y C'(CC')^{-1/2} \Gamma_1 = ABH(CC')^{1/2} \Gamma_1 + 1_p \gamma' X C'(CC')^{-1/2} \Gamma_1
+ (U Z + E) C'(CC')^{-1/2} \Gamma_1, \quad (5)
\]

\[
V_2 = Y C'(CC')^{-1/2} \Gamma_2 = ABH(CC')^{1/2} \Gamma_2 + 1_p \gamma' X C'(CC')^{-1/2} \Gamma_2
+ E C'(CC')^{-1/2} \Gamma_2. \quad (6)
\]

**Theorem 3.1** Let \( V_1, V_2 \) and \( W \) be as defined in (5), (6) and (4),
respectively. Then,

\[ V_1 \sim \mathcal{N}_{p,m} \left( M_1, \Sigma_u + \Sigma_e, I_m \right), \]
\[ V_2 \sim \mathcal{N}_{p,m-k} \left( M_2, \Sigma_e, I_{mk-m} \right), \]
\[ W \sim \mathcal{N}_{p,N-m} \left( M_3, \Sigma_e, I_{N-m} \right), \]

where

\[ M_1 = ABH (CC')^{1/2} \Gamma_1 + 1_p \gamma' X C'(CC')^{-1/2} \Gamma_1, \]
\[ M_2 = ABH (CC')^{1/2} \Gamma_2 + 1_p \gamma' X C'(CC')^{-1/2} \Gamma_2, \]
\[ M_3 = 1_p \gamma' X C'^{\alpha}. \]

The matrices \( V_1, V_2 \) and \( W \) are independently distributed.

**Proof 3.1** The independence of \( V_1, V_2 \) and \( W \) follows by the fact that the matrices \( \Gamma_1, \Gamma_2 \) and \( C'^{\alpha} \) are all of full rank and pairwise orthogonal. We observe that

\[ \Gamma_1 \Gamma_1' = \begin{bmatrix} I_m & 0 \\ 0 & I_{mk-m} \end{bmatrix}, \quad \Gamma_2 \Gamma_2' = \begin{bmatrix} I_m & 0 \\ 0 & I_{mk-m} \end{bmatrix} \]

From

\[ E \sim \mathcal{N}_{p,N}(0, \Sigma_e, I_N), \quad U \sim \mathcal{N}_{p,m}(0, \Sigma_u, I_m), \]

it follows that

\[ U Z C'(CC')^{-1/2} \Gamma_1 \sim \mathcal{N}_{p,m} \left( 0, \Sigma_u, \Gamma_1'(CC')^{-1/2} C Z' Z C'(CC')^{-1/2} \Gamma_1 \right), \]
\[ E C'(CC')^{-1/2} \Gamma_1 \sim \mathcal{N}_{p,m} \left( 0, \Sigma_e, \Gamma_1'(CC')^{-1/2} C C'(CC')^{-1/2} \Gamma_1 \right), \]
\[ E C'(CC')^{-1/2} \Gamma_2 \sim \mathcal{N}_{p,m-k} \left( 0, \Sigma_e, \Gamma_2'(CC')^{-1/2} C C'(CC')^{-1/2} \Gamma_2 \right), \]
\[ E C'^{\alpha} \sim \mathcal{N}_{p,N-m} \left( 0, \Sigma_e, (C'^{\alpha})' C'^{\alpha} \right). \]

Hence,

\[ (U Z + E) C'(CC')^{-1/2} \Gamma_1 \sim \mathcal{N}_{p,m} \left( 0, \Sigma_u + \Sigma_e, I_m \right), \]
\[ E C'(CC')^{-1/2} \Gamma_2 \sim \mathcal{N}_{p,m-k} \left( 0, \Sigma_e, I_{mk-m} \right), \]
\[ E C'^\alpha \sim \mathcal{N}_{p,N-m} \left( 0, \Sigma_e, I_{N-m} \right). \]
since

\[ \Gamma'(CC')^{-1/2}CZ'Z(CC')^{-1/2}\Gamma_1 = \Gamma'D\Gamma'\Gamma_1 = I_m, \]
\[ \Gamma'(CC')^{-1/2}CC'(CC')^{-1/2}\Gamma_1 = I_m, \]
\[ \Gamma_2'(CC')^{-1/2}CC'(CC')^{-1/2}\Gamma_2 = I_{mk-m}, \]
\[ (C^o)^'C^o = I_{N-mk}, \]

which completes the proof of the theorem.

