# Small Area Estimation for Multivariate Repeated Measures Data

**Innocent Ngaruye**

**Linköping University**
**INSTITUTE OF TECHNOLOGY**

Linköping Studies in Science and Technology.
Licentiate Thesis. No. 1684

**Small Area Estimation for Multivariate Repeated Measures Data**

Innocent Ngaruye

*innocent.ngaruye@liu.se*
*www.mai.liu.se*
*Mathematical Statistics*
*Department of Mathematics*
*Linköping University*
*SE–581 83 Linköping*
*Sweden*

*To my family,*

# Abstract

This thesis considers Small Area Estimation with a main focus on estimation and prediction theory for repeated measures data. The demand for small area statistics is for both cross-sectional and 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. In this work, we consider repeated surveys including the same subjects at different time points. The population from which a sample has been drawn is partitioned into several subpopulations and within all subpopulations there is the same number of group units. For this setting a multivariate linear regression model is formulated. The aim of the proposed model is to borrow strength across small areas and over time with a particular interest of growth profiles over time. The model accounts for repeated surveys, group individuals and random effects variations.

The estimation of model parameters is discussed with a restricted maximum likelihood based approach. The prediction of random effects and the prediction of small area means across time points, per group units and for all time points are derived. The theoretical results have also been supported by a simulation study and finally, suggestions for future research are presented.

# Populärvetenskaplig sammanfattning

I den här avhandling diskuteras *small area estimation* (SAE) med fokus på skattningar av parametrarna samt prediktion för slumpvariabler givet en modell för upprepade mätningar. Efterfrågan på statistisk inferens för undergrupper av en population (*small area statistics*) har ökat markant. Till exempel kan sådan inferens för upprepad mätningar användas av offentliga beslutsfattare vid tilldelning av ekonomiska resurser, planering av nya utbildnings-eller hälsoprogram och i vissa fall kan det vara av intresse att studera en specifik undergrupp av befolkningen.

Det har visat sig att den multivariata framställningen för modellbaserade metoder vid SAE kan uppnå betydande bättre resultat jämfört med det vanliga endimensionella modellantagandet. I detta arbete betraktar vi upprepade undersökningar på samma individer vid olika tidpunkter. Populationen från vilket ett urval har tagits är uppdelat i flera delpopulationer och inom alla undergrupper finns det samma antal observationer. För dessa förutsättningar, är en multivariat linjär regressionsmodell formulerad. Syftet med den förslagna modellen är att låna egenskaper mellan undergrupperna och över tid. I modellen ingår det upprepade undersökningar, grupper av individer samt slumpmässiga effekter.

Parametrarna i den föreslagna modellen skattas med hjälp av *restricted maximum likelihood* metoden. Prediktion av slumpmässiga effekter samt prediktion av de förväntade värdena i undergrupperna beräknas.

# Acknowledgments

*Linköping, November 4, 2014*
*Innocent Ngaruye*

# Contents

# Notation

Throughout this thesis, all vectors are column vectors and are written by lower case letters with bold font. Matrices are written by upper case letters with bold font and scalar variables are written by lower case letters with ordinary font. In general, parameters are written by Greek letters. Note that the same symbol can be used for different purposes. The principal notation is listed below and any deviations from this are explained in the text.

## Symbols and Operators

| | |
|---|---|
| $E[Y]$ | Expectation of random variable $Y$ |
| $E[Y|X]$ | Expectation of random variable $Y$ given the random variable $X$ |
| $\text{Var}[Y]$ | Variance of random variable $Y$ |
| $\text{Cov}[X, Y]$ | Covariance matrix between random variables $X$ and $Y$ |
| $|\boldsymbol{A}|$ | Determinant of a square matrix $\boldsymbol{A}$ |
| $\boldsymbol{A}'$ | Transpose of matrix $\boldsymbol{A}$ |
| $\boldsymbol{I}_n$ | Identity matrix of size $n$ |
| $\boldsymbol{1}_n$ | Column vector of ones of dimension $n$ |
| $\widehat{x}$ | Estimator or estimate of $x$, determined by the context |
| $\text{vec}(\boldsymbol{A})$ | Vectorization of matrix $\boldsymbol{A}$ |
| $\text{tr}\,\boldsymbol{A}$ | Trace of matrix $\boldsymbol{A}$ |
| $\exp\{\boldsymbol{A}\}$ | Exponential of matrix $\boldsymbol{A}$ |
| $\sim$ | denotes "distributed according to" |
| $\cong$ | denotes "approximately equal to" |
| $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ | Kronecker product of matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ |
| $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | Multivariate Normal (Gaussian) distribution of dimension $p$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ |

| | |
|---|---|
| $\mathcal{N}_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ | Matrix Normal distribution of dimension $p \times n$ with mean $\boldsymbol{M}$ and dispersion matrix $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$, which is equivalent to $\mathcal{N}_{pn}(\text{vec}(\boldsymbol{M}), \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma})$ |
| $l(\,\cdot\,)$ | Log-likelihood function |
| $L(\,\cdot\,)$ | Likelihood function |
| $\mathcal{C}(\boldsymbol{A})$ | Column space of matrix $\boldsymbol{A}$ |
| $\mathcal{C}(\boldsymbol{A})^{\perp}$ | Orthogonal complement of column space of matrix $\boldsymbol{A}$ |
| $\boldsymbol{A}^{o}$ | Matrix of full rank spanning $\mathcal{C}(\boldsymbol{A})^{\perp}$ |
| $\boldsymbol{A}'$ | Transpose of matrix $\boldsymbol{A}$ |
| $\boldsymbol{A}^{-1}$ | Inverse of matrix $\boldsymbol{A}$ |
| $\boldsymbol{A}^{-}$ | Generalized inverse of matrix $\boldsymbol{A}$ |
| $\frac{d(f)}{dX}$ | First derivative of the function $f$ with respect to the variable $X$ |
| $\frac{d^{2}(f)}{dX^{2}}$ | Second derivative of the function $f$ with respect to the variable $X$ |

# Abbreviations and Acronyms

| | |
|---|---|
| e.g. | for example |
| i.i.d. | independent identically distributed |
| w.r.t. | with respect to |
| SAE | Small Area Estimation |
| SRSWOR | Simple Random Sampling Without Replacement |
| MSE | Mean Square Error |
| MLE | Maximum Likelihood Estimator |
| RMLE | Restricted Maximum Likelihood Estimator |
| BP | Best Predictor |
| BLP | Best Linear Predictor |
| BLUP | Best Linear Unbiased Predictor |
| EBLUP | Empirical Best Linear Unbiased Predictor |
| BLUE | Best Linear Unbiased Estimator |
| GLS | General Least Squares |
| GC | Growth Curve |
| EGC | Extended Growth Curve |
| GMANOVA | Generalized Multivariate Analysis of Variance |

# 1

# Introduction

THIS thesis is concerned with the problem of *Small Area Estimation* (SAE) which is mainly about how to produce reliable estimates of characteristics of interest, (totals, means, proportions, quantiles, etc.) for small areas or domains based on few samples or even no samples taken from these areas and how to assess the estimation or prediction error.

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. The population surveys are usually designed to provide efficient estimates of parameters of interest for large populations. In most cases these 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.

The aim of this work is to provide an overview and a broad understanding of methods in SAE with a focus put on estimation and prediction of small area characteristics of interest for repeated measures data. Throughout this thesis, we are interested in small area means.

## 1.1   Background and problem formulation

Following the definition given by Rao (2003), the term "*small area*" or "*small domain*" is referred to a subpopulation for which the domain-specific sample is not large enough to produce direct estimates with reliable precision. This subpopulation can be a small geographical area (county, state, district, etc.), a demographic group within a geographical region (specific sex-age group, etc.) or any subdivision of the population. One possible

solution to the estimation problem to improve direct estimates is to "*borrow strength*" from other related data sets, either from similar areas, or relevant "*auxiliary information*" obtained from census or some other administrative records.

In recent years, 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) and 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.

In this work, we consider repeated surveys for the variable of interest on the same subjects at different time points. We assume that the target population from which the sample has been drawn is partitioned into several subpopulations and within all subpopulations, there is the same number of group units. In addition, we also assume that the sampling scheme is a simple random sampling scheme without replacement (SRSWOR) for which an element from the population is not chosen more than once, and the sampling design was not planned to estimate subpopulation level. Hence, direct estimates at this level are not of any high precision. The purpose with this work is to develop reliable estimates for a given subpopulation which we call "small area" by prediction of target quantities in a given above mentioned situation.

## 1.2   Example: Regional estimation of malnutrition

In order to make the formulation of the problem clear, we give a realistic motivating example. In a country, say C, having 8 subnational regions, surveys about child nutrition were carried out every 3 months during the whole year on the same subjects with intention of providing nutritional status of male and female children under 5 years and the factors affecting it. With a SRSWOR scheme, a random sample of children was drawn, and their height and weight were measured. A child is considered as stunting or underweight if she/he has a height which is below a median height for a reference child ($z_1$) or has a weight below the median weight for a reference child, ($z_2$) respectively. Therefore, the variable of interest $y$ is height or weight and direct estimates are available for sampled units. Since these surveys were not designed for regions, we assume that the sample sizes within regions are very small so that these direct estimates have large standard errors and hence are not reliable. We also assume that from census, some auxiliary information $x$

are available for the whole country. The auxiliary information can be the size and monthly expenditure of the household, sex and employment of the head of the household. The aim is to develop the yearly estimates of regional height and weight averages and deduce a yearly regional estimate of malnutrition.

## 1.3   Main contributions

The main contributions of this work are

- model formulation of SAE under a multivariate linear regression model for repeated measures data,

- derivation of explicit estimators of the model parameters, and

- prediction of random effects and small area means.

## 1.4   Thesis outline

Apart from the short Introduction which comprises the background, problem formulation, motivation example and main contributions, this thesis consists of 4 more chapters.

Chapter 2 discusses the general theory about multivariate linear models, specifically about estimation and prediction in linear mixed models, multivariate and matrix normal distribution and Growth Curve modeling. Chapter 3 gives a brief review of methods used in SAE with main focus on prediction of small area means. Chapter 4 is devoted to SAE under the multivariate linear regression model. After model formulation, likelihood based estimators are presented in Section 1 and 2. Section 3 and Section 4 deal with the prediction of random effects, the last two sections treat the prediction of small area means and a simulation study. Finally, Chapter 5 gives some concluding remarks and suggestions of future work.

# 2

# Multivariate linear models

MULTIVARIATE linear models play an important role in the analysis of repeated measures data. This class of models includes linear and mixed linear models mainly used for analysis of variance and regression analysis, the Growth Curve model and the Growth Curve model with random effects which are commonly used for studying between-individual differences and within-individual patterns over time. The theory of multivariate analysis based on the matrix normal distribution has been discussed by many authors, among others, we may cite Srivastava (2002); Kollo and von Rosen (2005); McCulloch and Neuhaus (2008); Muirhead (2009).

## 2.1 Multivariate normal distribution

**Definition 2.1 (Multivariate normal distribution).** An $N$-dimensional random vector $\boldsymbol{y}$ is said to have a $N$-variate normal distribution with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$, or simply that $\boldsymbol{y}$ is $\mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its density function is given by

$$f(\boldsymbol{y}) = (2\pi)^{-N/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})'}.$$

### 2.1.1 Estimation in Mixed linear models

The classical linear model for fixed effects is usually written in the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{2.1}$$

where $\boldsymbol{y}$ is an $N \times 1$ observation vector, $\boldsymbol{X}$ is a $N \times p$ known coefficient matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constants for fixed effects and $\boldsymbol{e}$ is a vector of random errors. It is assumed that $E[\boldsymbol{e}] = \boldsymbol{0}$ and $\text{Var}[\boldsymbol{e}] = \sigma_e^2 \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix.

In summary, the model (2.1) can be written as

$$\boldsymbol{y} \sim (\boldsymbol{X}\boldsymbol{\beta}, \sigma_e^2 \boldsymbol{I}_N),$$

and the normal distribution is often assumed so that

$$\boldsymbol{y} \sim \mathcal{N}_N(\boldsymbol{X}\boldsymbol{\beta}, \sigma_e^2 \boldsymbol{I}_N).$$

A general linear mixed model is obtained by incorporating an $m \times 1$ vector of random effects $\boldsymbol{u}$ and a suitable design matrix $\boldsymbol{Z} : N \times m$ in model (2.1), which yields

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}, \tag{2.2}$$

with assumptions

$$E[\boldsymbol{u}] = E[\boldsymbol{e}] = \boldsymbol{0},$$
$$\mathrm{Cov}[\boldsymbol{u}, \boldsymbol{e}] = \boldsymbol{0},$$
$$\mathrm{Var}[\boldsymbol{y}] = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R} = \boldsymbol{\Sigma},$$

for

$$\boldsymbol{G} = \mathrm{Var}[\boldsymbol{u}], \quad \boldsymbol{R} = \mathrm{Var}[\boldsymbol{e}].$$

It is convenient to notice that under these assumptions, the conditional mean of $\boldsymbol{y}$ given the realized $\boldsymbol{u}$ is $E[\boldsymbol{y}|\boldsymbol{u}] = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$ and so

$$\boldsymbol{y}|\boldsymbol{u} \sim \mathcal{N}_N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \boldsymbol{R}).$$

Under the normality assumption, with known $\boldsymbol{\Sigma}$, the *Generalized Least Squares* (GLS) estimator of $\boldsymbol{\beta}$ corresponds to its *Maximum Likelihood Estimator* (MLE) and is unbiased:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}, \tag{2.3}$$

with $\boldsymbol{X}$ assumed to be of full rank. For unknown $\boldsymbol{\Sigma}$, it is replaced by its estimator which may be obtained by the maximum likelihood approach and then

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y}.$$

### 2.1.2   Prediction of random effects

Hereafter, we present some techniques used to predict the random area effects values in mixed linear models.

