SAMPLE DELIVERABLES 12 and 16

FINAL SMALL AREA ESTIMATION DEVELOPMENTS AND SIMULATION RESULTS

Grant agreement No:	SSH - CT - 2007 - 217565
Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Funding Scheme:	Collaborative Project - Small or medium scale focused research project
Deliverable n.:	D12 and D16
Deliverable name:	Final small area estimation developments and simulation results
WP n.:	2
Lead Beneficiary:	1, 5
Nature:	Report
Dissemination level:	PU
Status:	Final
Due delivery date:	May 31^{st}, 2010
Actual delivery date:	June 7^{th}, 2010
Project co-ordinator name:	Monica Pratesi
Title:	Associate Professor of Statistics - University of Pisa
Organization:	Department of Statistics and Mathematics Applied to Economics of the University of Pisa (UNIPI-DSMAE)
Tel.:	+39-050-2216252, +39-050-2216492
Fax:	+39-050-2216375
E-mail:	coordinator@sample-project.eu
Project website address:	www.sample-project.eu

FINAL SMALL AREA ESTIMATION DEVELOPMENTS

AND SIMULATION RESULTS

Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Grant agreement No:	SSH - CT - 2007 - 217565
WP n.:	2
Deliverable n.:	D12 and D16
Document title:	Final small area estimation developments and simulation results
Status:	Final
Editors:	I. Molina, D. Morales, M. Pratesi and N. Tzavidis
Authors:	Monica Pratesi (m.pratesi@ec.unipi.it), UNIPI-DSMAE Caterina Giusti (caterina.giusti@ec.unipi.it), UNIPI-DSMAE Stefano Marchetti (s.marchetti@ds.unifi.it), UNIPI-DSMAE Nicola Salvati (salvati@ec.unipi.it), UNIPI-DSMAE
	Nikos Tzavidis (nikos.tzavidis@manchester.ac.uk), CCSR
	Isabel Molina (imolina@est-econ.uc3m.es), UC3M María Durbán (mdurban@est-econ.uc3m.es), UC3M Aurea Grané (agrane@est-econ.uc3m.es), UC3M Juan Miguel Marín (jmmarin@est-econ.uc3m.es), UC3M Maria Helena Veiga (mhveiga@est-econ.uc3m.es), UC3M
	Domingo Morales (d.morales@umh.es), UMH María Dolores Esteban (md.esteban@umh.es), UMH Ángel Sánchez (a.sanchez@umh.es), UMH Laureano Santamaría (l.santamaria@umh.es), UMH Yolanda Marhuenda (y.marhuenda@umh.es), UMH Agustín Pérez (agustin.perez@umh.es), UMH Maria Chiara Pagliarella (mc.pagliarella@unicas.it), UMH
	J.N.K. Rao (jrao@math.carleton.ca), Carleton University Caterina Ferretti (ferretti@ds.unifi.it), University of Florence

Contents

Prologue

This report contains the final small area developments of the partners of the WP2 in the SAMPLE project. The target of the report is to present the statistical methodology that has been developed within the SAMPLE project. The manuscript is organized in ten chapters.

Chapter 1 introduces the basic theory of linear mixed models (LMMs). Special attention is given to the model fitting methods and algorithms, to the calculation of EBLUP estimates and to the estimation of their mean squared errors.

Chapter 2 describes a methodology for obtaining empirical best predictors of general, possibly non-linear, domain parameters using unit level linear regression models. The proposed method (called EB method) is particularized to FGT poverty measures as particular cases of non-linear parameters. The mean squared error of the proposed estimators is obtained by a parametric bootstrap for finite populations.

Chapter 3 proposes a modification of the EB method, called fast EB method, which reduces drastically the computing time, making feasible the estimation of complex non-linear quantities under large populations, whereas loosing little efficiency.

Chapter 4 introduces an area-level linear mixed model with spatial correlation. For this model, the EBLUP, called here Spatial EBLUP, is introduced and ML and REML model fitting methods are described. Analytical approximations of the mean squared error (MSE) of the Spatial EBLUP are discussed, and parametric and nonparametric bootstrap procedures for estimating the MSE are proposed.

Chapter 5 treats the problem of specifying the weight matrix in area-level linear mixed model with spatial correlation, which is one of the challenges in analyzing spatial data. The literature on spatial econometrics and statistics specifies mainly two ways of modeling this matrix.