Now consider the two components \( W \) and \( V_2 \) and recall that the covariance matrix \( \Sigma_e = \sigma^2 I_p \) is known. Put

\[ K_1 = H(CC')^{1/2}\Gamma_1, \quad K_2 = H(CC')^{1/2}\Gamma_2, \]
\[ R_1 = C'(CC')^{-1/2}\Gamma_1, \quad R_2 = C'(CC')^{-1/2}\Gamma_2. \]

Then,

\[ W = YC^o = 1_p\gamma'XC^o + EC^o, \]
\[ V_2 = YR_2 = ABK_2 + 1_p\gamma'XR_2 + ER_2. \]

The corresponding log-likelihood of the joint density function is given by

\[ l(\gamma, B) = l_W(\gamma) + l_{V_2}(\gamma, B), \]

where \( l_W(\gamma) \) and \( l_{V_2}(\gamma, B) \) denote the log-likelihood functions for \( W \) and \( V_2 \), respectively. Let use the notation \((Y)(Y)'\) instead of \((Y)(Y)\)' in order to shorten matrix expressions. Therefore,

\[
l(\gamma, B) = -\frac{p(N - mk)}{2} \log (2\pi) - \frac{N - mk}{2} \log |\Sigma_e| - \frac{p(mk - m)}{2} \log (2\pi)
- \frac{mk - m}{2} \log |\Sigma_e| - \frac{1}{2} \text{tr} \left\{ \Sigma_e^{-1} \left( W - 1_p\gamma'XC^o \right) \right\} \\
- \frac{1}{2} \text{tr} \left\{ \Sigma_e^{-1} \left( V_2 - ABK_2 - 1_p\gamma'XR_2 \right) \right\} \\
= c_o - \frac{1}{2} \text{tr} \left\{ \Sigma_e^{-1} \left( W - 1_p\gamma'XC^o \right) \right\} \\
+ \Sigma_e^{-1} \left( V_2 - ABK_2 - 1_p\gamma'XR_2 \right) \left( \gamma \right)'
\]

where \( c_o \) is a constant not depending on the parameters and \( \text{tr} \) stands for the trace.
Taking the first and second derivatives with respect to $\gamma$ and $B$, the likelihood equations can be expressed as
\[
\left( XC'^o(C'^o)' + XR_2R'_2 \right) Y'1_p - p\left( XC'^o(C'^o)'X' + XR_2R'_2X' \right) \gamma \\
- XR_2K'_2B'\gamma A'1_p = 0, \quad (7)
\]
\[
A'(YR_2 - ABK_2 - 1_p\gamma'XR_2)K'_2 = 0. \quad (8)
\]

**Lemma 3.1** The likelihood equation (7) admits a unique solution for the parameter vector $\gamma$ if the matrix $X$ is of full rank and the likelihood equation (8) admits a non unique solution for the parameter matrix $B$.