**Best Predictor and Best Linear Predictor**

Consider the general linear mixed model (2.2),

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}.$$

The problem is to how predict the random effects $\boldsymbol{u}$. The *best predictor* (BP) is defined as the predictor with *minimum mean square error* (MSE), it is the conditional mean of the predictor given the data vector. That is

$$\widehat{\boldsymbol{u}}_{BP} = E[\boldsymbol{u}|\boldsymbol{y}].$$

If the best predictor $\widehat{\boldsymbol{u}}_{BP}$ is linear in $\boldsymbol{y}$, then it is the *best linear predictor* (BLP). With normality assumption

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{y} \end{pmatrix} \sim \mathcal{N}_{N+m}\Big[ \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{X\beta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{G} & \boldsymbol{GZ}' \\ \boldsymbol{ZG} & \boldsymbol{\Sigma} \end{pmatrix} \Big].$$

The best linear predictor of $\boldsymbol{u}$ is given by

$$\begin{aligned} \widehat{\boldsymbol{u}}_{BLP} =& E[\boldsymbol{u}|\boldsymbol{y}] = E[\boldsymbol{u}] + \mathrm{Cov}[\boldsymbol{u}, \boldsymbol{y}](\mathrm{Cov}[\boldsymbol{y}])^{-1}(\boldsymbol{y} - E[\boldsymbol{y}]) \\ =& \boldsymbol{GZ}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}). \end{aligned}$$

More details about derivation of these predictors can be found for example in Searle et al. (2009).

## Best Linear unbiased Predictor

If the predictor $\widehat{\boldsymbol{u}}$ is the "best", in the sense that it has minimum MSE; if it is "linear" in $\boldsymbol{y}$ and "unbiased", in the sense that $E[\widehat{\boldsymbol{u}}] = E[\boldsymbol{u}] = \boldsymbol{0}$; then it is the *best linear unbiased predictor* (BLUP). In this case, the unknown parameter $\boldsymbol{\beta}$ is replaced by its GLS estimate which is the *best linear unbiased estimator* (BLUE) of $\boldsymbol{\beta X}$, i.e.,

$$\widehat{\boldsymbol{u}}_{BLUP} = \boldsymbol{GZ}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}),$$

where

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}.$$

An alternative way of derivation of the BLUP is to solve the mixed model equations which do not involve the computation of the inverse of the covariance matrix $\boldsymbol{\Sigma}$. These equations were developed by Henderson (1973) and the prediction approach is known as Henderson's method of prediction. The solutions yield simultaneously the BLUE of $\boldsymbol{X\beta}$ and the BLUP of $\boldsymbol{u}$. The set of equations are obtained by maximizing the joint density of $\boldsymbol{y}$ and $\boldsymbol{u}$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{u}$ which is

$$f(\boldsymbol{y}, \boldsymbol{u}) = f(\boldsymbol{y}|\boldsymbol{u})f(\boldsymbol{u}),$$

with

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{u} &\sim \mathcal{N}_N(\boldsymbol{X\beta} + \boldsymbol{Zu}, \boldsymbol{R}), \\ \boldsymbol{u} &\sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{G}). \end{aligned}$$

Henderson maximized the joint density

$$f(\boldsymbol{y}, \boldsymbol{u}) = \frac{\exp\{-\frac{1}{2}\Big[(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu}) + \boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}\Big]\}}{\sqrt{(2\pi)^{N+m}|\boldsymbol{R}||\boldsymbol{G}|}}$$

taking it as it was a likelihood and got the mixed model equations written in matrix form
as

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix}.$$

The solutions of Henderson's mixed model equations are

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y},$$
$$\widehat{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}).$$

Note that in practice, the covariance matrices $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$ are unknown. Then, they are
replaced by their estimates from the observed data which yields *Empirical Best Linear
Unbiased Predictor* (EBLUP):

$$\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{G}}\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}).$$

## 2.2   Matrix normal distribution

**Definition 2.2 (Matrix normal distribution).**   Let $\boldsymbol{Y}$ be an $p \times n$ random matrix, $\boldsymbol{M}$ an
$p \times n$ matrix, $\boldsymbol{\Sigma} : p \times p$ and $\boldsymbol{\Psi} : n \times n$ positive definite matrices. Then the matrix $\boldsymbol{Y}$ has a
matrix normal distribution with mean $\boldsymbol{M}$ and covariances matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ denoted by

$$\boldsymbol{Y} \sim \mathcal{N}_{p,n}\Big(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\Big),$$

or equivalently

$$\mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_{pn}\Big(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}\Big),$$

if its density is given by

$$f(\boldsymbol{Y}) = (2\pi)^{-\frac{pn}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{\Psi}|^{-\frac{p}{2}} e^{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{M})\boldsymbol{\Psi}^{-1}(\boldsymbol{Y}-\boldsymbol{M})'\}},$$

where $\mathrm{vec}(\,\cdot\,)$ stands for the vectorization operator and $\otimes$ is the usual Kronecker product.

The covariance matrix $\boldsymbol{\Sigma}$ is interpreted as the covariance between rows of $\boldsymbol{Y}$ which
is the same for each column and the covariance matrix $\boldsymbol{\Psi}$ is interpreted as the covariance
between columns of $\boldsymbol{Y}$ which is the same for each row.

**Theorem 2.1**
*Suppose that $\boldsymbol{X} \sim \mathcal{N}_{p,n}\Big(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\Big)$. For $\boldsymbol{A} : q \times p$ and $\boldsymbol{B} : n \times m$ set $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$,
then, $\boldsymbol{Y} \sim \mathcal{N}_{q,m}\Big(\boldsymbol{A}\boldsymbol{M}\boldsymbol{B}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}', \boldsymbol{B}'\boldsymbol{\Psi}\boldsymbol{B}\Big).$*

The proof can be found for example in Kollo and von Rosen (2005). It is important to
note that if

$$\boldsymbol{X}_1 \sim \mathcal{N}_{p,n}\Big(\boldsymbol{M}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\Psi}_1\Big),$$

and

$$\boldsymbol{X}_2 \sim \mathcal{N}_{p,n}\Big(\boldsymbol{M}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\Psi}_2\Big),$$

then $\boldsymbol{X}_1 + \boldsymbol{X}_2$ is normally distributed but not necessarily matrix normally distributed. This is because it is not always true that there exist matrices $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Phi}$ such that

$$\boldsymbol{\Psi}_1 \otimes \boldsymbol{\Sigma}_1 + \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Sigma}_2 = \boldsymbol{\Upsilon} \otimes \boldsymbol{\Phi}.$$

See Kollo and von Rosen (2005) for more details.

### 2.2.1   Growth Curve Model

The *Growth Curve* (GC) model also known as *Generalized Multivariate Analysis of Variance* (GMANOVA) model is given by

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{C} + \boldsymbol{E}, \tag{2.4}$$

where $\boldsymbol{Y} : p \times n$ is the observation matrix, $\boldsymbol{B} : q \times k$ is the parameter matrix, $\boldsymbol{A} : p \times q$ is the within individual design matrix indicating the time dependency within the individuals, $\boldsymbol{C} : k \times n$ with rank$(\boldsymbol{C}) + p \leq n$ is the between individual design matrix accounting for group effects and $\boldsymbol{E} : p \times n$ is the error matrix whose columns are independently distributed as a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_e$.

The estimation, testing and model diagnostic problems for GC model and its extensions have been investigated by several authors, see for example, Potthoff and Roy (1964); von Rosen (1989); Khatri (1973); Rao (1958); Nzabanita et al. (2012); Ohlson and Srivastava (2010). Theorem 2.2 gives the MLEs for a classical GC model.

**Theorem 2.2**
*Consider the Growth Curve model defined in equation (2.4). The maximum likelihood estimators of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_e$ are respectively given by*

$$\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{Y}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1},$$
$$n\widehat{\boldsymbol{\Sigma}}_e = (\boldsymbol{Y} - \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C})(\boldsymbol{Y} - \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C})',$$

*where*

$$\boldsymbol{S} = \boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C})\boldsymbol{Y}'$$

*and the design matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ are assumed to have full rank.*

The proof of the theorem can be found for example in Kollo and von Rosen (2005).

### 2.2.2   Random effect Growth Curve Model

Consider the Growth Curve model given in equation (2.4) and suppose that there exist individual random effects $\boldsymbol{U} : p \times r$ assumed to be multivariate normal distributed with

mean zero and covariance matrix $\boldsymbol{\Sigma}_u$. The random effects Growth Curve model is defined as

$$\boldsymbol{Y} = \boldsymbol{ABC} + \boldsymbol{UZ} + \boldsymbol{E}, \tag{2.5}$$

where $\boldsymbol{Z}$ is a known design matrix for random effects. This class of model has been considered by many authors and estimation of parameters of interest has been discussed with different choices of $\boldsymbol{C}$ and $\boldsymbol{Z}$. See for example Nummi (1997); Ip et al. (2007); Lange and Laird (1989); Yokoyama and Fujikoshi (1993); Yokoyama (2001); Srivastava and Singull (2012) for more details.

### 2.2.3 Extended Growth Curve Model

The classical GC model defined in (2.4) relies on the assumption of the same profile of different individuals. If this does not hold, a more general model, sometimes called *sum of profiles* can be introduced. That is an *Extended Growth Curve* (EGC) model. The following EGC model is an extension of the GC model discussed by Filipiak and von Rosen (2012).

**Definition 2.3 (Extended growth curve model).** The Extended Growth Curve Model with nested subspace condition $\mathcal{C}(\boldsymbol{A}_i) \subseteq \mathcal{C}(\boldsymbol{A}_{i-1})$, $i = 2, 3, \ldots, m$, where $\mathcal{C}(\,\cdot\,)$ denotes the column space of a matrix, is given by

$$\boldsymbol{Y} = \sum_{i=1}^{m} \boldsymbol{A}_i \boldsymbol{B}_i \boldsymbol{C}_i + \boldsymbol{E},$$

where $\boldsymbol{Y} : p \times n$ is a data matrix, $\boldsymbol{A}_i : p \times q_i$, $\boldsymbol{C}_i : k_i \times n$ with $\mathrm{rank}(\boldsymbol{C}_1) + p \leq n$, $i = 1, 2, \ldots, m$ are design matrices, $\boldsymbol{B}_i : q_i \times k_i$ are unknown parameters and columns of $\boldsymbol{E}$ are assumed to be independently distributed as multivariate normal with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}$.

In summary,

$$\boldsymbol{Y} \sim N_{p,n}\left( \sum_{i=1}^{m} \boldsymbol{A}_i \boldsymbol{B}_i \boldsymbol{C}_i, \boldsymbol{\Sigma}, \boldsymbol{I}_n \right),$$

and when $m = 1$, the model reduces to the classical Growth Curve model.

#### Maximum Likelihood Estimators

The derivation of MLEs for the EGC model as defined in Definition 2.3 can be found in Filipiak and von Rosen (2012). Here we present the MLEs for EGC model with two terms ($m = 2$) which will be used later on in Chapter 4. In the following, we use the notation $\boldsymbol{A}^o$ for any matrix of full rank spanning $\mathcal{C}(\boldsymbol{A})^\perp$, i.e. $\mathcal{C}(\boldsymbol{A}^o) = \mathcal{C}(\boldsymbol{A})^\perp$. Moreover, $\boldsymbol{A}^-$ denotes an arbitrary generalized inverse of the matrix $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$. We also denote by $\boldsymbol{P_A} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$, $\boldsymbol{Q_A} = \boldsymbol{I} - \boldsymbol{P_A}$, $\boldsymbol{P_{A,B}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{A})^-\boldsymbol{A}'\boldsymbol{B}$ and $\boldsymbol{Q_{A,B}} = \boldsymbol{I} - \boldsymbol{P_{A,B}}$.

**Theorem 2.3**

*Consider the EGC model given in Definition (2.3). Set $m = 2$ and assume that $\mathcal{C}(\boldsymbol{A}_2) \subseteq \mathcal{C}(\boldsymbol{A}_1)$. The maximum likelihood estimators for the parameter matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are given by*

$$
\begin{aligned}
\widehat{\boldsymbol{B}}_2 &= (\boldsymbol{A}_2'\boldsymbol{S}_1^{-1}\boldsymbol{A}_2)^{-}\boldsymbol{A}_2'\boldsymbol{S}_1^{-1}\boldsymbol{Y}\boldsymbol{Q}_{\boldsymbol{C}_1'}\boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{Q}_{\boldsymbol{C}_1'}\boldsymbol{C}_2')^{-} \\
&\quad + (\boldsymbol{A}_2')^o\boldsymbol{T}_{21}\boldsymbol{C}_2\boldsymbol{Q}_{\boldsymbol{C}_1'} + \boldsymbol{A}_2'\boldsymbol{T}_{22}(\boldsymbol{C}_2\boldsymbol{Q}_{\boldsymbol{C}_1'})^o \\
\widehat{\boldsymbol{B}}_1 &= (\boldsymbol{A}_1'\boldsymbol{S}_2^{-1}\boldsymbol{A}_1)^{-}\boldsymbol{A}_1'\boldsymbol{S}_2^{-1}(\boldsymbol{Y} - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-} \\
&\quad + (\boldsymbol{A}_1')^o\boldsymbol{T}_{11}\boldsymbol{C}_1' + \boldsymbol{A}_1'\boldsymbol{T}_{12}\boldsymbol{C}_1^{o'}, \\
n\widehat{\boldsymbol{\Sigma}} &= (\boldsymbol{Y} - \boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)(\boldsymbol{Y} - \boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)',
\end{aligned}
$$

*where*

$$
\begin{aligned}
\boldsymbol{S}_1 &= \boldsymbol{Y}\boldsymbol{Q}_{(\boldsymbol{C}_1':\boldsymbol{C}_2')}\boldsymbol{Y}', \\
\boldsymbol{S}_2 &= \boldsymbol{S}_1 + \boldsymbol{Q}_{\boldsymbol{A}_2,\boldsymbol{S}_1^{-1}}\boldsymbol{Y}\boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_1'}\boldsymbol{C}_2'}\boldsymbol{Y}'\boldsymbol{Q}'_{\boldsymbol{A}_2,\boldsymbol{S}_1^{-1}},
\end{aligned}
$$