Chapter 6 gives a semiparametric version of the basic Fay-Herriot model that is based on P-splines and can also handle situations where the functional form of the relationship between the variable of interest and the covariates cannot be specified a priori. This is often the case when the data are supposed to be affected by spatial proximity effects. In these cases P-spline bivariate smoothing can easily introduce spatial effects in the area level model.

Chapter 7 deals with area-level time models. Two basic models are presented. The first one contains time random effects following an auto-regressive process AR(1) and the second one is a simplification where these effects are independent. Complete theoretical developments are presented as well as some simulations to study the behavior of the fitting algorithms and to investigate when it is worthwhile to employ AR(1) random effects. Extension of the basic models for partitioned populations are also given.

Chapter 8 introduces two area-level linear mixed models with time and spatial correlations. For these models, the EBLUP is introduced and the REML model fitting methods are described. Parametric bootstrap procedures for estimating the MSE are proposed.

Chapter 9 describes two unit-level time models. As in the case of area-level models, two models are presented. The first one contains time random effects following an auto-regressive process AR(1) and the second one is a simplification where these effects are independent.

Chapter 10 presents M-quantile regression, nonparametric M-quantile regression and M-quantile Geographically Weighted regression and describes how quantile or M-quantile models can be employed for measuring area effects and estimators of cumulative distribution function. This chapter also discusses mean squared error estimation for M-quantile small area predictors. It also reports several simulation studies and empirical evaluations of the introduced estimation methods.

This report has been coordinated by Domingo Morales (UMH). He has also been in charge of writing Chapters 1, 7-9. Isabel Molina (UC3M) has been responsible for the elaboration of Chapters 2-5. Finally, Nikos Tzavidis (CCSR) and Monica Pratesi (UNIPI-DSMAE) have coordinated the production of the contents of Chapters 6 and 10.

Chapter 1

Linear mixed models

1.1 Linear mixed models with known variance

1.1.1 Introduction

We consider the model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}, \qquad (1.1)$$

where $\mathbf{y}_{n\times 1}$ is the vector of observations, $\beta_{p\times 1}$ is the vector of fixed effects, $\mathbf{u}_{q\times 1}$ is the vector of random effects, $\mathbf{X}_{n\times p}$ and $\mathbf{Z}_{n\times q}$ are the incidence matrices and $\mathbf{e}_{n\times 1}$ is the vector of sampling errors. We assume that sampling errors and random effects are independent and normally distributed with mean equal to zero and known matrices of variances,

$$\text{var}[\mathbf{u}] = E[\mathbf{u}\mathbf{u}'] = \mathbf{V}_u \qquad \text{and} \qquad \text{var}[\mathbf{e}] = E[\mathbf{e}\mathbf{e}'] = \mathbf{V}_e,$$

depending on a parameter θ containing the variance components. From (1.1) we obtain

$$\mathbf{V} = \text{var}[\mathbf{y}] = \mathbf{Z}\mathbf{V}_u\mathbf{Z}' + \mathbf{V}_e,$$

where $\mathbf{V}$ is assumed to be not singular.

1.1.2 Least squared estimation of β

In this section we assume that the variance components of model (1.1) are known. The random term is $Z\mathbf{u} + \mathbf{e}$, with variance $\text{var}[Z\mathbf{u} + \mathbf{e}] = \mathbf{Z}\mathbf{V}_u\mathbf{Z}' + \mathbf{V}_e = \mathbf{V}$. We transform the model to have uncorrelated random terms and common variance equal to 1, i.e.

$$\mathbf{V}^{-1/2}\mathbf{y} = \mathbf{V}^{-1/2}\mathbf{X}\beta + \mathbf{V}^{-1/2}(\mathbf{Z}\mathbf{u} + \mathbf{e}).$$

Assuming that $\mathbf{y}^* = \mathbf{V}^{-1/2}\mathbf{y}$, $\mathbf{e}^* = \mathbf{V}^{-1/2}(\mathbf{Z}\mathbf{u} + \mathbf{e})$ and $\mathbf{X}^* = \mathbf{V}^{-1/2}\mathbf{X}$; the model is

$$\mathbf{y}^* = \mathbf{X}^*\beta + \mathbf{e}^*$$

with $\mathrm{var}[\mathbf{e}^*] = \mathbf{V}^{-1/2}\mathrm{var}[\mathbf{Zu}+\mathbf{e}]\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \mathbf{I}_n$. Therefore, one can apply the ordinary least squared method, i.e.