The proof of the lemma can be found in Appendix. The second equation (8) gives
\[
B = (A'A)^{-1}A'YR_2K'_2(K_2K_2')^{-} - (A'A)^{-1}A'1_p\gamma'XR_2K'_2(K_2K_2')^{-} \\
+ T_1(K_2K_2')o',
\]
for an arbitrary matrix $T_1$. Plugging in the value of $B$ into equation (7) yields
\[
\left( XC'^o(C'^o)' + XR_2R'_2 \right) Y'1_p - p\left( XC'^o(C'^o)'X' + XR_2R'_2X' \right) \gamma \\
- XR_2K'_2(K_2K_2')^{-}K_2R'_2Y'1_p + pXR_2K'_2(K_2K_2')^{-}K_2R'_2X'\gamma = 0.
\]
Then, the restricted maximum likelihood estimators (RMLEs) for $\gamma$ and $B$ are obtained by
\[
\hat{\gamma} = \left( pXC'^o(C'^o)'X' + pXR_2R'_2X' - pXR_2K'_2(K_2K_2')^{-}K_2R'_2X' \right)^{-} \\
\times \left( XC'^o(C'^o)'Y' + XR_2R'_2Y' - XR_2K'_2(K_2K_2')^{-}K_2R'_2Y' \right)1_p \\
+ \left( pXC'^o(C'^o)'X' + pXR_2R'_2X' - pXR_2K'_2(K_2K_2')^{-}K_2R'_2X' \right)^{a_o}t_2,
\]
\[
\hat{B} = (A'A)^{-1}A'YR_2K'_2(K_2K_2')^{-} - (A'A)^{-1}A'1_p\hat{\gamma}'XR_2K'_2(K_2K_2')^{-} \\
+ T_1(K_2K_2')o',
\]
for an arbitrary vector $t_2$ and an arbitrary matrix $T_1$. Put $P = XC'^o(C'^o)' + XR_2R'_2 - XR_2K'_2(K_2K_2')^{-}K_2R'_2$. Then
\[
\hat{\gamma} = \frac{1}{p}(PX')^oPY'1_p + (PX')^o t_2, \quad (9)
\]
\[
\hat{B} = (A'A)^{-1}A'YR_2K'_2(K_2K_2')^{-} \\
- \frac{1}{p}(A'A)^{-1}A'1_p1_p'(XP')^oXR_2K'_2(K_2K_2')^{-} \\
- (A'A)^{-1}A'1_p t'_2(PX')^o XR_2K'_2(K_2K_2')^{-} + T_1(K_2K_2')o'. \quad (10)
\]
The estimators $\hat{\gamma}$ of $\gamma$ and $\hat{B}$ of $B$ depend on arbitrary vector $t_2$ and arbitrary matrix $T_1$, respectively. However, there is also information about $B$ and $\gamma$ in $V_1$ which we now try to utilize in order to find the expressions of $T_1$ and $t_2$. Recall that

$$V_1 = Y R_1 = A B K_1 + 1_p \gamma' X R_1 + (U Z + E) R_1,$$

$$E R_1 \sim N_{p,m}(0, \Sigma_u + \Sigma_e, I_m).$$

Inserting the values of $\gamma$ and $B$ in the model $V_1$ above, yields

$$V_1 = A (A' A)^{-1} A' Y R_2 K_2' (K_2 K_2')^{-1} K_1$$

$$- \frac{1}{p} 1_p 1_p' Y P' (X P')^{-1} X R_2 K_2' (K_2 K_2')^{-1} K_1$$

$$- 1_p t_2' (P X')' X R_2 K_2' (K_2 K_2')^{-1} K_1 + A T_1 (K_2 K_2')' K_1$$

$$+ \frac{1}{p} 1_p 1_p' Y P' (X P')^{-1} X R_1 + 1_p t_2' (P X')' X R_1 + (U Z + E) R_1.$$

Set

$$V_3 = Y R_1 - A (A' A)^{-1} A' Y R_2 K_2' (K_2 K_2')^{-1} K_1$$

$$+ \frac{1}{p} 1_p 1_p' Y P' (X P')^{-1} X R_2 K_2' (K_2 K_2')^{-1} K_1 - \frac{1}{p} 1_p 1_p' Y P' (X P')^{-1} X R_1,$$

with model equation

$$V_3 = A T_1 (K_2 K_2')' K_1 + 1_p t_2' (P X')' X R_1$$

$$- 1_p t_2' (P X')' X R_2 K_2' (K_2 K_2')^{-1} K_1 + (U Z + E) R_1,$$

or equivalently,

$$V_3 = A T_1 \tilde{C}_1 + 1_p t_2' \tilde{C}_2 + \tilde{E}, \quad (11)$$

where

$$\tilde{C}_1 = (K_2 K_2')' K_1,$$

$$\tilde{C}_2 = (P X')' X R_1 - (P X')' X R_2 K_2' (K_2 K_2')^{-1} K_1,$$

$$\tilde{E} = (U Z + E) R_1, \text{ with } \tilde{E} \sim N_{p,m}(0, \Sigma_u + \Sigma_e, I_m).$$