*and $\boldsymbol{T}_{ij}, i,j = 1,2$, are arbitrary matrices.*

Because of arbitrary matrices $\boldsymbol{T}_{ij}, i,j = 1,2$, the maximum likelihood estimators for the parameter matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are not unique. However, the estimated mean

$$
\widehat{E[\boldsymbol{Y}]} = \boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \boldsymbol{A}_2\widehat{\boldsymbol{B}}_1\boldsymbol{C}_2
$$

is always unique. The maximum likelihood estimators for the parameter matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are unique if the following conditions

$$
\text{rank}(\boldsymbol{C}_1) = k_1, \quad \text{rank}(\boldsymbol{A}_1) = q_1, \quad \mathcal{C}(\boldsymbol{C}_1) \cap \mathcal{C}(\boldsymbol{C}_2) = \{\boldsymbol{0}\},
$$

and

$$
\text{rank}(\boldsymbol{C}_2) = k_2, \quad \text{rank}(\boldsymbol{A}_2) = q_2, \quad \mathcal{C}(\boldsymbol{C}_1) \cap \mathcal{C}(\boldsymbol{C}_2) = \{\boldsymbol{0}\},
$$

hold, respectively. In such a case, the maximum likelihood estimators for the parameter matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ and $\boldsymbol{\Sigma}$ are given by

$$
\begin{aligned}
\widehat{\boldsymbol{B}}_2 &= (\boldsymbol{A}_2'\boldsymbol{S}_1^{-1}\boldsymbol{A}_2)^{-1}\boldsymbol{A}_2'\boldsymbol{S}_1^{-1}\boldsymbol{Y}\boldsymbol{Q}_{\boldsymbol{C}_1'}\boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{Q}_{\boldsymbol{C}_1'}\boldsymbol{C}_2')^{-1}, \\
\widehat{\boldsymbol{B}}_1 &= (\boldsymbol{A}_1'\boldsymbol{S}_2^{-1}\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1'\boldsymbol{S}_2^{-1}(\boldsymbol{Y} - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-1}, \\
n\widehat{\boldsymbol{\Sigma}} &= (\boldsymbol{Y} - \boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)(\boldsymbol{Y} - \boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)',
\end{aligned}
$$

where $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are defined in Theorem (2.3) (see Filipiak and von Rosen (2012) for more details).

# 3

# Methods in Small Area Estimation

M ETHODS used in Small Area Estimation are mainly divided into "design-based" and "model-based" methods. For the first category, the inference is fully based on the used sampling design and for the latter methods, also called "model-dependent", the inference is involved with statistical methods based on the frequentist or Bayesian approaches or the combination of the two (Pfeffermann 2002; Rahman 2008). Both methods use auxiliary information to "borrow strength" from related neighboring areas, from censuses, surveys or registers. Throughout this chapter, we suppose that the sampling design for the population as a whole does not correspond well to the divisions in the population. Therefore, the samples sizes for subpopulations may become small which leads to the small area estimation problem.

## 3.1   Design-based methods

Design-based estimation or *randomization* approach in SAE is based on traditional probability sampling theory. One can refer to Fuller (2009) for probability sampling theory in statistics. Under this estimation approach, the randomness is only induced by the sampling design used to select the sample with population measurements regarded as fixed (e.g., Lehtonen and Veijanen 2009; Pfeffermann 2002). These methods make use of survey weights and associated statistical inferences are based on the sample selection probabilities. Two types of estimators are derived according to the related data sources used; either direct estimators when only domain-specific data are used or indirect estimators if the estimation procedure "borrows strength" from related areas. A common feature of design-based estimators is that there is no explicit model assumptions used for their derivation, and the variance and bias are calculated under the randomization distribution.

### 3.1.1   Direct estimators

The direct estimators of small area means under a design-based approach is often based on estimation of population means in classical probability theory by using for example Horvitz-Thompson form of direct estimator (see Cochran 1977; Särndal et al. 1992).

Let $U$ be a finite population of size $N$ consisting on $m$ disjoint subpopulations or small areas $U_i$ each with population size $N_i, (i = 1, \ldots, m)$ such that $U = \bigcup_{i=1}^{m} U_i$ and $N = \sum_{i=1}^{m} N_i$.

We assume that a sample $s$ of size $n$ is selected by SRSWOR so that we have equal selection probabilities $\pi_j = \frac{n}{N}$ and sampling weights $w_j = \frac{1}{\pi_j} = \frac{N}{n}; \; j \in s$.

Let $Y$ denote the characteristic of interest, $y_{ij}$ the outcome value of the $j^{th}$ population unit coming from the small area $i$ (with $i = 1, \ldots, m; \; j = 1, \ldots, N_i$) and $s_i$ be the corresponding sample of size $n_i$ taken from the small area $i$ such that $s = s_1 \cup \cdots \cup s_m$; $n = \sum_{i=1}^{m} n_i$ and $w_{ij} = \frac{N_i}{n_i}$ .

Then, if there is no auxiliary information available, the true area mean $\overline{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$ can be estimated using Horvitz-Thompson form:

$$\widehat{\overline{Y}}_i = \frac{1}{N_i} \sum_{j=1}^{n_i} w_{ij} y_{ij} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} = \overline{y}_i.$$

Its variance is given by

$$\mathrm{Var}[\widehat{\overline{Y}}_i] = (1 - f_i) \frac{S_i^2}{n_i},$$

with $f_i = \frac{n_i}{N_i}$, where $(1 - f_i)$ is the finite population correction factor and

$$S_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (y_{ij} - \overline{y}_i)^2, \; N_i \geq 2.$$

It follows that for small $n_i$, the variance will be larger unless the variability of the $y$-values is sufficiently small. In order to reduce the variance, we suppose in addition that we have $r$ auxiliary variables known for every sample units. Denote by $\boldsymbol{x}_{ij}$, the $r$-vector of covariables for the $j^{th}$ unit in area $i$ and $\overline{\boldsymbol{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$ the corresponding sample mean. We assume that the population mean $\overline{\boldsymbol{X}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{x}_{ij}$ are also known. Then, a more efficient design-based estimator is the *regression estimator* given by

$$\overline{Y}_i^{reg} = \overline{y}_i + (\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{x}}_i)' \boldsymbol{\beta}_i,$$

which variance equals

$$\mathrm{Var}[\overline{Y}_i^{reg}] = (1 - \rho_i^2)(1 - f_i) \frac{S_i^2}{n_i} = (1 - \rho_i^2) \mathrm{Var}[\widehat{\overline{Y}}_i],$$

where $\boldsymbol{\beta}_i$ and $\rho_i$ are respectively the vector of regression coefficients and the multiple correlation between the survey variable $Y$ and auxiliary variables $\boldsymbol{x}_{ij}$ in area $i$.

We note that the use of auxiliary information reduces the variance by the factor $(1 - \rho_i^2)$, the reason why it increases the prediction power in SAE. However, in practice the regression coefficient $\boldsymbol{\beta}_i$ is replaced by its ordinary least squares estimator from the sample $s_i$ which may not be effective because of a small sample size.

### 3.1.2   Indirect estimators

By assuming that all small areas are similar with respect to the quantity to be estimated, an estimator obtained from a large domain covering several small areas can be used to derive an indirect estimator for separate areas comprising that domain (Gonzalez, 1973). It is equivalent to put an assumption on $\boldsymbol{\beta}_i$ and on intercepts $(\overline{Y}_i - \overline{\boldsymbol{X}}_i' \boldsymbol{\beta}_i)$ to be similar across the small areas. An effective *synthetic regression estimator* is obtained by

$$\overline{Y}_{syn,i}^{reg} = \overline{y}_i + (\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{x}}_i)' \widehat{\boldsymbol{\beta}},$$

where $\widehat{\boldsymbol{\beta}}$ is the sample estimator computed using data from all samples $s_i$. The synthetic estimator is more efficient if the assumption of homogeneity within the larger domain holds, however even though it reduces the variance, it can lead to severe biases if there is a strong individual effect on the regression coefficient.

The bias reduction is then improved by the use of a *composite estimator* which is a weighted sum of the area direct estimator. Let $\theta_i$ be the small area characteristic of interest, the composite estimator has the general form

$$\widehat{\theta}_{com,i} = \phi_i \widehat{\theta}_i + (1 - \phi_i) \widehat{\theta}_{syn,i},$$

where $\widehat{\theta}_i$ is the direct estimator, $\widehat{\theta}_{syn,i}$ is the synthetic estimator and $\phi_i = f_i - n_i/N_i$ is a suitable weight chosen to minimize MSE.

A composite estimator has small or no bias but large variance. The more weight is given to the direct unbiased estimator as the sampling fraction $f_i$ increases. However, the sampling fractions are usually very small and thus the use of this weight in practice implies the use of the synthetic estimator.

## 3.2   Model-based methods

In recent years, model-based approaches in SAE have received much attention due to their usefulness for estimating small area characteristics. For example model-based methods are now being extensively used to find indirect estimates for small area means. As stated previously, even though the design-based estimators are simple to implement and can provide efficient estimators not dependent on an assumed model, they can lead to severe biases if the assumption of homogeneity within larger domains is violated. Furthermore, they are relying on assumption of similarity of all areas with respect to the variable of interest and do not account for area specific variability. The model-based methods are based on the use of explicitly linking of a model for incorporating random area effects to overcome these underlying problems. However, model-based methods can lead to severe bias if the assumed model is not correct.

### 3.2.1   Prediction in finite populations

Model-based sampling theory applied to finite population treats the problem of estimating finite population characteristics of interest as a prediction problem. For example, the estimation of a finite population mean from the sample returns to the prediction of a mean of non-sampled values. Model-based methods use the superpopulation approach which considers the finite population as a random sample from a larger population (e.g. see, Bolfarine and Zacks, 1992). The observed population vector $\boldsymbol{y} = (y_1, \ldots, y_N)'$ is considered as a realization of a random variable characterized by a model, say $\varphi$. Under the superpopulation approach, the finite quantities $\theta$ to be estimated are random and then we do not estimate them, but predict them.

By denoting $s$ and $r$ the sampled and remainder of the finite population $U$ so that $U = s \cup r$ and $\boldsymbol{y} = (\boldsymbol{y}_s', \boldsymbol{y}_r')'$, we can express the quantities of interest $\theta = \boldsymbol{l}'\boldsymbol{y}$ by

$$\theta = \theta_s + \theta_r,$$

or equivalently

$$\theta = \boldsymbol{l}_s'\boldsymbol{y}_s + \boldsymbol{l}_r'\boldsymbol{y}_r,$$

where $\boldsymbol{l} = (\boldsymbol{l}_s', \boldsymbol{l}_r')'$. Therefore, the predictor $\widehat{\theta}$ of $\theta$ is given by

$$\widehat{\theta} = \theta_s + \widehat{\theta}_r,$$

where $\widehat{\theta}_r$ is the predictor of $\theta_r$.

The predictor $\widehat{\theta}$ is $\varphi$-unbiased if

$$E_\varphi[\widehat{\theta} - \theta] = 0, \text{ for the given model } \varphi,$$

where $E_\varphi[\,\cdot\,]$ denotes the expectation with respect to the superpopulation model. The MSE of $\widehat{\theta}$ is given by

$$E_\varphi[\widehat{\theta} - \theta]^2 = \text{Var}_\varphi[\widehat{\theta} - \theta] + \left(E_\varphi[\widehat{\theta} - \theta]\right)^2,$$

and if $\widehat{\theta}$ is unbiased with respect to $\varphi$, then

$$\text{Var}_\varphi[\widehat{\theta} - \theta] = E_\varphi[\widehat{\theta} - \theta]^2,$$

where $\text{Var}_\varphi[\,\cdot\,]$ stands for the variance with respect to the superpopulation model.

Model-based methods in SAE largely use linear mixed models involving suitable auxiliary variables in the fixed part of the model and random area-specific effects. These area-specific effects account for between-variations in data that are not explained by auxiliary variables. Commonly used models in SAE are area-level model and unit-level models (Rao, 2003). In the following two subsections, we present a general estimation and prediction overview about area-level and unit-level models and keep the same population settings as defined in Section 1 of this chapter. In the last subsection we present a version of a mixed linear model for panel data with specific population settings.

### 3.2.2  Area level model

The area level model is a widely used small area model originally presented by Fay III and Herriot (1979) for prediction of mean per capital income in small geographical areas within counties. It is used when the values $\boldsymbol{x}_i$ for auxiliary variables are only available at the area level. This model consists of two models, one is the sampling model comprising the direct estimates and the sampling error, the other one is the linking model relating the population value to some auxiliary variables with unknown random area effects. Suppose that we are interested in prediction of unknown small area means $\overline{Y}_i$. Denote by $\widetilde{\theta}_i$ the direct estimate of $\theta_i$ with $\theta_i = f(\overline{Y}_i)$ some function of small area means. The sampling model and linking model are respectively defined as

$$\widetilde{\theta}_i = \theta_i + e_i, \tag{3.1}$$
$$\theta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i,$$

where $e_i$ are sampling errors assumed to be independent with $E(e_i|\theta_i) = 0$ and known $\mathrm{Var}(e_i|\theta_i) = \sigma_i^2$. The $u_i$'s are uncorrelated area-specific random effects assumed to be independent with $E(u_i) = 0$ and $\mathrm{Var}(u_i) = \sigma_u^2$. Further, $u_i$ and $e_i$ are assumed to be independent.  Combining the two models of (3.1) leads to the area level linear mixed model

$$\widetilde{\theta}_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i + e_i.$$

From the prediction of random effects in linear mixed model under normality discussed in Chapter 2, the BLUP of $u_i$ is given by

$$\widehat{u}_i = E[u_i|\widetilde{\theta}_i] = E[u_i] + \mathrm{Cov}[\widetilde{\theta}_i, u_i](\mathrm{Cov}[\widetilde{\theta}_i])^{-1}(\widetilde{\theta}_i - E[\widetilde{\theta}_i])$$
$$= \frac{\sigma_u^2}{\sigma_i^2 + \sigma_u^2}(\widetilde{\theta}_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}).$$

Put $\gamma_i = \frac{\sigma_u^2}{\sigma_i^2 + \sigma_u^2}$, for known $\sigma_i^2$ and $\sigma_u^2$, the BLUP is then given by

$$\widehat{\theta}_i = \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}} + \widehat{u}_i = \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}} + \gamma_i(\widetilde{\theta}_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}) = \gamma_i\widetilde{\theta}_i + (1 - \gamma_i)\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}},$$

where $\widehat{\boldsymbol{\beta}}$ is the GLS estimator of $\boldsymbol{\beta}$. In practice the variances $\sigma_i^2$ and $\sigma_u^2$ are unknown and are replaced by their sample estimates which yields the corresponding EBLUP.