$$\widehat{\beta} = argmin_{\beta}(\mathbf{e}^{*\prime}\mathbf{e}^*).$$

We observe that

$$\begin{aligned}
\mathbf{e}^{*\prime}\mathbf{e}^* &= \left(\mathbf{V}^{-1/2}\mathbf{y} - \mathbf{V}^{-1/2}\mathbf{X}\beta\right)'\left(\mathbf{V}^{-1/2}\mathbf{y} - \mathbf{V}^{-1/2}\mathbf{X}\beta\right) \\
&= (\mathbf{y}-\mathbf{X}\beta)'\,\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta) = \mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + \beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta.
\end{aligned}$$

By taking derivatives, we obtain

$$\frac{\partial \mathbf{e}^{*\prime}\mathbf{e}^*}{\partial\beta} = -2\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta.$$

The normal equations are

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{1.2}$$

and the solution is

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{1.3}$$

when $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ and $\mathbf{V}$ are invertible. Under normality $\widehat{\beta}$ is also the *maximum likelihood estimator* (MLE) of β, i.e.

$$\widehat{\beta} = argmax_{\beta}\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta)\right).$$

1.1.3 Best linear unbiased prediction of a linear combination of effects

We look at the model (1.1) and define $\tau = \mathbf{a}_r'(\mathbf{X}_r\beta + \mathbf{Z}_r\mathbf{u})$, where $\mathbf{a}_r$ $(k\times 1)$, $\mathbf{X}_r$ $(k\times p)$ and $\mathbf{Z}_r$ $(k\times q)$ are known vectors and matrices. Let $\widehat{\tau} = \mathbf{g}'\mathbf{y} + g_0$ be a linear estimator (predictor) of τ, where $\mathbf{g}$ $(n\times 1)$ and g_0 (1×1) are such that

1. $\widehat{\tau}$ is unbiased, i.e.

   $$E[\tau] = \mathbf{a}_r'\mathbf{X}_r\beta \quad \text{and} \quad E[\widehat{\tau}] = \mathbf{g}'\mathbf{X}\beta + g_0$$

   are equal. Thus $g_0 = 0$ and $\mathbf{a}_r'\mathbf{X}_r = \mathbf{g}'\mathbf{X}$.

2. $\widehat{\tau}$ minimizes the prediction error

   $$\begin{aligned}
   E[(\widehat{\tau}-\tau)^2] &= V(\widehat{\tau}-\tau) = V(\mathbf{g}'\mathbf{y} - \mathbf{a}_r'\mathbf{X}_r\beta - \mathbf{a}_r'\mathbf{Z}_r\mathbf{u}) = V(\mathbf{g}'\mathbf{y} - \mathbf{a}_r'\mathbf{Z}_r\mathbf{u}) \\
   &= \mathbf{g}'\mathbf{V}\mathbf{g} + \mathbf{a}_r'\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_r'\mathbf{a}_r - 2\mathbf{g}'\mathbf{C}\mathbf{Z}_r'\mathbf{a}_r,
   \end{aligned}$$

   where $\mathbf{C} = \mathrm{cov}(\mathbf{y},\mathbf{u}) = \mathbf{Z}\mathbf{V}_u$.

Therefore, the problem to be solved is

$$\text{minimize } V(\widehat{\tau}-\tau), \quad \text{restricted to } \mathbf{a}_r'\mathbf{X}_r = \mathbf{g}'\mathbf{X}.$$

Since $\mathbf{a}_r'\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_r'\mathbf{a}_r$ does not depend on $\mathbf{g}$, the Lagrangian function is

$$L(\mathbf{g},\lambda) = \mathbf{g}'\mathbf{V}\mathbf{g} - 2\mathbf{g}'\mathbf{C}\mathbf{Z}_r'\mathbf{a}_r + 2(\mathbf{g}'\mathbf{X} - \mathbf{a}_r'\mathbf{X}_r)\lambda.$$