Hence, the model obtained in (11) is an extended multivariate linear growth curve model with nested subspace condition $C(1_p) \subseteq C(A)$ on the within design matrices. One can refer to Filipiak and von Rosen (2012) for more details about extended multivariate linear growth curve model and the corresponding maximum likelihood estimators.
Lemma 3.2 Let $V_3$ be defined as in (11). The RMLEs for the parameter matrices $T_1$, $t_2$ and $\Sigma_u$ are given by

\[
\hat{t}_2 = (1_p S_1^{-1} 1_p)^{-1} (\tilde{C}_2 Q C_1 \tilde{C}_1') - \tilde{C}_2 Q C_1 V_3' S_1^{-1} 1_p + (\tilde{C}_2 Q C_1)^\prime t_2^1 1_p, \tag{12}
\]

\[
\hat{T}_1 = (A' S_2^{-1} A)^{-1} A' S_2^{-1} (V_3 - 1_p \tilde{t}_2 \tilde{C}_2) \tilde{C}_1 (\tilde{C}_1 \tilde{C}_1')^{-1} + A T_{11} \tilde{C}_1, \tag{13}
\]

\[
\tilde{\Sigma}_u = \frac{1}{m} (V_3 - A \hat{T}_1 \tilde{C}_1 - 1_p \tilde{t}_2 \tilde{C}_2) (V_3 - A \hat{T}_1 \tilde{C}_1 - 1_p \tilde{t}_2 \tilde{C}_2)' - \Sigma_e, \tag{14}
\]

for an arbitrary vector $t_{21}$ and an arbitrary matrix $T_{11}$, where $S_1 = V_3 Q (C_1', \tilde{C}_2') V_3$, $S_2 = S_1 + Q_{1_p} s_1^{-1} V_3 P Q_{C_1'} C_2' V_3' Q_{1_p} s_1^{-1}$.

The proof follows similarly to Theorem 1 in Filipiak and von Rosen (2012) in a particular case of two profiles.

Following Filipiak and von Rosen (2012), from model (11), $\hat{T}_1$ and $\hat{t}_2$ are unique if and only if the matrices $\tilde{C}_1$ and $\tilde{C}_2$ are of full rank respectively and $C(\tilde{C}_1) \cap C(\tilde{C}_2) = \{0\}$. However, $A \hat{T}_1 \tilde{C}_1$ and $1_p \tilde{t}_2 \tilde{C}_2$ are unique and hence $\tilde{\Sigma}_u$ given in (14) is unique.

With the above calculations, we are now ready to give a theorem which summarize the main results about estimation of the formulated model.

Theorem 3.2 Consider the model given by (2). Then, the RMLEs of $\gamma$, $B$ and $\Sigma_u$ can be expressed as

\[
\hat{\gamma} = \frac{1}{p} (P X')^{-1} P Y' 1_p + (P X')^\prime \hat{t}_2,
\]

\[
\hat{B} = (A' A)^{-1} A Y R_2 K_2' (K_2 K_2')^{-1}
- \frac{1}{p} (A' A)^{-1} A 1_p 1_p' Y P' (X P')^{-1} X R_2 K_2' (K_2 K_2')^{-1}
- (A' A)^{-1} A 1_p \tilde{t}_2 (P X')^\prime X R_2 K_2' (K_2 K_2')^{-1} + \hat{T}_1 (K_2 K_2')^\prime,
\]

\[
\tilde{\Sigma}_u = \frac{1}{m} (V_3 - A \hat{T}_1 \tilde{C}_1 - 1_p \tilde{t}_2 \tilde{C}_2) (V_3 - A \hat{T}_1 \tilde{C}_1 - 1_p \tilde{t}_2 \tilde{C}_2)' - \Sigma_e,
\]

where $\tilde{t}_2$ and $\hat{T}_1$ are given by (12) and (13), respectively.

This Theorem follows from relations (9), (10) and Lemma 3.2.

4 Prediction of random effects

The small area means in a given area $i$ are defined as the conditional mean given the realized area effects (Battese et al., 1988). Thus, estimates of random area effects are needed in the estimation of small area means. As pointed
out by Pinheiro and Bates (2000), although technically the random effects are not model parameters, in some ways they do behave like parameters and since they are unobservable we want to predict their values. The idea is to predict unobservable random variable from some realized values. Robinson (1991) discusses the need of prediction of random effects and summarizes the theory best linear unbiased predictor (BLUP). Searle et al. (2009) present the theory of prediction of random variables. As stated by these authors, the minimum variance is used for estimation of a fixed parameter while the minimum mean square is used for estimation of the realized value of a random variable. In this section, we use the approach developed by Henderson (1973) for prediction of random effects which consists of maximizing the joint density between the observable random variable and non observable random variable.