### 3.2.3  Unit level model

Unit level Model also called "*Nested error unit regression model*" is used when the values $\boldsymbol{x}_{ij}$ of auxiliary variables are known for every unit $j$ in the sample from small area $i$ and the true area means $\overline{\boldsymbol{X}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\boldsymbol{x}_{ij}$ are also known. It was proposed by Battese et al. (1988). The model is defined as

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + u_i + \varepsilon_{ij}, \tag{3.2}$$

where $y_{ij}$ denotes the value of variable of interest $Y$; the random effect $u_i$ are the effect of area characteristic that are not accounted for by the auxiliary variables and $\varepsilon_{ij}$ the

individual random error. The two error terms $u_i$ and $\varepsilon_{ij}$ are assumed to be mutually independent with zero means and variances $\sigma_u^2$ and $\sigma_e^2$. Normality of $u_i$ and $\varepsilon_{ij}$ is often assumed to hold. Under this model, the true small area means for the variable of interest $y$ are given by

$$\overline{Y}_i = \overline{\boldsymbol{X}}_i'\boldsymbol{\beta} + u_i + \overline{\varepsilon}_{ij} = \mu_i + \overline{\varepsilon}_{ij} \cong \mu_i,$$

where $\mu_i = \overline{\boldsymbol{X}}_i'\boldsymbol{\beta} + u_i$ and $\overline{\varepsilon}_{ij} = \frac{1}{N_i} \sum_{j=1}^{N_i} \varepsilon_{ij} \cong 0$ for sufficiently large $N_i$.

For known $\sigma_i^2$ and $\sigma_u^2$, the BLUP for the small area means $\overline{Y}_i$ is expressed as

$$\begin{aligned}
\widehat{\mu}_i &= \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} + \widehat{u}_i \\
&= \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} + \gamma_i(\overline{y}_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}) \\
&= (1 - \gamma_i)\boldsymbol{X}_i'\widehat{\boldsymbol{\beta}} + \gamma_i\left[\overline{y}_i + (\boldsymbol{X}_i - \boldsymbol{x}_i)'\widehat{\boldsymbol{\beta}}\right],
\end{aligned}$$

where $\widehat{\boldsymbol{\beta}}$ is the GLS estimator of $\boldsymbol{\beta}$, $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2/n_i}$ and the BLUP of $u_i$ is similarly obtained following the prediction of random effects discussed in Chapter 2 (Henderson, 1975).

### 3.2.4   Unit level linear mixed model for panel data

Some studies about longitudinal surveys in small area estimation have been carried out. For example, Ferrante and Pacei (2004) considered rotating samples and derived two-stage longitudinal composite and one-stage longitudinal composite estimators. Fabrizi et al. (2007) present EBLUPs for unit level models using panel survey data for both design-based and model-based SAE methods. Singh and Sisodia (2011) have developed the design-based estimates of population mean for small areas in longitudinal surveys. In a PhD thesis, Nissinen (2009) gives an overview of small area estimation from unit-level survey data collected with a panel or a rotating panel design and presents the model based-estimators of small area total and its MSE. In The EURAREA Consortium (2004) project, section C4, various models that borrow strength across time and space are discussed.

We consider the population $U$ of size $N$ divided into $m$ non-overlapping small areas. Let $Y$ be the survey characteristic of interest and assume that the observations on $Y$ are recorded for selected units $n$ over different time points $t = 1, \ldots, p$ so that $\sum_{i=1}^{m} \sum_{t=1}^{p} N_{it} = N$, where $N_{it}$ stands for size of the population $U_{it}$ of the $i$th area at time $t$. We define the longitudinal population of each area $i$ as

$$U_i = U_{i1} \cup U_{i2} \cup \cdots \cup U_{ip}$$

of size

$$N_i = \sum_{t=1}^{p} N_{it}$$

and define the cross-sectional population at each time $t$

$$U_t = U_{1t} \cup U_{2t} \cup \cdots \cup U_{mt}$$

of size

$$N_t = \sum_{i=1}^{m} N_{it}.$$

The overall size of population is then given by

$$\sum_{i=1}^{m} N_i + \sum_{t=1}^{p} N_t = N.$$

We also assume that the sampled units change at each time occasion and denote by $s_{it}$ and $r_{it}$, respectively the sample and remainder of the finite population $U_{it}$ of the area $i$ at time $t$ so that $U_{it} = s_{it} + r_{it}$.

If we suppose that there is no unit effect and the correlation between observations from the same small area is kept constant across different time points, the linear mixed model for a random observation $y_{ijt}$ for unit $j$ in small area $i$ at time $t$ can be defined as

$$y_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + u_i + v_t + e_{ijt}, \quad i = 1, \ldots, m; t = 1, \ldots, p; j = 1, \ldots, N_{it}, \quad (3.3)$$

where $u_i$ and $v_t$ are area effects and time effects assumed to be mutually independent and to follow normal distributions with mean zero and variances $\sigma_u^2$ and $\sigma_v^2$. Furthermore, the row vector $\boldsymbol{x}'_{ijt}$ represents the $r$-covariates (or auxiliary variables), the vector $\boldsymbol{\beta}$ stands for the regression coefficients of $\boldsymbol{x}$ variables and $e_{ijt}$ is the error term assumed to be mutually independent of $u_i$ and $v_t$ and follows the normal distribution with mean zero and variance $\sigma_e^2$.

Put

$$\boldsymbol{u} = [u_1, \cdots, u_m, v_1, \cdots, v_p]',$$
$$\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2),$$

where

$$\boldsymbol{Z}_1 = \text{diag}(\boldsymbol{1}_{N_1}, \cdots, \boldsymbol{1}_{N_m}),$$
$$\boldsymbol{Z}_2 = \begin{pmatrix} \boldsymbol{Z}_{21} \\ \vdots \\ \boldsymbol{Z}_{2m} \end{pmatrix},$$

for $\boldsymbol{Z}_{2i} = \text{diag}(\boldsymbol{1}_{N_{i1}}, \cdots, \boldsymbol{1}_{N_{ip}})$ and $\boldsymbol{1}_{N_{it}}$ a vector of ones of dimension $N_{it}$. The model (3.3) becomes

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}, \quad (3.4)$$

where

$$\text{Cov}(\boldsymbol{y}) = \boldsymbol{\Sigma} = \frac{\sigma_u^2}{\sigma_e^2} \boldsymbol{Z}_1 \boldsymbol{Z}'_1 + \frac{\sigma_v^2}{\sigma_e^2} \boldsymbol{Z}_2 \boldsymbol{Z}'_2 + \boldsymbol{I}_N.$$

We decompose the model (3.4) into a sample and a remainder parts

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_s \\ \boldsymbol{y}_r \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_s \\ \boldsymbol{X}_r \end{bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{ss} & \boldsymbol{\Sigma}_{sr} \\ \boldsymbol{\Sigma}_{rs} & \boldsymbol{\Sigma}_{rr} \end{bmatrix},$$

where the subscripts $s$ and $r$ corresponding to the sample and non-sample population units. The corresponding model for the sample values is given by

$$\boldsymbol{y}_s = \boldsymbol{X}_s \boldsymbol{\beta} + \boldsymbol{Z}_s \boldsymbol{u} + \boldsymbol{e}_s \text{ with } \boldsymbol{\Sigma}_{ss} = \frac{\sigma_u^2}{\sigma_e^2} \boldsymbol{Z}_{s1} \boldsymbol{Z}_{s1}' + \frac{\sigma_v^2}{\sigma_e^2} \boldsymbol{Z}_{s2} \boldsymbol{Z}_{s2}' + \boldsymbol{I}_n,$$

where $n$ denotes the overall number of sampled units across all time points and all areas, $\boldsymbol{Z}_{s1}$ and $\boldsymbol{Z}_{s2}$ are corresponding sample versions of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ respectively. More details about model formulation and estimation can be found for example in The EURAREA Consortium (2004).

For known covariance matrix $\boldsymbol{\Sigma}$, the MLE for $\boldsymbol{\beta}$ corresponds to its GLS estimator

$$\widehat{\boldsymbol{\beta}} = \left( \boldsymbol{X}_s' \boldsymbol{\Sigma}_{ss}^{-1} \boldsymbol{X}_s \right)^{-1} \boldsymbol{X}_s' \boldsymbol{\Sigma}_s^{-1} \boldsymbol{y}_s,$$

as mentioned in Section 2.1.1 and referring to the Section 2.1.2, the BLUP of $\boldsymbol{u}$ is given by

$$\widehat{\boldsymbol{u}} = \boldsymbol{G} \boldsymbol{Z}_s' \boldsymbol{\Sigma}_{ss}^{-1} \left( \boldsymbol{y}_s - \boldsymbol{X}_s \widehat{\boldsymbol{\beta}} \right),$$

where

$$\boldsymbol{G} = \begin{pmatrix} \frac{\sigma_u^2}{\sigma_e^2} \boldsymbol{I}_m & \boldsymbol{0} \\ \boldsymbol{0} & \frac{\sigma_v^2}{\sigma_e^2} \boldsymbol{I}_p \end{pmatrix}.$$

In this situation, the variance components $\sigma_u^2$ and $\sigma_v^2$ for random effects are assumed to be known. However, in practice, they are unknown and have to be estimated from the sample data. They can be estimated using the method of moments, maximum likelihood, restricted maximum likelihood or Henderson's method, among others. More details can be found for example in The EURAREA Consortium (2004), Nissinen (2009), Searle et al. (2009) and González-Manteiga et al. (2008).

Following the theory of prediction presented in Chapter 2, the BLUP $\widehat{\delta}$ of the linear target parameter

$$\delta = \boldsymbol{l}_s' \boldsymbol{y}_s + \boldsymbol{l}_r' \boldsymbol{y}_r,$$

is obtained by minimizing the MSE under the assumption that $\widehat{\delta}$ is model unbiased and is given as

$$\widehat{\delta} = \boldsymbol{l}_s' \boldsymbol{y}_s + \boldsymbol{l}_r' \widehat{\boldsymbol{y}}_r,$$

where

$$\widehat{\boldsymbol{y}}_r = \boldsymbol{X}_r \widehat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{rs} \boldsymbol{\Sigma}_{ss}^{-1} \left( \boldsymbol{y}_s - \boldsymbol{X}_s \widehat{\boldsymbol{\beta}} \right).$$

For small area mean at time $t$, $\delta = \overline{Y}_{it}$, the BLUP will have the form

$$\widetilde{\overline{Y}}_{it} = \frac{1}{N_{it}} \Big( \sum_{j \in s_{it}} y_{ijt} + \sum_{j \in r_{it}} \widetilde{y}_{ijt} \Big),$$

for

$$\widetilde{y}_{ijt} = \boldsymbol{x}'_{ijt} \widehat{\boldsymbol{\beta}} + \widehat{u}_i + \widehat{v}_t,$$

where $\widehat{u}_i$ and $\widehat{v}_t$ the BLUPs of the area effect and time effect, respectively, are obtained from

$$\widehat{\boldsymbol{u}} = \boldsymbol{G} \boldsymbol{Z}'_s \boldsymbol{\Sigma}_{ss}^{-1} \Big( \boldsymbol{y}_s - \boldsymbol{X}_s \widehat{\boldsymbol{\beta}} \Big).$$

---

**Example 3.1**

Suppose that surveys are repeated 3 times on the survey variable $y$ from a finite population having 4 disjoint regions. Suppose in addition that there are 2 auxiliary variables associated to $y$ which values are available for all units in the sample and the true regional means for these variables are also known (see Table 3.1 for population and sample sizes). Therefore,

**Table 3.1:** *Population and Sample sizes*

| Area (region) | Population sizes | | | | Sample sizes | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| $i$ | $N_{it}$ | | | | $n_{it}$ | | | |
| | $t=1$ | $t=2$ | $t=3$ | Total $(N_i)$ | $t=1$ | $t=2$ | $t=3$ | Total $(n_i)$ |
| 1 | $N_{11}$ | $N_{12}$ | $N_{13}$ | $N_1.$ | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_1.$ |
| 2 | $N_{21}$ | $N_{22}$ | $N_{23}$ | $N_2.$ | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_2.$ |
| 3 | $N_{31}$ | $N_{32}$ | $N_{33}$ | $N_3.$ | $n_{31}$ | $n_{32}$ | $n_{33}$ | $n_3.$ |
| 4 | $N_{41}$ | $N_{42}$ | $N_{43}$ | $N_4.$ | $n_{41}$ | $n_{42}$ | $n_{43}$ | $n_4.$ |
| Total $(N_t)$ | $N._1$ | $N._2$ | $N._3$ | $N$ | $n._1$ | $n._2$ | $n._3$ | $n$ |

The linear mixed model for a random observation $y_{ijt}$ of unit $j$ coming from area $i$ at time $t$ is assumed:

$$y_{ijt} = \beta_0 + \boldsymbol{x}'_{ijt} \boldsymbol{\beta}_t + u_i + v_t + e_{ijt}, \quad i = 1, \ldots, 4 \ \ j = 1, \ldots, N \ \ t = 1, 2, 3,$$

where $\beta_0$ is the intercept and $\boldsymbol{x}'_{ijt} = [\boldsymbol{x}'_{1ijt}, \boldsymbol{x}'_{2ijt}]$. The response vector for all time points is given by

$$\boldsymbol{y}_{ij} = \begin{pmatrix} y_{ij1} \\ y_{ij2} \\ y_{ij3} \end{pmatrix},$$

with model equation

$$\boldsymbol{y}_{ij} = \boldsymbol{X}_{ij} \boldsymbol{\beta} + \boldsymbol{1} u_i + \boldsymbol{v} + \boldsymbol{e}_{ij},$$

where

$$X_{ij} = \begin{pmatrix} x'_{ij1} \\ x'_{ij2} \\ x'_{ij3} \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad e_{ij} = \begin{pmatrix} e_{ij1} \\ e_{ij2} \\ e_{ij3} \end{pmatrix}.$$