By taking partial derivatives with respect to $\mathbf{g}$ and λ, we obtain

$$\begin{aligned} 0 &= \frac{\partial L(\mathbf{g},\lambda)}{\partial \mathbf{g}} = 2\mathbf{V}\mathbf{g} - 2\mathbf{C}\mathbf{Z}_r'\mathbf{a}_r + 2\mathbf{X}\lambda \iff \mathbf{V}\mathbf{g} + \mathbf{X}\lambda = \mathbf{C}\mathbf{Z}_r'\mathbf{a}_r \\ 0 &= \frac{\partial L(\mathbf{g},\lambda)}{\partial \lambda} = 2\mathbf{g}'\mathbf{X} - 2\mathbf{a}_r'\mathbf{X}_r \iff \mathbf{g}'\mathbf{X} = \mathbf{a}_r'\mathbf{X}_r \end{aligned}$$

In matrix form, the above equations are

$$\begin{pmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{g} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{C}\mathbf{Z}_r'\mathbf{a}_r \\ \mathbf{X}_r'\mathbf{a}_r \end{pmatrix}$$

If we apply the formula

$$\begin{bmatrix} A & B \\ B' & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} \left(C - B'A^{-1}B\right)^{-1} \left[-B'A^{-1}, I\right],$$

with $A = \mathbf{V}$, $B = \mathbf{X}$, $C = \mathbf{0}$, then we obtain

$$\begin{aligned} \begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{V}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} - \begin{bmatrix} -\mathbf{V}^{-1}\mathbf{X} \\ I \end{bmatrix} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \left[-\mathbf{X}'\mathbf{V}^{-1}, I\right] \\ &= \begin{pmatrix} \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} & \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} & -(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \end{pmatrix} \end{aligned}$$

Therefore

$$\begin{pmatrix} \mathbf{g} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{C}\mathbf{Z}_r'\mathbf{a}_r \\ \mathbf{X}_r'\mathbf{a}_r \end{pmatrix},$$

with

$$\mathbf{g} = \mathbf{V}^{-1}\mathbf{C}\mathbf{Z}_r'\mathbf{a}_r - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{C}\mathbf{Z}_r'\mathbf{a}_r + \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}_r'\mathbf{a}_r.$$

The best linear unbiased predictor (BLUP) of τ is

$$\begin{aligned} \widehat{\tau} &= \mathbf{g}'\mathbf{y} = \mathbf{a}_r'\mathbf{X}_r\{(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\} + \mathbf{a}_r'\mathbf{Z}_r\mathbf{C}'\mathbf{V}^{-1}\mathbf{y} \\ &- \mathbf{a}_r'\mathbf{Z}_r\mathbf{C}'\mathbf{V}^{-1}\mathbf{X}\{(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\} \\ &= \mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta} + \mathbf{Z}_r\mathbf{C}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta})\right], \end{aligned}$$

where

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

is the least squared estimator of β.

As $\mathbf{C} = \mathrm{cov}(\mathbf{y},\mathbf{u}) = \mathbf{Z}\mathbf{V}_u$, by taking $\mathbf{X}_r = \mathbf{0}$, $\mathbf{a}_r = \mathbf{1}_{(i)} = (0,\ldots,0,1^{(i)},0,\ldots,0)'$ and $\mathbf{Z}_r = \mathbf{I}$ we obtain

$$\widehat{u}_i = \mathbf{1}_{(i)}'\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}), \quad i = 1,\ldots,q,$$

or equivalently

$$\widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}).$$

1.1.4 Best linear unbiased prediction of u

The *best linear unbiased predictor* (BLUP) of $\mathbf{u}$ is

$$\widehat{\mathbf{u}} = \mathbf{V}_u \mathbf{Z}' \mathbf{V}^{-1} \left(\mathbf{y} - \mathbf{X}\widehat{\beta} \right). \tag{1.4}$$

The predictor (1.4) has the following properties:

- "Best" in the sense that minimizes $E[(\widehat{\mathbf{u}} - \mathbf{u})' \mathbf{A} (\widehat{\mathbf{u}} - \mathbf{u})]$ for any given positive definite matrix $\mathbf{A}$.
- Linear with respect to $\mathbf{y}$.
- Unbiased: $E[\widehat{\mathbf{u}} - \mathbf{u}] = \mathbf{0}$.