Consider the model in (2) given by
\[ Y = ABHC + 1_p\gamma'X + UZ + E, \]
and maximize the joint density \( f(Y, U) \) with respect to \( U \) assuming the covariance matrices \( \Sigma_u \) and \( \Sigma_e \) to be known. We get
\[
f(Y, U) = f(U)f(Y|U) = \lambda \exp \left\{ -\frac{1}{2} \text{tr} \{ U'\Sigma_u^{-1}U \} \right\} \times \exp \left\{ -\frac{1}{2} \text{tr} \{ \Sigma_e^{-1}(Y - ABHC - 1_p\gamma'X - UZ)(Y') \} \right\},
\]
where \( \lambda \) is a constant. Then, the estimating equation for \( U \) equals
\[
\Sigma_e^{-1}(Y - ABHC - 1_p\gamma'X - UZ)Z' - \Sigma_u^{-1}U = 0,
\]
which is equivalent to
\[
\Sigma_e \Sigma_u^{-1}U + UZZ' = (Y - ABHC - 1_p\gamma'X)Z',
\]
and since \( ZZ' = I_m \), it follows that
\[
(\Sigma_e \Sigma_u^{-1} + I_p)U = (Y - ABHC - 1_p\gamma'X)Z'.
\]
Thus, we get the following theorem about the prediction of random effects

**Theorem 4.1** Consider the model defined by (2). Then, the prediction of random effects is given by
\[
\hat{U} = \left( \Sigma_e \hat{\Sigma}_u^{-1} + I_p \right)^{-1}(Y - ABHC - 1_p\hat{\gamma}'X)Z',
\]
where \( \hat{\gamma}, \hat{B} \) and \( \hat{\Sigma}_u \) are given in Theorem 3.2.
5 Prediction of target small area means

We assume that the small area model holds for the sample data which means that no sample selection bias and that the sampling design is not informative. The model in (2) comprises in theory, sampled and non sampled units. That is

\[ \mathbf{Y} = (\mathbf{Y}^{(s)}_1, \ldots, \mathbf{Y}^{(s)}_m, \mathbf{Y}^{(r)}_1, \ldots, \mathbf{Y}^{(r)}_m) : p \times N, \]

where \( \mathbf{Y}^{(s)}_i = (y_{i1}, \ldots, y_{in_i}) : p \times n_i, \) represents the sampled \( n_i \) observations from the \( i \)-th small area and \( \mathbf{Y}^{(r)}_i = (y_{in_i+1}, \ldots, y_{iN_i}) : p \times (N_i - n_i, \) corresponds to the non sampled \( (N_i - n_i) \) units from the \( i \)-th small area.

Then, split the sample \( s_i \) into \( s_{ig}, g = 1, \ldots, k \) with corresponding sample sizes \( n_{ig} \) for \( k \) groups. Therefore, the target vector in small area \( i \) which elements are area means at each time point is given by

\[ \bar{\mathbf{y}}_i = f_i \mathbf{Y}^{(s)}_i + (1 - f_i) \hat{\mathbf{Y}}^{(r)}_i, \]

where \( \mathbf{Y}^{(s)}_i \) is the vector of small area means corresponding to sampled units, \( \hat{\mathbf{Y}}^{(r)}_i \) is the vector of predicted small area means corresponding to non-sampled units and \( 1 - f_i \) is the finite population correction factor with \( f_i = \frac{n_i}{N_i} \) the sampling fraction, that is the fraction of the population that is sampled.

Therefore,

\[ \bar{\mathbf{y}}_i = \frac{f_i}{n_i} \mathbf{Y}^{(s)}_i 1_{n_i} + \frac{1 - f_i}{N_i - n_i} \hat{\mathbf{Y}}^{(r)}_i 1_{N_i - n_i} = \frac{1}{N_i} \left( \mathbf{Y}^{(s)}_i 1_{n_i} + \hat{\mathbf{Y}}^{(r)}_i 1_{N_i - n_i} \right), \]

where

\[ \hat{\mathbf{Y}}^{(r)}_i = \mathbf{A} \hat{\mathbf{B}} \mathbf{C}_i + 1_p \hat{\gamma}' \mathbf{X}_i + \hat{\mathbf{u}}_i \mathbf{z}'_i. \]  

(15)