Collecting vectors $y_{ij}$ for all units yields

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iN_i} \end{pmatrix},$$

with model equation

$$y_i = X_i \beta + z_i u_i + Z_{2i} v + e_i,$$

where

$$X_i = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{iN_i} \end{pmatrix}, \quad z_i = \mathbf{1}_{N_i}, \quad Z_{2i} = \begin{pmatrix} \mathbf{1}_{N_{i1}} & & \\ & \mathbf{1}_{N_{i2}} & \mathbf{0} \\ \mathbf{0} & & \\ & & \mathbf{1}_{N_{i3}} \end{pmatrix}, \quad e_i = \begin{pmatrix} e_{i1} \\ \vdots \\ e_{iN_i} \end{pmatrix}.$$

Combining all areas, we obtain

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_4 \end{pmatrix},$$

with corresponding model equation

$$y = X\beta + Zu + e,$$

where

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_4 \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad Z = [Z_1, Z_2], \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_4 \end{pmatrix}$$

for

$$Z_1 = \text{diag}(z_1, \cdots, z_4) = \begin{pmatrix} \mathbf{1}_{N_1} & 0 & 0 & 0 \\ 0 & \mathbf{1}_{N_2} & 0 & 0 \\ 0 & 0 & \mathbf{1}_{N_3} & 0 \\ 0 & 0 & 0 & \mathbf{1}_{N_4} \end{pmatrix}, \quad Z_2 = \begin{pmatrix} Z_{21} \\ \vdots \\ Z_{24} \end{pmatrix}.$$

# 4

# Small Area Estimation and a multivariate linear regression model

$\mathbf{A}$s stated previously, the main part of this work is devoted to the SAE for repeated surveys with consideration of different groups within the 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 The EURAREA 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 estimators, we propose a model which borrows strength across both small areas and over time by incorporating simultaneously the effects of areas and time correlation. This 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 model investigated in this chapter is connected with the realistic motivating example about regional estimation of malnutrition presented in the Introduction.

## 4.1  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 $N$ and 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 $\sum_{i}^{m} \sum_{g=1}^{k} N_{ig} = N$. Suppose that a sample $s = s_1, \ldots, s_m$ is selected from the population using SRSWOR where $s_i$ is the sample observed for area $i$ of size $n_i$. The sample remains the same over time.

Let $\boldsymbol{y}_{ij}$ be the $p$-vector of measurements on the $j$-th unit, in the $i$-th area. That is

$$\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijp})', \quad j = 1, \ldots, N_i, \ i = 1, \ldots, m. \tag{4.1}$$

We assume the mean growth of the $j$th unit in area $i$ in each group to be the polynomial in time of degree $q - 1$. Furthermore, we suppose that we have auxiliary data $\boldsymbol{x}_{ij}$ 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 $\boldsymbol{x}_{ij}$ 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, it can be also register based information or the data measured for characteristics of interest in other similar areas.

The relationship between $\boldsymbol{y}_{ij}$ and $\boldsymbol{x}_{ij}$ 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 $\boldsymbol{y}_{ij}$. 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

$$y_{ijt} = \beta_0 + \beta_1 t + \cdots + \beta_q t^{q-1} + \boldsymbol{\gamma}' \boldsymbol{x}_{ij} + u_{it} + e_{ijt}, \tag{4.2}$$
$$j = 1, \ldots, N_i; i = 1, \ldots, m; t = 1, \ldots, p,$$

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 $u_{it}$'s. $\boldsymbol{\gamma}$ is a vector of fixed regression coefficients of auxiliary variables. $\beta_0, \ldots, \beta_q$ 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

$$\boldsymbol{y}_{ij} = \boldsymbol{A}\boldsymbol{\beta} + \mathbf{1}_p \boldsymbol{\gamma}' \boldsymbol{x}_{ij} + \boldsymbol{u}_i + \boldsymbol{e}_{ij}, \tag{4.3}$$

where

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

The vector $\boldsymbol{u}_i$ is assumed to be distributed multivariate normally with zero mean and unknown covariance matrix $\boldsymbol{\Sigma}_u$. Collecting the vectors $\boldsymbol{y}_{ij}$ for all units in small area $i$ coming from the group $g$ gives

$$\boldsymbol{Y}_{ig} = \boldsymbol{A}\boldsymbol{B}_g + \mathbf{1}_p \boldsymbol{\gamma}' \boldsymbol{X}_{ig} + \boldsymbol{u}_i \boldsymbol{z}_{ig} + \boldsymbol{E}_{ig}, \ g = 1, \ldots, k, \tag{4.4}$$

where $\boldsymbol{Y}_{ig} = (\boldsymbol{y}_{i1}, \cdots, \boldsymbol{y}_{iN_{ig}})$; $\boldsymbol{B}_g = (\boldsymbol{\beta}_g, \cdots, \boldsymbol{\beta}_g)_{N_{ig}}$; $\boldsymbol{X}_{ig} = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iN_{ig}})$; $\boldsymbol{z}_{ig} = \frac{1}{\sqrt{N_{ig}}} \mathbf{1}'_{N_{ig}}$ and $\boldsymbol{E}_{ig} = (\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iN_{ig}})$.

The model presented in (4.4) holds for each group unit in area $i$. Since we are interested in every group unit, we need to include a known design matrix $\boldsymbol{C}_i : k \times N_i$ 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

$$\boldsymbol{Y}_i = \boldsymbol{ABC}_i + \boldsymbol{1}_p \boldsymbol{\gamma}' \boldsymbol{X}_i + \boldsymbol{u}_i \boldsymbol{z}_i + \boldsymbol{E}_i, \tag{4.5}$$

where $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}, \cdots, \boldsymbol{Y}_{ik})$; $\boldsymbol{B} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_k) : q \times k$; $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{ik})$; $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}, \cdots, \boldsymbol{z}_{ik})$, $\boldsymbol{E}_{ig} = (\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{ik})$,

$$\boldsymbol{C}_i = \begin{pmatrix} \boldsymbol{1}'_{N_{i1}} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{1}'_{N_{ik}} \end{pmatrix},$$

where $\boldsymbol{E}_i \sim \mathcal{N}_{p, N_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_e, \boldsymbol{I}_{N_i})$ with $\boldsymbol{\Sigma}_e = \sigma_e^2 \boldsymbol{I}_p$. $\boldsymbol{Y}_i$ is a $p \times N_i$ data matrix; $\boldsymbol{A}$ is a $p \times q, q \leq p$ known matrix called *within individuals design matrix for fixed effects*; $\boldsymbol{B}$ is $q \times k$ unknown parameter matrix; $\boldsymbol{C}_i, \mathrm{rank}(\boldsymbol{C}_i) + p \leq N_i$ is a $k \times N_i$ known matrix called *between individuals design matrix for fixed effects* and $\boldsymbol{X}_i$ is a $r \times N_i$ known matrix taking the values of covariables.

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

$$\underbrace{\boldsymbol{Y}}_{p \times N} = \underbrace{\boldsymbol{A}}_{p \times q} \underbrace{\boldsymbol{B}}_{q \times k} \underbrace{\boldsymbol{H}}_{k \times mk} \underbrace{\boldsymbol{C}}_{mk \times N} + \underbrace{\boldsymbol{1}_p \boldsymbol{\gamma}'}_{p \times r} \underbrace{\boldsymbol{X}}_{r \times N} + \underbrace{\boldsymbol{U}}_{p \times m} \underbrace{\boldsymbol{Z}}_{m \times N} + \underbrace{\boldsymbol{E}}_{p \times N}, \tag{4.6}$$

where

$$\boldsymbol{Y} = [\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_m], \ \boldsymbol{H} = \begin{pmatrix} \boldsymbol{I}_k : \cdots : \boldsymbol{I}_k \end{pmatrix}, \ \boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{C}_m \end{pmatrix},$$

$$\boldsymbol{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_m], \ \boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m], \ \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{z}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{z}_m \end{pmatrix},$$

and

$$\boldsymbol{E} = [\boldsymbol{E}_1, \cdots, \boldsymbol{E}_m].$$

It is assumed that

$$\boldsymbol{E} \sim \mathcal{N}_{p,N}(\boldsymbol{0}, \boldsymbol{\Sigma}_e, \boldsymbol{I}_N), \quad \boldsymbol{U} \sim \mathcal{N}_{p,m}(\boldsymbol{0}, \boldsymbol{\Sigma}_u, \boldsymbol{I}_m), \ p \leq m.$$

Furthermore,

$$\mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_{pN} \Big( \mathrm{vec}(\boldsymbol{ABHC} + \boldsymbol{1}_p \boldsymbol{\gamma}' \boldsymbol{X}), \boldsymbol{\Sigma} \Big),$$

for

$$\boldsymbol{\Sigma} = \boldsymbol{Z}'\boldsymbol{Z} \otimes \boldsymbol{\Sigma}_u + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}_e,$$

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

where the covariance between random effects $\boldsymbol{u}_i$'s at two distinct times $t, s$ is $\mathrm{Cov}(\boldsymbol{u}_t, \boldsymbol{u}_s) = \sigma_{uts} = \rho_{t,s}\sigma_{ut}\sigma_{us}$ with $\rho_{t,s}$ standing for the correlation.

## 4.2   Estimation of the mean and covariance parameters

The model defined in (4.6) 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 matrix normal distribution.

First of all, we observe that the matrices $\boldsymbol{Z}, \boldsymbol{H}, \boldsymbol{C}$ are of full row rank and the matrix $\boldsymbol{A}$ is of full column rank. Moreover,

$$\mathcal{C}(\boldsymbol{Z}') \subseteq \mathcal{C}(\boldsymbol{C}') \quad \text{and} \quad \boldsymbol{Z}\boldsymbol{Z}' = \boldsymbol{I}_m.$$

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

$$(\boldsymbol{C}\boldsymbol{C}')^{-1/2}\boldsymbol{C}\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1/2} = \boldsymbol{\Gamma}\boldsymbol{D}\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}\begin{pmatrix} \boldsymbol{I}_m & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{\Gamma}'.$$

Since each column of $\boldsymbol{\Gamma}$ is an eigenvector associated to the corresponding eigenvalue in $\boldsymbol{D}$, we can partition $\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_1 : \boldsymbol{\Gamma}_2 \end{bmatrix}$ such that $\boldsymbol{\Gamma}_1$ corresponds to the block $\boldsymbol{I}_m$ and $\boldsymbol{\Gamma}_2$ corresponds to the block $\boldsymbol{0}$, with $\boldsymbol{\Gamma}_1 : mk \times m$ and $\boldsymbol{\Gamma}_2 : mk \times (mk - m)$. It follows that $\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_1 = \boldsymbol{I}_m$ and $\boldsymbol{\Gamma}_2'\boldsymbol{\Gamma}_2 = \boldsymbol{I}_{mk-m}$.

Choose $\boldsymbol{C}'^o$ to be a matrix which columns are unit length eigenvectors corresponding to non zero eigenvalues of $\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^-\boldsymbol{C}$. Then $(\boldsymbol{C}'^o)'\boldsymbol{C}'^o = \boldsymbol{I}_{N-mk}$ and we can make a one to one transformation on (4.6) to get $[\boldsymbol{V} : \boldsymbol{W}] = \boldsymbol{Y}[\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1/2} : \boldsymbol{C}'^o]$, where

$$\boldsymbol{V} = \boldsymbol{Y}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1/2} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{H}(\boldsymbol{C}\boldsymbol{C}')^{1/2} + \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1/2}$$
$$+ (\boldsymbol{U}\boldsymbol{Z} + \boldsymbol{E})\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1/2}, \tag{4.7}$$
$$\boldsymbol{W} = \boldsymbol{Y}\boldsymbol{C}'^o = \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{X}\boldsymbol{C}'^o + \boldsymbol{E}\boldsymbol{C}'^o, \tag{4.8}$$

since $\boldsymbol{CC'}^{o} = \boldsymbol{0}$ and $\mathcal{C}(\boldsymbol{Z'}) \subseteq \mathcal{C}(\boldsymbol{C'})$ implies that $\boldsymbol{ZC'}^{o} = \boldsymbol{0}$. Make a further transformation on (4.7) so that $[\boldsymbol{V}_1 : \boldsymbol{V}_2] = \boldsymbol{Y}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}[\boldsymbol{\Gamma}_1 : \boldsymbol{\Gamma}_2]$, where

$$
\begin{aligned}
\boldsymbol{V}_1 = \boldsymbol{Y}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1 &= \boldsymbol{ABH}(\boldsymbol{CC'})^{1/2}\boldsymbol{\Gamma}_1 + \boldsymbol{1}_p\boldsymbol{\gamma'}\boldsymbol{X}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1 \\
&\quad + (\boldsymbol{UZ} + \boldsymbol{E})\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1, \quad\quad (4.9) \\
\boldsymbol{V}_2 = \boldsymbol{Y}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_2 &= \boldsymbol{ABH}(\boldsymbol{CC'})^{1/2}\boldsymbol{\Gamma}_2 + \boldsymbol{1}_p\boldsymbol{\gamma'}\boldsymbol{X}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_2 \\
&\quad + \boldsymbol{E}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_2. \quad\quad (4.10)
\end{aligned}
$$

**Theorem 4.1**
*Let $\boldsymbol{V}_1, \boldsymbol{V}_2$ and $\boldsymbol{W}$ be as defined in (4.9), (4.10) and (4.8) respectively. Then, these component models are independent and*

$$
\begin{aligned}
\boldsymbol{V}_1 &\sim \mathcal{N}_{p,m}\Big(\boldsymbol{M}_1, \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_e, \boldsymbol{I}_m\Big), \\
\boldsymbol{V}_2 &\sim \mathcal{N}_{p,mk-m}\Big(\boldsymbol{M}_2, \boldsymbol{\Sigma}_e, \boldsymbol{I}_{mk-m}\Big), \\
\boldsymbol{W} &\sim \mathcal{N}_{p,N-mk}\Big(\boldsymbol{M}_3, \boldsymbol{\Sigma}_e, \boldsymbol{I}_{N-mk}\Big),
\end{aligned}
$$