For more details see Searle (1971), 458-462, or chapter 7 of Searle et al. (1992).

1.2 Linear mixed models with unknown variances

Let us consider the mixed model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1 \mathbf{u}_1 + \ldots + \mathbf{Z}_m \mathbf{u}_m + \mathbf{e}, \tag{1.5}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$ is the vector of sample observations, $\beta = (\beta_1, \ldots, \beta_p)'$ is the vector of fixed effects, and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iq_i})'$ is the vector containing the effects of the q_i levels of the i-th random factor. The expression i-th random factor is used to denote the vector $\mathbf{u}_i$. Finally, $\mathbf{e} = (e_1, \ldots, e_n)'$ is the vector of sampling errors, and $\mathbf{X}$, $\mathbf{Z}_1, \ldots, \mathbf{Z}_m$ are design matrices with dimensions $n \times p$, $n \times q_1, \ldots, n \times q_m$ respectively.

The model (1.5) can be written in the form (1.1) if we define

$$\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_m] \quad \text{and} \quad \mathbf{u} = [\mathbf{u}_1', \ldots, \mathbf{u}_m']', \quad q = \sum_{i=1}^{m} q_i.$$

The following assumptions ensure that the model parameters are estimable.

(F1) $\mathbf{u}_1, \ldots, \mathbf{u}_m$, $\mathbf{e}$ are independent, and

$$\mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, \sigma_0^2 \Sigma_e), \quad \mathbf{u}_i \sim \mathcal{N}_{q_i}(\mathbf{0}, \sigma_i^2 \Sigma_{u_i}), \; i = 1, \ldots, m,$$

with Σ_e and Σ_{u_i}, $i = 1, \ldots, m$, known.

(F2) $\mathrm{r}(\mathbf{X}) = p$.

Note The assumption (F2) always holds if an adequate re-parametrization of the model is made.

The next hypothesis states that the number of observations should be greater than the number of parameters.

(F3) $n \geq p+m+1$.

If assumption (F4) holds, then the fix effects are not confused with the random effects of any factors.

(F4) $\mathrm{r}(\mathbf{X} : \mathbf{Z}_i) > p,\ i = 1, \ldots, m$.

Assumption (F5) ensures that random effects of a factor are not confused with random effects of other factors. Let $\mathbf{G}_0 = \Sigma_e$ and $\mathbf{G}_i = \mathbf{Z}_i \Sigma_{u_i} \mathbf{Z}_i'$, $i = 1, \ldots, m$.

(F5) $\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_m$ are linearly independent, then,

$$\sum_{i=0}^{m} \alpha_i \mathbf{G}_i = \mathbf{0} \Longrightarrow \alpha_i = 0,\ i = 0, 1, \ldots, m\,.$$

Finally, assumption (F6) states that $\mathbf{Z}_i,\ i = 1, \ldots, m$, are standard design matrices.

(F6) $\mathbf{Z}_i$ has only 0's and 1's. In each row there is exactly one 1, and in each column there is at least one 1, $i = 1, \ldots, m$.

This assumption implies that $\mathbf{Z}_i'\mathbf{Z}_i$ is a $q_i \times q_i$ nonsingular diagonal matrix, $\mathrm{r}(\mathbf{Z}_i) = q_i$ and $q_i \leq n$, $i = 1, \ldots, m$.

Another consequence of the previous assumption is that

$$\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\beta, \mathbf{V}), \quad \text{with} \quad \mathbf{V} = \sum_{i=0}^{m} \sigma_i^2 \mathbf{G}_i.$$

Let $\sigma = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_m^2)'$. When necessary, we will emphasize the dependency of $\mathbf{V}$ on σ by writing $\mathbf{V}(\sigma)$. Let $M = p+m+1$ and let $\theta' = (\beta', \sigma')$ be the vector of unknown parameters. The parameter space is

$$\Theta = \{\theta' = (\beta', \sigma'); \beta \in R^p; \sigma_0^2 > 0; \sigma_i^2 \geq 0,\ i = 1, \ldots, m\}\,. \tag{1.6}$$

The likelihood of θ, given a vector of observations $\mathbf{y}$, is denoted in the same way as the joint density function of $\mathbf{y}$ given θ, i.e.