It is convenient to note that \( \hat{\gamma}, \hat{\mathbf{B}} \) and \( \hat{\mathbf{u}}_i \) used in (15) are estimators computed from Theorem 3.2 and Theorem 4.1 , respectively using observed data. Then, the target vector of small area means for each group \( g \) across all time points is given by

\[ \bar{\mathbf{y}}_{ig} = \frac{1}{N_{ig}} \left( \mathbf{Y}^{(s)}_{ig} 1_{n_{ig}} + \hat{\mathbf{Y}}^{(r)}_{ig} 1_{N_{ig} - n_{ig}} \right), \quad g = 1, \ldots, k. \]
Equivalently,
\[
\bar{y}_{ig} = \frac{1}{N_{ig}} \sum_{j \in s_{ig}} y_{ij} + (1 - \frac{n_{ig}}{N_{ig}}) A \hat{\beta}_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} 1_p \hat{\gamma}' x_{ij} \\
+ (1 - \frac{n_{ig}}{N_{ig}}) \hat{u}_i + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} e_{ij},
\]
where \( \hat{\beta}_g \) is the column of the estimated parameter matrix \( \hat{B} \) corresponding to the group \( g \) and the predicted vectors \( \hat{u}_i \)'s are the columns of the predicted matrix \( \hat{U} \). The first term of this expression on the right side is known and by the strong law of large numbers, if \( N_{ig} \) is large, the last term is approximately equal to zero. Following Henderson (1975), the linear predictor of
\[
y_{ig} = (1 - \frac{n_{ig}}{N_{ig}}) A \beta_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} 1_p \gamma' x_{ij} + (1 - \frac{n_{ig}}{N_{ig}}) u_i,
\]
is given by
\[
\hat{y}_{ig} = (1 - \frac{n_{ig}}{N_{ig}}) A \hat{\beta}_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} 1_p \hat{\gamma}' x_{ij} + (1 - \frac{n_{ig}}{N_{ig}}) \hat{u}_i,
\]
where \( \hat{\beta}_g \) are columns of the estimated parameter matrix \( \hat{B} \). Hence,
\[
\bar{y}_{ig} = \frac{1}{N_{ig}} \sum_{j \in s_{ig}} y_{ij} + (1 - \frac{n_{ig}}{N_{ig}}) A \hat{\beta}_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} 1_p \hat{\gamma}' x_{ij} + (1 - \frac{n_{ig}}{N_{ig}}) \hat{u}_i.
\]
Note that the population means of auxiliary variables \( x \) in area \( i \) at time \( t \) must be known so that the non-sampled mean \( \sum_{j \notin s_{ig}} x_{ijt} \) is then obtained by substracting the corresponding sample means from the population mean.

**Theorem 5.1** Given the multivariate linear regression model as defined in (2). The target small area means at each time point are the elements of the vectors
\[
\bar{y}_i = \frac{1}{N_i} (Y_i^{(s)} 1_{n_i} + \hat{Y}_i^{(r)} 1_{N_i-n_i}),
\]
where \( \hat{Y}_i^{(r)} = A \hat{B} C_i + 1_p \hat{\gamma}' X_i + \hat{u}_i z_i \). The target small area means for each group across all time points are the elements of the vectors
\[
\bar{y}_{ig} = \frac{1}{N_{ig}} \sum_{j \in s_{ig}} y_{ij} + (1 - \frac{n_{ig}}{N_{ig}}) A \hat{\beta}_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} 1_p \hat{\gamma}' x_{ij} + (1 - \frac{n_{ig}}{N_{ig}}) \hat{u}_i.
\]
6 Simulation study example

In this section, we present a simulation study where we consider four times repeated surveys on a population of size having 8 small areas and draw a sample of size $n = 450$ with the following small area sample sizes given in Table 1,