*where*

$$
\begin{aligned}
\boldsymbol{M}_1 &= \boldsymbol{ABH}(\boldsymbol{CC'})^{1/2}\boldsymbol{\Gamma}_1 + \boldsymbol{1}_p\boldsymbol{\gamma'}\boldsymbol{X}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1, \\
\boldsymbol{M}_2 &= \boldsymbol{ABH}(\boldsymbol{CC'})^{1/2}\boldsymbol{\Gamma}_2 + \boldsymbol{1}_p\boldsymbol{\gamma'}\boldsymbol{X}\boldsymbol{C'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_2, \\
\boldsymbol{M}_3 &= \boldsymbol{1}_p\boldsymbol{\gamma'}\boldsymbol{X}\boldsymbol{C'}^{o}.
\end{aligned}
$$

**Proof:** The independence of $\boldsymbol{V}_1, \boldsymbol{V}_2$ and $\boldsymbol{W}$ follows by the fact that the matrices $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2$ and $\boldsymbol{C'}^{o}$ are all of full rank and pairwise orthogonal. We observe that

$$
\boldsymbol{\Gamma'}\boldsymbol{\Gamma} = \boldsymbol{I}_{mk},
$$

and

$$
\boldsymbol{\Gamma'}\boldsymbol{\Gamma}_1 = \begin{bmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{\Gamma'}\boldsymbol{\Gamma}_2 = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_{mk-m} \end{bmatrix}, \quad \boldsymbol{\Gamma}_1'\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{\Gamma}_2'\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{mk-m} \end{bmatrix}.
$$

From

$$
\boldsymbol{E} \sim \mathcal{N}_{p,N}(\boldsymbol{0}, \boldsymbol{\Sigma}_e, \boldsymbol{I}_N), \quad \boldsymbol{U} \sim \mathcal{N}_{p,m}(\boldsymbol{0}, \boldsymbol{\Sigma}_u, \boldsymbol{I}_m),
$$

it follows that

$$
\begin{aligned}
\boldsymbol{UZC'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1 &\sim \mathcal{N}_{p,m}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_u, \boldsymbol{\Gamma}_1'(\boldsymbol{CC'})^{-1/2}\boldsymbol{CZ'ZC'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1\Big), \\
\boldsymbol{EC'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1 &\sim \mathcal{N}_{p,m}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_e, \boldsymbol{\Gamma}_1'(\boldsymbol{CC'})^{-1/2}\boldsymbol{CC'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_1\Big), \\
\boldsymbol{EC'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_2 &\sim \mathcal{N}_{p,mk-m}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_e, \boldsymbol{\Gamma}_2'(\boldsymbol{CC'})^{-1/2}\boldsymbol{CC'}(\boldsymbol{CC'})^{-1/2}\boldsymbol{\Gamma}_2\Big), \\
\boldsymbol{EC'}^{o} &\sim \mathcal{N}_{p,N-mk}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_e, (\boldsymbol{C'}^{o})'\boldsymbol{C'}^{o}\Big).
\end{aligned}
$$

Hence,

$$(\boldsymbol{UZ} + \boldsymbol{E})\boldsymbol{C}'(\boldsymbol{CC}')^{-1/2}\boldsymbol{\Gamma}_1 \sim \mathcal{N}_{p,m}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_e, \boldsymbol{I}_m\Big),$$

$$\boldsymbol{EC}'(\boldsymbol{CC}')^{-1/2}\boldsymbol{\Gamma}_2 \sim \mathcal{N}_{p,mk-m}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_e, \boldsymbol{I}_{mk-m}\Big),$$

$$\boldsymbol{EC}'^{o} \sim \mathcal{N}_{p,N-mk}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_e, \boldsymbol{I}_{N-mk}\Big),$$

since

$$\boldsymbol{\Gamma}_1'(\boldsymbol{CC}')^{-1/2}\boldsymbol{CZ}'\boldsymbol{ZC}'(\boldsymbol{CC}')^{-1/2}\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}\boldsymbol{D}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}_1 = \boldsymbol{I}_m,$$

$$\boldsymbol{\Gamma}_1'(\boldsymbol{CC}')^{-1/2}\boldsymbol{CC}'(\boldsymbol{CC}')^{-1/2}\boldsymbol{\Gamma}_1 = \boldsymbol{I}_m,$$

$$\boldsymbol{\Gamma}_2'(\boldsymbol{CC}')^{-1/2}\boldsymbol{CC}'(\boldsymbol{CC}')^{-1/2}\boldsymbol{\Gamma}_2 = \boldsymbol{I}_{mk-m},$$

$$(\boldsymbol{C}'^{o})'\boldsymbol{C}'^{o} = \boldsymbol{I}_{N-mk}.$$

$\square$

Now consider the two components $\boldsymbol{W}$ and $\boldsymbol{V}_2$ and recall that the covariance matrix $\boldsymbol{\Sigma}_e = \sigma_e^2 \boldsymbol{I}_p$ is known. Put

$$\boldsymbol{K}_1 = \boldsymbol{H}(\boldsymbol{CC}')^{1/2}\boldsymbol{\Gamma}_1, \quad \boldsymbol{K}_2 = \boldsymbol{H}(\boldsymbol{CC}')^{1/2}\boldsymbol{\Gamma}_2,$$

$$\boldsymbol{R}_1 = \boldsymbol{C}'(\boldsymbol{CC}')^{-1/2}\boldsymbol{\Gamma}_1, \quad \boldsymbol{R}_2 = \boldsymbol{C}'(\boldsymbol{CC}')^{-1/2}\boldsymbol{\Gamma}_2.$$

Then,

$$\boldsymbol{W} = \boldsymbol{YC}'^{o} = \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XC}'^{o} + \boldsymbol{EC}'^{o},$$

$$\boldsymbol{V}_2 = \boldsymbol{YR}_2 = \boldsymbol{ABK}_2 + \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XR}_2 + \boldsymbol{ER}_2.$$

The corresponding loglikelihood of the joint density function is given by

$$l(\boldsymbol{\gamma}, \boldsymbol{B}) = l_{\boldsymbol{W}}(\boldsymbol{\gamma}) + l_{\boldsymbol{V}_2}(\boldsymbol{\gamma}, \boldsymbol{B}),$$

where $l_{\boldsymbol{W}}(\boldsymbol{\gamma})$ and $l_{\boldsymbol{V}_2}(\boldsymbol{\gamma}, \boldsymbol{B})$ denote the loglikelihood functions for $\boldsymbol{W}$ and $\boldsymbol{V}_2$ respectively. Therefore,

$$
\begin{aligned}
l(\boldsymbol{\gamma}, \boldsymbol{B}) = & -\frac{p(N-mk)}{2}\log(2\pi) - \frac{(N-mk)}{2}\log|\boldsymbol{\Sigma}_e| - \frac{p(mk-m)}{2}\log(2\pi) \\
& -\frac{mk-m}{2}\log|\boldsymbol{\Sigma}_e| - \frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{\Sigma}_e^{-1}\Big(\boldsymbol{W} - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XC}'^{o}\Big)\Big(\boldsymbol{W} - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XC}'^{o}\Big)'\Big\} \\
& -\frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{\Sigma}_e^{-1}\Big(\boldsymbol{V}_2 - \boldsymbol{ABK}_2 - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XR}_2\Big)\Big(\boldsymbol{V}_2 - \boldsymbol{ABK}_2 - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XR}_2\Big)'\Big\} \\
= & c_o - \frac{(N-m)}{2}\log|\boldsymbol{\Sigma}_e| - \frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{\Sigma}_e^{-1}\Big(\boldsymbol{W} - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XC}'^{o}\Big)\Big(\boldsymbol{W} - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XC}'^{o}\Big)' \\
& + \boldsymbol{\Sigma}_e^{-1}\Big(\boldsymbol{V}_2 - \boldsymbol{ABK}_2 - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XR}_2\Big)\Big(\boldsymbol{V}_2 - \boldsymbol{ABK}_2 - \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{XR}_2\Big)'\Big\},
\end{aligned}
$$

where $c_o$ is a constant not depending on the parameters.

The first and second derivatives with respect to $\gamma$ and $B$ are

$$\frac{dl(\gamma, B)}{d\gamma} = \sigma_e^{-2}\left[\left(XC'^o(C'^o)' + XR_2R_2'\right)Y'1_p - p\left(XC'^o(C'^o)'X' + XR_2R_2'X'\right)\gamma\right.$$
$$\left. - XR_2K_2'B'A'1_p\right],$$

$$\frac{d^2l(\gamma, B)}{d\gamma^2} = -p\left(XC'^o(C'^o)'X' + XR_2R_2'X'\right),$$

$$\frac{dl(\gamma, B)}{dB} = \sigma_e^{-2}A'\left(YR_2 - ABK_2 - 1_p\gamma'XR_2\right)K_2',$$

$$\frac{d^2l(\gamma, B)}{dB^2} = -\sigma_e^{-2}(K_2K_2') \otimes A'A.$$

Therefore, the likelihood equations can be expressed as

$$\left(XC'^o(C'^o)' + XR_2R_2'\right)Y'1_p - p\left(XC'^o(C'^o)'X' + XR_2R_2'X'\right)\gamma$$
$$- XR_2K_2'B'A'1_p = 0, \tag{4.11}$$
$$A'\left(YR_2 - ABK_2 - 1_p\gamma'XR_2\right)K_2' = 0. \tag{4.12}$$

**Lemma 4.1**
*The likelihood equation given by (4.12) admits a non unique solution for the parameter matrix $B$.*

**Proof:** The likelihood equation (4.12) is equivalent to

$$A'ABK_2K_2' = A'YR_2K_2' - A'1_p\gamma'XR_2K_2'.$$

This equation 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

$$\mathcal{C}(\Gamma_2) = \mathcal{C}\left((CC')^{1/2}(CZ')^o\right).$$

But for any matrices of proper sizes, if $\mathcal{C}(F) = \mathcal{C}(G)$, then $\mathcal{C}(EF) = \mathcal{C}(EG)$ (see for example Harville (1998) for more details). Therefore,

$$\mathcal{C}\left(H(CC')^{1/2}\Gamma_2\right) = \mathcal{C}\left(HCC'(CZ')^o\right).$$

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)

$$\text{rank}(F\colon G) = \text{rank}(F'G^o) + \text{rank}(G),$$

we get

$$\text{rank}(H(CC')^{1/2}\Gamma_2) = \text{rank}(HCC'(CZ')^o) = \text{rank}(CC'H'\colon CZ') - \text{rank}(CZ').$$

Since $\mathcal{C}(\boldsymbol{Z}') \subseteq \mathcal{C}(\boldsymbol{C}')$, then $\boldsymbol{Z}' = \boldsymbol{C}'\boldsymbol{Q}$ for

$$
\boldsymbol{Q} = \begin{pmatrix} \frac{1}{\sqrt{N_1}}\mathbf{1}_k & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{\sqrt{N_m}}\mathbf{1}_k \end{pmatrix}.
$$

It follows that

$$
\begin{aligned}
\operatorname{rank}(\boldsymbol{H}(\boldsymbol{CC}')^{1/2}\boldsymbol{\Gamma}_2) &= \operatorname{rank}(\boldsymbol{HCC}'(\boldsymbol{CZ}')^o) \\
&= \operatorname{rank}(\boldsymbol{CC}'\boldsymbol{H}' : \boldsymbol{CC}'\boldsymbol{Q}) - \operatorname{rank}(\boldsymbol{CC}'\boldsymbol{Q}) \\
&= \operatorname{rank}(\boldsymbol{H}' : \boldsymbol{Q}) - \operatorname{rank}(\boldsymbol{Q}) \\
&= \operatorname{rank}(\boldsymbol{H}') + \operatorname{rank}(\boldsymbol{Q}) - \dim(\mathcal{C}(\boldsymbol{H}') \cap \mathcal{C}(\boldsymbol{Q})) - \operatorname{rank}(\boldsymbol{Q}) \\
&= \operatorname{rank}(\boldsymbol{H}') - \dim(\mathcal{C}(\boldsymbol{H}') \cap \mathcal{C}(\boldsymbol{Q})),
\end{aligned}
$$

since the matrix $\boldsymbol{CC}'$ is of full rank and where $dim$ denotes the dimension of a subspace. It remains to show that $\mathcal{C}(\boldsymbol{H}')$ and $\mathcal{C}(\boldsymbol{Q})$ are not disjoint.