$$f_\theta(\mathbf{y}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \right\}\,. \tag{1.7}$$

1.3 Maximum likelihood estimation

1.3.1 Description of the method

The maximum likelihood estimator $\widehat{\theta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p, \widehat{\sigma}_0^2, \ldots, \widehat{\sigma}_m^2)'$ of θ is the vector satisfying

$$\widehat{\theta} = \mathrm{argmax}_{\theta \in \Theta} f_\theta(\mathbf{y}) = \mathrm{argmax}_{\theta \in \Theta} \log f_\theta(\mathbf{y})\,.$$

Note that $l(\theta) = \log f_\theta(\mathbf{y})$. We denote the vector of derivatives as $\mathbf{S}(\theta) = (S_\beta, S_{\sigma_0^2}, \ldots, S_{\sigma_m^2})'$, where

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \left(\frac{\partial l(\theta)}{\partial \beta}, \frac{\partial l(\theta)}{\partial \sigma_0^2}, \ldots, \frac{\partial l(\theta)}{\partial \sigma_m^2} \right)'.$$

If $\widehat{\theta}$ exists in the interior Θ, then it is the solution of the likelihood equations which are obtained by equating to zero the components of the vector of scores. By deriving the log-likelihood with respect to the parameters we obtain the score components of model (1.5), i.e.

$$S_\beta = \mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta), \tag{1.8}$$
$$S_{\sigma_i^2} = -\frac{1}{2}\frac{\partial \log|\mathbf{V}|}{\partial \sigma_i^2} - \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)'\frac{\partial \mathbf{V}^{-1}}{\partial \sigma_i^2}(\mathbf{y} - \mathbf{X}\beta),\ i = 0, 1, \ldots, m.$$

We know that

$$\frac{\partial \log|\mathbf{V}|}{\partial \sigma_i^2} = \mathrm{tr}\left\{ \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2} \right\}, \tag{1.9}$$
$$\frac{\partial \mathbf{V}^{-1}}{\partial \sigma_i^2} = -\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}. \tag{1.10}$$

Since $\partial \mathbf{V}/\partial \sigma_i^2 = \mathbf{G}_i$,we have

$$S_{\sigma_i^2} = -\frac{1}{2}\mathrm{tr}\{\mathbf{V}^{-1}\mathbf{G}_i\} + \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta),\ i = 0, 1, \ldots, m. \tag{1.11}$$

When we equate (1.8) and (1.11) to zero, we obtain the likelihood equations

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{1.12}$$
$$\mathrm{tr}\{\mathbf{V}^{-1}\mathbf{G}_i\} = (\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta),\ i = 0, 1, \ldots, m. \tag{1.13}$$

These equations cannot be solved to obtain explicit expressions of the maximum likelihood estimators. The Newton-Raphson or the Fisher-Scoring algorithms calculate them iteratively, starting with an initial value θ^0. In each iteration, the Newton-Raphson method updates the estimator of θ by using the formula

$$\theta^{i+1} = \theta^i - \mathbf{H}(\theta^i)^{-1}\mathbf{S}(\theta^i),$$

where $\mathbf{S}(\theta^i)$ is the vector of derivatives and $\mathbf{H}(\theta^i)$ is the Hessian matrix of $l(\theta)$, both calculated with the estimator obtained at the last iteration θ^i. The elements of the Hessian matrix are obtained by taking new derivatives, using (1.10) and applying the property that the derivative of the trace of a matrix is the the trace of the derivative of the matrix, i.e.

$$\frac{\partial^2 l(\theta)}{\partial \beta \partial \beta'} = -\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}, \tag{1.14}$$
$$\frac{\partial^2 l(\theta)}{\partial \sigma_i^2 \partial \beta} = \frac{\partial^2 l(\theta)}{\partial \beta \partial \sigma_i^2} = -\mathbf{X}'\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta), \tag{1.15}$$
$$\frac{\partial^2 l(\theta)}{\partial \sigma_j^2 \sigma_i^2} = \frac{1}{2}\mathrm{tr}\{\mathbf{V}^{-1}\mathbf{G}_j\mathbf{V}^{-1}\mathbf{G}_i\} - (\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}\mathbf{G}_j\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta), \tag{1.16}$$