<table>
<thead>
<tr>
<th>Area</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$n_{11} = 12$</td>
<td>$n_{12} = 18$</td>
<td>$n_{13} = 16$</td>
<td>$n_1 = 46$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{21} = 21$</td>
<td>$n_{22} = 23$</td>
<td>$n_{23} = 12$</td>
<td>$n_2 = 56$</td>
</tr>
<tr>
<td>3</td>
<td>$n_{31} = 10$</td>
<td>$n_{32} = 20$</td>
<td>$n_{33} = 15$</td>
<td>$n_3 = 45$</td>
</tr>
<tr>
<td>4</td>
<td>$n_{41} = 16$</td>
<td>$n_{42} = 24$</td>
<td>$n_{43} = 17$</td>
<td>$n_4 = 57$</td>
</tr>
<tr>
<td>5</td>
<td>$n_{51} = 24$</td>
<td>$n_{52} = 26$</td>
<td>$n_{53} = 21$</td>
<td>$n_5 = 71$</td>
</tr>
<tr>
<td>6</td>
<td>$n_{61} = 20$</td>
<td>$n_{62} = 12$</td>
<td>$n_{63} = 28$</td>
<td>$n_6 = 60$</td>
</tr>
<tr>
<td>7</td>
<td>$n_{71} = 27$</td>
<td>$n_{72} = 13$</td>
<td>$n_{73} = 14$</td>
<td>$n_7 = 54$</td>
</tr>
<tr>
<td>8</td>
<td>$n_{81} = 20$</td>
<td>$n_{82} = 14$</td>
<td>$n_{83} = 27$</td>
<td>$n_8 = 61$</td>
</tr>
</tbody>
</table>

$m = 8$  $g_1 = 150$  $g_2 = 150$  $g_3 = 150$  $n = 450$

We assume that we have $r = 3$ covariables.

The design matrices are chosen to be

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}, \quad C = \begin{pmatrix} C_1 & 0 \\ \vdots & \ddots \\ 0 & C_8 \end{pmatrix}$$

for

$$C_i = \left( 1'_{n_{i1}} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} : 1'_{n_{i2}} \otimes \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} : 1'_{n_{i3}} \otimes \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right), \quad i = 1, \ldots, 8.$$

The parameter matrices, the sampling variance and the covariance for the random effects are

$$B = \begin{pmatrix} 8 & 9 & 10 \\ 11 & 12 & 13 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \sigma_e^2 = 0.16 \quad \text{and} \quad \Sigma_u = \begin{pmatrix} 4.1 & 1.8 & 1.2 & 2.4 \\ 1.8 & 3.6 & 2.4 & 1.4 \\ 1.2 & 2.4 & 6.0 & 2.2 \\ 2.4 & 1.4 & 2.2 & 9.6 \end{pmatrix}.$$
Then, the data are randomly generated from
\[ \text{vec}(Y) \sim N_{pn}(\text{vec}(ABHC + 1_p \gamma'X), \Sigma, I_n), \]
using MATLAB Version 8.3.0.532 (The MathWorks, Inc. USA), where the matrix of covariates \( X \) is randomly generated as uniformly distributed on the interval \([0, 1]\) and then taken as fixed.

The simulations were repeated 500 times using the formulas presented in Theorem 3.2 and the following average estimates were obtained:

\[
\hat{B} = \begin{pmatrix} 8.0226 & 9.0551 & 9.9728 \\ 11.0002 & 11.9997 & 13.0002 \end{pmatrix}, \quad \hat{\gamma} = \begin{pmatrix} 0.9681 \\ 1.9743 \\ 3.0222 \end{pmatrix},
\]

\[
\hat{\Sigma}_u = \begin{pmatrix} 4.1683 & 1.8835 & 1.2804 & 2.5056 \\ 1.8835 & 3.6705 & 2.4316 & 1.4471 \\ 1.2804 & 2.4316 & 5.9460 & 2.1355 \\ 2.5056 & 1.4471 & 2.1355 & 9.3509 \end{pmatrix}.
\]

From the above simulations, we see that these estimates are close to the true values and thus, the proposed estimators support the theoretical results.

7 Concluding remarks

The main task in the present paper has been the prediction of small area means for repeated measures data using the model-based approach under SAE techniques. We have considered longitudinal surveys under simple random sampling without replacement repeated over time whose target population is divided into non-overlapping groups available in all small areas.

In order to address the problem of SAE under these settings, we have proposed a multivariate linear regression model that borrows strength across both small areas and over time. This model accounts for repeated measures data, group individuals and random effect variations over time. The estimation of model parameters has been discussed within a restricted maximum likelihood based approach. The model is split into three component models, some algebraic transformations are performed to achieve the matrix normal distribution of each component thereby follows the derivation of explicit restricted maximum likelihood estimators. Prediction of small area means is presented at all time points, at each time point and by group units. These theoretical results have also been illustrated in a simulation study.