Let $\boldsymbol{v}_1 = \mathbf{1}_k$ and $\boldsymbol{v}_2 = (\sqrt{N_1}, \cdots, \sqrt{N_m})'$ and recall that $\boldsymbol{H} = \big(\boldsymbol{I}_k : \cdots : \boldsymbol{I}_k\big)$. Then we have

$$
\boldsymbol{H}'\boldsymbol{v}_1 = \mathbf{1}_{mk} \quad \text{and} \quad \boldsymbol{Q}\boldsymbol{v}_2 = \mathbf{1}_{mk}.
$$

Thus, the two spaces are not disjoint since they include a common vector. This completes the proof of the lemma. $\qquad\square$

The second equation (4.12) gives

$$
\begin{aligned}
\boldsymbol{B} =& (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-} - (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\mathbf{1}_p\boldsymbol{\gamma}'\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-} \\
&+ \boldsymbol{T}_1(\boldsymbol{K}_2\boldsymbol{K}_2')^{o'},
\end{aligned}
$$

for an arbitrary matrix $\boldsymbol{T}_1$. Plugging in the value of $\boldsymbol{B}$ into equation (4.11) yields

$$
\begin{aligned}
&\Big(\boldsymbol{X}\boldsymbol{C}'^o(\boldsymbol{C}'^o)' + \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{R}_2'\Big)\boldsymbol{Y}'\mathbf{1}_p - p\Big(\boldsymbol{X}\boldsymbol{C}'^o(\boldsymbol{C}'^o)'\boldsymbol{X}' + \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{R}_2'\boldsymbol{X}'\Big)\boldsymbol{\gamma} \\
&- \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-}\boldsymbol{K}_2\boldsymbol{R}_2'\boldsymbol{Y}'\mathbf{1}_p + p\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-}\boldsymbol{K}_2\boldsymbol{R}_2'\boldsymbol{X}'\boldsymbol{\gamma} = \mathbf{0}.
\end{aligned}
$$

Then, the restricted maximum likelihood (RMLE) for $\boldsymbol{\gamma}$ and $\boldsymbol{B}$ are obtained by

$$
\begin{aligned}
\widehat{\boldsymbol{\gamma}} =& \Big(p\boldsymbol{X}\boldsymbol{C}'^o(\boldsymbol{C}'^o)'\boldsymbol{X}' + p\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{R}_2'\boldsymbol{X}' - p\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-}\boldsymbol{K}_2\boldsymbol{R}_2'\boldsymbol{X}'\Big)^{-} \\
&\times \Big(\boldsymbol{X}\boldsymbol{C}'^o(\boldsymbol{C}'^o)'\boldsymbol{Y}' + \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{R}_2'\boldsymbol{Y}' - \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-}\boldsymbol{K}_2\boldsymbol{R}_2'\boldsymbol{Y}'\Big)\mathbf{1}_p \\
&+ \Big(p\boldsymbol{X}\boldsymbol{C}'^o(\boldsymbol{C}'^o)'\boldsymbol{X}' + p\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{R}_2'\boldsymbol{X}' - p\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-}\boldsymbol{K}_2\boldsymbol{R}_2'\boldsymbol{X}'\Big)^{o}\boldsymbol{t}_2, \\
\widehat{\boldsymbol{B}} =& (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-} - (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\mathbf{1}_p\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-} \\
&+ \boldsymbol{T}_1(\boldsymbol{K}_2\boldsymbol{K}_2')^{o'},
\end{aligned}
$$

for an arbitrary vector $\boldsymbol{t}_2$ and an arbitrary matrix $\boldsymbol{T}_1$.

Put

$$\boldsymbol{P} = \boldsymbol{X}\boldsymbol{C}'^o(\boldsymbol{C}'^o)' + \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{R}_2' - \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-\boldsymbol{K}_2\boldsymbol{R}_2'. \qquad (4.13)$$

Then

$$\widehat{\boldsymbol{\gamma}} = \frac{1}{p}(\boldsymbol{P}\boldsymbol{X}')^-\boldsymbol{P}\boldsymbol{Y}'\boldsymbol{1}_p + (\boldsymbol{P}\boldsymbol{X}')^o\boldsymbol{t}_2,$$

$$\begin{aligned}
\widehat{\boldsymbol{B}} =& (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^- - \frac{1}{p}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{1}_p\boldsymbol{1}_p'\boldsymbol{Y}\boldsymbol{P}'(\boldsymbol{X}\boldsymbol{P}')^-\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^- \\
& - (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{1}_p\boldsymbol{t}_2'(\boldsymbol{P}\boldsymbol{X}')^{o'}\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^- + \boldsymbol{T}_1(\boldsymbol{K}_2\boldsymbol{K}_2')^{o'}.
\end{aligned}$$

According to Lemma 4.1, the parameter matrix $\boldsymbol{B}$ is not unique. However, there is also information about $\boldsymbol{B}$ in $\boldsymbol{V}_1$ as given by (4.9) which we now try to utilize. Recall that

$$\boldsymbol{V}_1 = \boldsymbol{Y}\boldsymbol{R}_1 = \boldsymbol{A}\boldsymbol{B}\boldsymbol{K}_1 + \boldsymbol{1}_p\boldsymbol{\gamma}'\boldsymbol{X}\boldsymbol{R}_1 + \boldsymbol{E}\boldsymbol{R}_1,$$

$$\boldsymbol{E}\boldsymbol{R}_1 \sim \mathcal{N}_{p,m}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_e, \boldsymbol{I}_m\Big).$$

Inserting the values of $\boldsymbol{\gamma}$ and $\boldsymbol{B}$ in $\boldsymbol{V}_1$ yields

$$\begin{aligned}
\boldsymbol{Y}\boldsymbol{R}_1 =& \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-\boldsymbol{K}_1 \\
& - \frac{1}{p}\boldsymbol{1}_p\boldsymbol{1}_p'\boldsymbol{Y}\boldsymbol{P}'(\boldsymbol{X}\boldsymbol{P}')^-\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-\boldsymbol{K}_1 \\
& - \boldsymbol{1}_p\boldsymbol{t}_2'(\boldsymbol{P}\boldsymbol{X}')^{o'}\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-\boldsymbol{R}_1 + \boldsymbol{A}\boldsymbol{T}_1(\boldsymbol{K}_2\boldsymbol{K}_2')^{o'}\boldsymbol{K}_1 \\
& + \frac{1}{p}\boldsymbol{1}_p\boldsymbol{1}_p'\boldsymbol{Y}\boldsymbol{P}'(\boldsymbol{X}\boldsymbol{P}')^-\boldsymbol{X}\boldsymbol{R}_1 + \boldsymbol{1}_p\boldsymbol{t}_2'(\boldsymbol{P}\boldsymbol{X}')^{o'}\boldsymbol{X}\boldsymbol{R}_1 + (\boldsymbol{U}\boldsymbol{Z} + \boldsymbol{E})\boldsymbol{R}_1.
\end{aligned}$$

Set

$$\begin{aligned}
\boldsymbol{V}_3 =& \boldsymbol{Y}\boldsymbol{R}_1 - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-\boldsymbol{K}_1 \\
& + \frac{1}{p}\boldsymbol{1}_p\boldsymbol{1}_p'\boldsymbol{Y}\boldsymbol{P}'(\boldsymbol{X}\boldsymbol{P}')^-\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-\boldsymbol{K}_1 - \frac{1}{p}\boldsymbol{1}_p\boldsymbol{1}_p'\boldsymbol{Y}\boldsymbol{P}'(\boldsymbol{X}\boldsymbol{P}')^-\boldsymbol{X}\boldsymbol{R}_1,
\end{aligned}$$

with model equation

$$\begin{aligned}
\boldsymbol{V}_3 =& \boldsymbol{A}\boldsymbol{T}_1(\boldsymbol{K}_2\boldsymbol{K}_2')^{o'}\boldsymbol{K}_1 + \boldsymbol{1}_p\boldsymbol{t}_2'(\boldsymbol{P}\boldsymbol{X}')^{o'}\boldsymbol{X}\boldsymbol{R}_1 \\
& - \boldsymbol{1}_p\boldsymbol{t}_2'(\boldsymbol{P}\boldsymbol{X}')^{o'}\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-\boldsymbol{R}_1 + (\boldsymbol{U}\boldsymbol{Z} + \boldsymbol{E})\boldsymbol{R}_1,
\end{aligned}$$

or equivalently,

$$\boldsymbol{V}_3 = \boldsymbol{A}\boldsymbol{T}_1\underset{\sim}{\boldsymbol{C}}_1 + \boldsymbol{1}_p\boldsymbol{t}_2'\underset{\sim}{\boldsymbol{C}}_2 + \underset{\sim}{\boldsymbol{E}},$$

where

$$\begin{aligned}
\underset{\sim}{\boldsymbol{C}}_1 &= (\boldsymbol{K}_2\boldsymbol{K}_2')^{o'}\boldsymbol{K}_1, \\
\underset{\sim}{\boldsymbol{C}}_2 &= (\boldsymbol{P}\boldsymbol{X}')^{o'}\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^-)\boldsymbol{R}_1, \\
\underset{\sim}{\boldsymbol{E}} &= (\boldsymbol{U}\boldsymbol{Z} + \boldsymbol{E})\boldsymbol{R}_1,
\end{aligned}$$

and

$$\underset{\sim}{\boldsymbol{E}} \sim \mathcal{N}_{p,m}\Big(\boldsymbol{0}, \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_e, \boldsymbol{I}_m\Big).$$

Hence, we have obtained an EGC model with nested subspace condition $\mathcal{C}(\boldsymbol{1}_p) \subseteq \mathcal{C}(\boldsymbol{A})$. Referring to the Theorem 2.3, the MLEs for the parameter matrices $\boldsymbol{T}_1$, $\boldsymbol{t}_2$ and $\boldsymbol{\Sigma}_u$ are given by

$$\widehat{\boldsymbol{t}}_2 = (\boldsymbol{1}_p' \boldsymbol{S}_1^{-1} \boldsymbol{1}_p)^{-1} (\underset{\sim}{\boldsymbol{C}}_2 \boldsymbol{Q}_{\underset{\sim}{\boldsymbol{C}}_1'} \underset{\sim}{\boldsymbol{C}}_2')^{-} \underset{\sim}{\boldsymbol{C}}_2 \boldsymbol{Q}_{\underset{\sim}{\boldsymbol{C}}_1'} \boldsymbol{V}_3' \boldsymbol{S}_1^{-1} \boldsymbol{1}_p + (\underset{\sim}{\boldsymbol{C}}_2 \boldsymbol{Q}_{\underset{\sim}{\boldsymbol{C}}_1'})^{o\prime} \boldsymbol{t}_{21}' \boldsymbol{1}_p, \quad (4.14)$$

$$\widehat{\boldsymbol{T}}_1 = (\boldsymbol{A}' \boldsymbol{S}_2^{-1} \boldsymbol{A})^{-1} \boldsymbol{A}' \boldsymbol{S}_2^{-1} (\boldsymbol{V}_3 - \boldsymbol{1}_p \widehat{\boldsymbol{t}}_2 \underset{\sim}{\boldsymbol{C}}_2) \boldsymbol{C}_1' (\boldsymbol{C}_1 \boldsymbol{C}_1')^{-} + \boldsymbol{A}' \boldsymbol{T}_{11} \underset{\sim}{\boldsymbol{C}}_1^o, \quad (4.15)$$

$$\widehat{\boldsymbol{\Sigma}}_u = \frac{1}{m} (\boldsymbol{V}_3 - \boldsymbol{A}\widehat{\boldsymbol{T}}_1 \underset{\sim}{\boldsymbol{C}}_1 - \boldsymbol{1}_p \widehat{\boldsymbol{t}}_2 \underset{\sim}{\boldsymbol{C}}_2)(\boldsymbol{V}_3 - \boldsymbol{A}\widehat{\boldsymbol{T}}_1 \underset{\sim}{\boldsymbol{C}}_1 - \boldsymbol{1}_p \widehat{\boldsymbol{t}}_2 \underset{\sim}{\boldsymbol{C}}_2)' - \boldsymbol{\Sigma}_e, \quad (4.16)$$

for an arbitrary vector $\boldsymbol{t}_{21}$ and an arbitrary matrix $\boldsymbol{T}_{11}$, where

$$\begin{aligned}
\boldsymbol{S}_1 &= \boldsymbol{V}_3 \boldsymbol{Q}_{(\underset{\sim}{\boldsymbol{C}}_1' : \underset{\sim}{\boldsymbol{C}}_2')} \boldsymbol{V}_3', \\
\boldsymbol{S}_2 &= \boldsymbol{S}_1 + \boldsymbol{Q}_{\boldsymbol{1}_p, \boldsymbol{S}_1^{-1}} \boldsymbol{V}_3 \boldsymbol{P}_{\boldsymbol{Q}_{\boldsymbol{C}_1'} \boldsymbol{C}_2'} \boldsymbol{V}_3' \boldsymbol{Q}_{\boldsymbol{1}_p, \boldsymbol{S}_1^{-1}}'
\end{aligned}$$

for

$$\begin{aligned}
\boldsymbol{P_A} &= \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}', \\
\boldsymbol{Q_A} &= \boldsymbol{I} - \boldsymbol{P_A}, \\
\boldsymbol{P_{A,B}} &= \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{B}, \\
\boldsymbol{Q_{A,B}} &= \boldsymbol{I} - \boldsymbol{P_{A,B}}.
\end{aligned}$$

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

**Theorem 4.2**
*Consider the model given by (4.6). Then, the RMLEs of $\boldsymbol{\gamma}$, $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_u$ can be expressed as*

$$\begin{aligned}
\widehat{\boldsymbol{\gamma}} &= \frac{1}{p}(\boldsymbol{P}\boldsymbol{X}')^{-}\boldsymbol{P}\boldsymbol{Y}'\boldsymbol{1}_p + (\boldsymbol{P}\boldsymbol{X}')^o \widehat{\boldsymbol{t}}_2, \\
\widehat{\boldsymbol{B}} &= (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-} \\
&\quad - \frac{1}{p}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{1}_p\boldsymbol{1}_p'\boldsymbol{Y}\boldsymbol{P}'(\boldsymbol{X}\boldsymbol{P}')^{-}\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-} \\
&\quad - (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{1}_p\widehat{\boldsymbol{t}}_2'(\boldsymbol{P}\boldsymbol{X}')^{o\prime}\boldsymbol{X}\boldsymbol{R}_2\boldsymbol{K}_2'(\boldsymbol{K}_2\boldsymbol{K}_2')^{-} + \widehat{\boldsymbol{T}}_1(\boldsymbol{K}_2\boldsymbol{K}_2')^{o\prime}, \\
\widehat{\boldsymbol{\Sigma}}_u &= \frac{1}{m}(\boldsymbol{V}_3 - \boldsymbol{A}\widehat{\boldsymbol{T}}_1 \underset{\sim}{\boldsymbol{C}}_1 - \boldsymbol{1}_p\widehat{\boldsymbol{t}}_2 \underset{\sim}{\boldsymbol{C}}_2)(\boldsymbol{V}_3 - \boldsymbol{A}\widehat{\boldsymbol{T}}_1 \underset{\sim}{\boldsymbol{C}}_1 - \boldsymbol{1}_p\widehat{\boldsymbol{t}}_2 \underset{\sim}{\boldsymbol{C}}_2)' - \boldsymbol{\Sigma}_e,
\end{aligned}$$

*where $\widehat{\boldsymbol{T}}_1$ and $\widehat{\boldsymbol{t}}_2$ are given by (4.15) and (4.14), respectively.*

## 4.3   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 presents the theory of 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 Henderson's approach as presented in Chapter 2 which consists of maximizing the joint density between the observable random variable and non observable random variable.