for $i,j=0,1,\ldots,m$. We illustrate the calculation of the second sum on (1.16). Let $Q=\frac{1}{2}\mathbf{y}'\mathbf{A}^{-1}\mathbf{y}$, where $\mathbf{A}^{-1}=\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}$. Then $\mathbf{A}=\mathbf{V}\mathbf{G}_i^{-1}\mathbf{V}$ and $\frac{\partial\mathbf{A}}{\partial\sigma_j^2}=\mathbf{V}\mathbf{G}_i^{-1}\mathbf{G}_j+\mathbf{G}_j\mathbf{G}_i^{-1}\mathbf{V}$. Therefore

$$\begin{aligned}\frac{\partial Q}{\partial\sigma_j^2} &= -\frac{1}{2}\mathbf{y}'\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial\sigma_j^2}\mathbf{A}^{-1}\mathbf{y}=-\frac{1}{2}\mathbf{y}'(\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1})[\mathbf{V}\mathbf{G}_i^{-1}\mathbf{G}_j+\mathbf{G}_j\mathbf{G}_i^{-1}\mathbf{V}](\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1})\mathbf{y}\\ &= -\frac{1}{2}\mathbf{y}'\mathbf{V}^{-1}\mathbf{G}_j\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}\mathbf{y}-\frac{1}{2}\mathbf{y}'\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}\mathbf{G}_j\mathbf{V}^{-1}\mathbf{y}=-\mathbf{y}'\mathbf{V}^{-1}\mathbf{G}_j\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}\mathbf{y}\end{aligned}$$

The Fisher-scoring method replaces the Hessian matrix by its expectation with the sign changed, that is, the information of Fisher matrix. The updating formula is

$$\theta^{i+1}=\theta^i+F(\theta^i)^{-1}S(\theta^i),$$

and $\mathbf{F}(\theta^i)$ is the Fisher information matrix defined by

$$\mathbf{F}(\theta)=-E[\mathbf{H}(\theta)],$$

and evaluated in θ^i. Taking expectations in (1.14)-(1.16), changing the sign and using the result

$$E[(\mathbf{y}-\mathbf{X}\beta)'\mathbf{A}(\mathbf{y}-\mathbf{X}\beta)]=\operatorname{tr}\{\mathbf{A}\mathbf{V}\},$$

for any not random matrix $\mathbf{A}$, we get the elements of the Fisher information matrix

$$F_{\beta\beta}=\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}, \tag{1.17}$$

$$F_{\sigma_i^2\beta}=F_{\beta\sigma_i^2}=\mathbf{0},\ i=0,1,\ldots,m, \tag{1.18}$$

$$F_{\sigma_j^2\sigma_i^2}=\frac{1}{2}\operatorname{tr}\{\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}\mathbf{G}_j\},\ i,j=0,1,\ldots,m. \tag{1.19}$$

We get

$$F(\theta)=\begin{pmatrix}\mathbf{F}_{\beta\beta} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & F_{\sigma_0^2\sigma_0^2} & F_{\sigma_0^2\sigma_1^2} & \cdots & F_{\sigma_0^2\sigma_m^2}\\ \mathbf{0} & F_{\sigma_1^2\sigma_0^2} & F_{\sigma_1^2\sigma_1^2} & \cdots & F_{\sigma_1^2\sigma_m^2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & F_{\sigma_m^2\sigma_0^2} & F_{\sigma_m^2\sigma_1^2} & \cdots & F_{\sigma_m^2\sigma_m^2}\end{pmatrix}=\begin{pmatrix}F(\beta) & \mathbf{0}\\ \mathbf{0} & F(\sigma)\end{pmatrix}.$$

The block structure of matrix $F(\theta)$ allows to separate the updating equation separately in two equations

$$\beta^{i+1}=\beta^i+F(\beta^i)^{-1}S(\beta^i),\qquad \sigma^{i+1}=\sigma^i+F(\sigma^i)^{-1}S(\sigma^i).$$

Finally

$$\beta^{i+1}=\beta^i+(\mathbf{X}'\mathbf{V}^{-1}(\sigma^i)\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\sigma^i)(\mathbf{y}-\mathbf{X}\beta^i)=(\mathbf{X}'\mathbf{V}^{-1}(\sigma^i)\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\sigma^i)\mathbf{y}.$$