In future work we wish to study properties of all proposed estimators. In particular, the moments and approximation of the distribution of estimators for this SAE multivariate linear regression model and Mean Square Error
estimation. We also wish to have some specific real data set and apply the results of this work.

A Appendix

A.1 Proof of Lemma 3.1

Proof A.1 Obviously, the likelihood equation (7) admits a unique solution for the parameter vector \( \gamma \) if the matrix \( XC'(C')'X' + XR_2R'_2X' \) is of full rank. We have \( \text{rank}(C'(C')') = N - mk \) (with \( N > mk \)) and \( \text{rank}(R_2R'_2) = mk - m \).

Moreover,
\[
C'(C')'R_2R'_2 = R_2R'_2C'(C')' = 0.
\]
So, \( C(C'(C')') \cap C(R_2R'_2) = \{0\} \) and then
\[
\text{rank}(C'(C')' + R_2R'_2) = \text{rank}(C'(C')') + \text{rank}(R_2R'_2) = N - m.
\]
Therefore,
\[
\text{rank}(XC'(C')'X' + XR_2R'_2X') = \text{rank} \left( X(C'(C')' + R_2R'_2)X' \right) = \text{rank}(X)
\]
provided that \( \text{rank}(X) \leq N - m \).

The likelihood equation (8) is equivalent to
\[
A'ABK_2K'_2 = A'YR_2K'_2 - A'1_p\gamma'XR_2K'_2.
\]
This equation in \( B \) admits a non unique solution if and only if one or both matrices \( A \) and \( K_2 \) are not of full rank. Since \( A \) is a full rank matrix, we need to show that the matrix \( K_2 = H(CC')^{1/2}\Gamma_2 \) is not of full rank. It follows from the construction of \( \Gamma_2 \) that \( \text{C}(\Gamma_2) = \text{C} \left( (CC')^{1/2}(CZ')^o \right) \). But for any matrices of proper sizes, if \( \text{C}(F) = \text{C}(G) \), then \( \text{C}(EF) = \text{C}(EG) \) (see for example Harville (1997) for more details). Therefore, \( \text{C} \left( H(CC')^{1/2}\Gamma_2 \right) = \text{C} \left( HCC'(CZ')^o \right) \).

Moreover, if two matrices have the same column space then they have the same rank. Using this result and the following rank formula which can be found, for example in Kollo and von Rosen (2005), \( \text{rank}(F:G) = \text{rank}(F'G') + \text{rank}(G) \), we get
\[
\text{rank}(H(CC')^{1/2}\Gamma_2) = \text{rank}(HCC'(CZ')^o) = \text{rank}(CC'H':CZ') - \text{rank}(CZ').
\]
Since \( Z' = C'Q \) for
\[
Q = \begin{pmatrix}
\frac{1}{\sqrt{N_1}}1_k & 0 \\
\vdots & \ddots & \ddots \\
0 & & \frac{1}{\sqrt{N_m}}1_k
\end{pmatrix},
\]
it follows that
\[
\text{rank}(H(\text{C}\text{C}')^{1/2}\Gamma_2) = \text{rank}(H\text{C}\text{C}'(CZ')^o) \\
= \text{rank}(\text{C}\text{C}'H': CC'Q) - \text{rank}(CC'Q) \\
= \text{rank}(H': Q) - \text{rank}(Q) \\
= \text{rank}(H') + \text{rank}(Q) - \dim(C(H') \cap C(Q)) - \text{rank}(Q) \\
= \text{rank}(H') - \dim(C(H') \cap C(Q)),
\]
since the matrix \( CC' \) is of full rank and where \( \dim \) denotes the dimension of a subspace. It remains to show that \( C(H') \) and \( C(Q) \) are not disjoint.

Let \( v_1 = 1_k \) and \( v_2 = (\sqrt{N_1}, \cdots, \sqrt{N_m})' \) and recall that \( H = (I_k : \cdots : I_k) \).

Then we have \( H'v_1 = 1_{mk} \) and \( Qv_2 = 1_{mk} \). Thus, the two spaces are not disjoint since they include a common vector. This completes the proof of the lemma.

**Acknowledgements**

The research of Innocent Ngaruye is supported by the Swedish International Development and Cooperation Agency (SIDA) in collaboration with the University of Rwanda. Both institutions are acknowledged.

**References**