Consider the model in (4.6) 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

$$
\begin{aligned}
f(Y, U) =& h(U)g(Y|U)\\
=& \lambda \exp\left\{ -\frac{1}{2}\mathrm{tr}\{U'\Sigma_u^{-1}U\}\right\} \exp\left\{ -\frac{1}{2}\mathrm{tr}\{\Sigma_e^{-1}(Y - ABHC - 1_p\gamma'X - UZ)\right.\\
&\left. \times (Y - ABHC - 1_p\gamma'X - UZ)'\}\right\},
\end{aligned}
$$

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

$$\Sigma_e\Sigma_u^{-1}U + UZZ' = (Y - ABHC - 1_p\gamma'X)Z', \qquad (4.17)$$

and since $ZZ' = I_m$

$$(\Sigma_e\Sigma_u^{-1} + I_p)U = (Y - ABHC - 1_p\gamma'X)Z'.$$

Thus, we get the following theorem about the prediction of random effects

**Theorem 4.3**
*Consider the model defined by (4.6). Then, the prediction of random effects is given by*

$$\widehat{U} = \left(\Sigma_e\widehat{\Sigma}_u^{-1} + I_p\right)^{-1}(Y - A\widehat{B}HC - 1_p\widehat{\gamma}'X)Z'.$$

*where $\widehat{\gamma}$, $\widehat{B}$ and $\widehat{\Sigma}_u$ are given in Theorm 4.2.*

## 4.4   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 (4.6) comprises two important parts of sampled and non sampled units. That is

$$\underbrace{\boldsymbol{Y}}_{p \times N} = (\boldsymbol{Y}_1^{(s)}, \cdots, \boldsymbol{Y}_m^{(s)}, \boldsymbol{Y}_1^{(r)}, \cdots, \boldsymbol{Y}_m^{(r)}),$$

where

$$\underbrace{\boldsymbol{Y}_i^{(s)}}_{p \times n_i} = (\boldsymbol{y}_{i1}, \cdots, \boldsymbol{y}_{in_i}),$$

represents the sampled $n_i$ observations from the $i$-th small area, and

$$\underbrace{\boldsymbol{Y}_i^{(r)}}_{p \times (N_i - n_i)} = (\boldsymbol{y}_{in_{i+1}}, \cdots, \boldsymbol{y}_{iN_i}),$$

which 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

$$\overline{\boldsymbol{y}}_i = f_i \overline{\boldsymbol{Y}}_i^{(s)} + (1 - f_i) \widehat{\overline{\boldsymbol{Y}}}_i^{(r)},$$

where $\overline{\boldsymbol{Y}}_i^{(s)}$ is the vector of small area means corresponding to sampled units, $\widehat{\overline{\boldsymbol{Y}}}_i^{(r)}$ 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,

$$\begin{aligned}
\overline{\boldsymbol{y}}_i &= \frac{f_i}{n_i} \boldsymbol{Y}_i^{(s)} \boldsymbol{1}_{n_i} + \left(\frac{1 - f_i}{N_i - n_i}\right) \widehat{\boldsymbol{Y}}_i^{(r)} \boldsymbol{1}_{(N_i - n_i)} \\
&= \frac{1}{N_i} \left( \boldsymbol{Y}_i^{(s)} \boldsymbol{1}_{n_i} + \widehat{\boldsymbol{Y}}_i^{(r)} \boldsymbol{1}_{(N_i - n_i)} \right),
\end{aligned}$$

where

$$\widehat{\boldsymbol{Y}}_i^{(r)} = \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C}_i + \boldsymbol{1}_p \widehat{\boldsymbol{\gamma}}' \boldsymbol{X}_i + \widehat{\boldsymbol{u}}_i \boldsymbol{z}_i, \tag{4.18}$$

for non sampled units. It is convenient to not that $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{u}}_i$ used in (4.18) are estimators computed from Theorem 4.2 and Theorem 4.3, respectively using observed data where the predicted vectors $\widehat{\boldsymbol{u}}_i$'s are the the columns of the predicted matrix $\widehat{\boldsymbol{U}}$. Then, the target vector of small area means for each group $g$ across all time points is given by

$$\overline{\boldsymbol{y}}_{ig} = \frac{1}{N_{ig}} \left( \boldsymbol{Y}_{ig}^{(s)} \boldsymbol{1}_{n_{ig}} + \widehat{\boldsymbol{Y}}_{ig}^{(r)} \boldsymbol{1}_{(N_{ig} - n_{ig})} \right), \ g = 1, \ldots, k.$$

Equivalently,

$$\overline{\boldsymbol{y}}_{ig} = \frac{1}{N_{ig}} \sum_{j \in s_{ig}} \boldsymbol{y}_{ij} + (1 - \frac{n_{ig}}{N_{ig}})\boldsymbol{A}\boldsymbol{\beta}_g + \frac{1}{N_{ig}} \sum_{j \notin s_i} \mathbf{1}_p \boldsymbol{\gamma}' \boldsymbol{x}_{ij} + (1 - \frac{n_{ig}}{N_{ig}})\boldsymbol{u}_i + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} \boldsymbol{e}_{ij}.$$

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) as discussed in section (3.2.1) of Chapter 3, the BLUP of

$$\boldsymbol{y}_{ig} = (1 - \frac{n_{ig}}{N_{ig}})\boldsymbol{A}\boldsymbol{\beta}_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} \mathbf{1}_p \boldsymbol{\gamma}' \boldsymbol{x}_{ij} + (1 - \frac{n_{ig}}{N_{ig}})\boldsymbol{u}_i$$

is given by

$$\widehat{\boldsymbol{y}}_{ig} = (1 - \frac{n_{ig}}{N_{ig}})\boldsymbol{A}\widehat{\boldsymbol{\beta}}_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} \mathbf{1}_p \widehat{\boldsymbol{\gamma}}' \boldsymbol{x}_{ij} + (1 - \frac{n_{ig}}{N_{ig}})\widehat{\boldsymbol{u}}_i. \qquad (4.19)$$

where $\widehat{\boldsymbol{\beta}}_g$ are columns of the estimated parameter matrix $\widehat{\boldsymbol{B}}$. Therefore,

$$\overline{\boldsymbol{y}}_{ig} = \frac{1}{N_{ig}} \sum_{j \in s_{ig}} \boldsymbol{y}_{ij} + (1 - \frac{n_{ig}}{N_{ig}})\boldsymbol{A}\widehat{\boldsymbol{\beta}}_g + \frac{1}{N_{ig}} \sum_{j \notin s_{ig}} \mathbf{1}_p \widehat{\boldsymbol{\gamma}}' \boldsymbol{x}_{ij} + (1 - \frac{n_{ig}}{N_{ig}})\widehat{\boldsymbol{u}}_i. \quad (4.20)$$

Note that the population means of auxiliary variables $\boldsymbol{x}$ in area $i$ at time $t$ must be known so that the non-sampled mean $\sum_{j \notin s_i} \boldsymbol{x}'_{ijt}$ is then obtained by substracting the corresponding sample means from the population mean.

## 4.5 Simulation study example

In this section, we present a simulation study where we consider four times repeated surveys on a population of size $N = 15000$ having 8 small areas and draw a sample of size $n = 450$ with the following small area sample sizes given in Table 4.1,

***Table 4.1:*** *Sample sizes*

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

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

for

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

and the parameter matrices are

$$B = \begin{pmatrix} 8 & 9 & 10 \\ 11 & 12 & 13 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

The sampling variance is equal to $\sigma_e^2 = 0.16$ and the covariance for the random effects is

$$\Sigma_u = \begin{pmatrix} 4.1 & 1.8 & 1.2 & 2.4 \\ 1.8 & 3.6 & 2.4 & 1.4 \\ 1.2 & 2.4 & 6.0 & 2.2 \\ 2.4 & 1.4 & 2.2 & 9.6 \end{pmatrix}$$

Then, the data are randomly generated from

$$\mathrm{vec}(Y) \sim \mathcal{N}_{pn}(\mathrm{vec}(ABHC + \mathbf{1}_p \gamma' X), \Sigma, I_n),$$

using MATLAB Version 8.3.0.532 (The MathWorks, Inc. USA), where the matrix of covariates $X$ is randomly generated and then taken as fixed.

The simulations are repeated 500 times using the formulas presented in Theorem 4.2 and the following average estimates are obtained :

$$\widehat{B} = \begin{pmatrix} 8.0226 & 9.0551 & 9.9728 \\ 11.0002 & 11.9997 & 13.0002 \end{pmatrix},$$

$$\widehat{\gamma} = \begin{pmatrix} 0.9681 \\ 1.9743 \\ 3.0222 \end{pmatrix},$$

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

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

# 5

# Concluding Remarks

T HE overall conclusion and some suggestions for further research are presented below.

## 5.1  Conclusion

The main task in the present thesis 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 SRSWOR 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 with a 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.

## 5.2  Further research

In future work we wish to study properties of all proposed estimators. In particular, the uniqueness of the estimators, the moments and approximation of the distribution of estimators for this SAE multivariate linear regression model.

The formulation of the multivariate linear regression model investigated in this work was based on the assumption that the sampling units remain the same at different time

occasions. However, in repeated surveys, it often happens to consider different sampling units at each time occasion where some units come in and others are dropped out. This study can be extended to such kind of panel data set.

Moreover, we have assumed the sampling scheme to be simple random sampling without replacement with equal selection probabilities. The model can be adapted to other more complex sampling schemes.

The assumption of independence of random effects is also questionable and therefore, some studies in SAE find reasonable to assume the correlation of random effects of the neighborhood areas in term of distance between them. This work can be extended to the modeling of the spatial correlation among small area random effects as an additional input consideration.

# Bibliography

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401):28–36.

Bolfarine, H. and Zacks, S. (1992). *Prediction Theory for Finite Populations.* Springer-Verlag, New York.

Chambers, R. and Clark, R. G. (2012). *An introduction to Model-Based Survey Sampling with Applications.* Oxford University Press, Oxford New York.

Cochran, W. G. (1977). *Sampling Techniques.* John Wiley and Sons, New York.

Datta, G. S., Lahiri, P., Maiti, T., and Lu, K. L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the us. *Journal of the American Statistical Association*, 94(448):1074–1082.

Fabrizi, E., Ferrante, M. R., and Pacei, S. (2007). Small area estimation of average household income based on unit level models for panel data. *Survey methodology*, 33(2):187–198.

Fay III, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74(366a):269–277.

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

Filipiak, K. and von Rosen, D. (2012). On MLEs in an extended multivariate linear growth curve model. *Metrika*, 75(8):1069–1092.

Fuller, W. A. (2009). *Sampling Statistics*, volume 560. John Wiley and Sons, New Jersey.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: an appraisal. *Statistical science*, pages 55–76.

Gonzalez, M. E. (1973). Use and evaluation of synthetic estimators. *Proceedings of the Social Statistics Section*, pages 33–36.

González-Manteiga, W., Lombardia, M., Molina, I., Morales, D., and Santamaria, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis*, 52:5242–5252.

Harville, D. A. (1998). Matrix algebra from a statistician's perspective. *Technometrics*, 40(2):164–164.

Henderson, C. R. (1973). Sire evaluation and genetic trends. *Journal of Animal Science*, 1973(Symposium):10–41.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, pages 423–447.

Ip, W. C., Wu, M. X., Wang, S. G., and Wong, H. (2007). Estimation for parameters of interest in random effects growth curve models. *Journal of multivariate analysis*, 98(2):317–327.

Khatri, C. G. (1973). Testing some covariance structures under a growth curve model. *Journal of Multivariate Analysis*, 3(1):102–116.

Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, volume 579. Springer, Dordrecht.

Lange, N. and Laird, N. M. (1989). The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *Journal of the American Statistical Association*, 84(405):241–247.

Lehtonen, R. and Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. *Handbook of statistics*, 29:219–249.

McCulloch, C. E. and Neuhaus, J. M. (2008). *Generalized Linear Mixed Models*, volume 2. John Wiley and Sons, New York.

Muirhead, R. J. (2009). *Aspects of Multivariate Statistical Theory*, volume 197. John Wiley and Sons, New York.

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.

Nzabanita, J., Singull, M., and von Rosen, D. (2012). Estimation of parameters in the extended growth curve model with a linearly structured covariance matrix. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 16(1):13–32.

Ohlson, M. and Srivastava, M. S. (2010). Profile analysis for a growth curve model. *Journal of the Japan Statistical Society*, 40(1):001–021.

Pfeffermann, D. (2002). Small area estimation-new developments and directions. *International Statistical Review*, 70(1):125–143.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28(1):40–68.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-plus*. Springer-Verlag, New York.

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.

Rahman, A. (2008). A review of small area estimation problems and methodological developments. Discussion paper, NATSEM-University of Canberra.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1):1–17.

Rao, J. N. K. (2003). *Small Area Estimation*, volume 331. John Wiley and Sons, New York.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical science*, pages 15–32.

Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model-Assisted Survey Sampling*. Springer-Verlag, New York.

Searle, S. R., Casella, G., and McCulloch, C. E. (2009). *Variance Components*, volume 391. 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. S. (2002). *Methods of Multivariate Statistics*. Wiley-Interscience, New York.

Srivastava, M. S. and Singull, M. (2012). Profile analysis with random-effects covariance structure. *Journal of the Japan Statistical Society*, 42(2):145–164.

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

von Rosen, D. (1989). Maximum likelihood estimators in multivariate linear normal models. *Journal of multivariate analysis*, 31(2):187–200.

Yokoyama, T. (2001). Estimation in a random effects model with parallel polynomial growth curves. *Hiroshima Mathematical Journal*, 31(3):425–433.

Yokoyama, T. and Fujikoshi, Y. (1993). A parallel profile model with random-effects covariance structure. *J. Japan Statist. Soc.*, 23:83–89.