1.3.2 Maximum likelihood with alternative parametrization

We consider the model (1.5) and the parameters

$$\sigma^2 = \sigma_0^2, \quad \varphi_i = \sigma_i^2/\sigma_0^2, \quad i = 1, \dots, m.$$

Let $\sigma' = (\sigma^2, \varphi_1, \dots, \varphi_m)$, $\theta' = (\beta', \sigma')$ and $\mathbf{V} = \sigma^2 (\Sigma_e + \sum_{i=1}^m \varphi_i \mathbf{G}_i) = \sigma^2 \Sigma$. The likelihood of θ, for a known observation vector, is

$$f_\theta(\mathbf{y}) = (2\pi)^{-n/2} (\sigma^2)^{-n/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta) \right\}.$$

The likelihood function is

$$l(\theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |\Sigma| - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

The components of the vector of scores are

$$S_\beta = \frac{1}{\sigma^2} \mathbf{X}' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta), \tag{1.20}$$

$$S_{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta), \tag{1.21}$$

$$S_{\varphi_i} = -\frac{1}{2} tr(\Sigma^{-1} \mathbf{G}_i) + \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{G}_i \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta), \quad i = 1, \dots, m. \tag{1.22}$$

By making $S_\beta = \mathbf{0}$ and $\mathbf{S}_{\sigma^2} = \mathbf{0}$ we obtain

$$\beta = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{y} \quad \text{and} \quad \sigma^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

Partial derivatives of the log-likelihood function are

$$\begin{array}{ll}
H_{\beta\beta} = -\frac{1}{\sigma^2} \mathbf{X}' \Sigma^{-1} \mathbf{X}, & H_{\beta\sigma^2} = -\frac{1}{\sigma^4} \mathbf{X}' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta), \\
H_{\beta\varphi_i} = -\frac{1}{\sigma^2} \mathbf{X}' \Sigma^{-1} \mathbf{G}_i \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta), & H_{\sigma^2\sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta), \\
H_{\sigma^2\varphi_i} = -\frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{G}_i \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta), & \\
H_{\varphi_i\varphi_j} = \frac{1}{2} tr(\Sigma^{-1} \mathbf{G}_j \Sigma^{-1} \mathbf{G}_i) - \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{G}_j \Sigma^{-1} \mathbf{G}_i \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta). &
\end{array}$$

Taking expectations and changing the sign, we obtain the elements of the Fisher information matrix, i.e.

$$\begin{array}{lll}
F_{\beta\beta} = \frac{1}{\sigma^2} \mathbf{X}' \Sigma^{-1} \mathbf{X}, & F_{\beta\sigma^2} = \mathbf{0}, & F_{\beta\varphi_i} = \mathbf{0}, \\
F_{\sigma^2\sigma^2} = \frac{n}{2\sigma^4}, & F_{\sigma^2\varphi_i} = \frac{1}{2\sigma^2} tr(\Sigma^{-1} \mathbf{G}_i), & F_{\varphi_i\varphi_j} = \frac{1}{2} tr(\Sigma^{-1} \mathbf{G}_j \Sigma^{-1} \mathbf{G}_i).
\end{array}$$

1.4 Residual maximum likelihood estimation

1.4.1 Description of the method

Residual maximum likelihood estimation (REML) is introduced to reduce the bias of the maximum likelihood estimators of the variance components. For this sake, it transforms the vector **y** in two independent vectors $\mathbf{y}_1^\star = \mathbf{K}_1\mathbf{y}$ and $\mathbf{y}_2^\star = \mathbf{K}_2\mathbf{y}$, with the condition that the distribution of $\mathbf{y}_1^\star$ does not depend on the fixed effect β. Let $\mathbf{K}_1$ be a matrix such that $\mathbf{K}_1\mathbf{X} = \mathbf{0}$. Therefore

$$E[\mathbf{y}_1^\star] = E[\mathbf{K}_1\mathbf{y}] = E[\mathbf{K}_1(\mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \ldots \mathbf{Z}_m u_m + \mathbf{e})] = \mathbf{0}.$$

The vector $\mathbf{y}_2^\star$ is selected to be independent of $\mathbf{y}_1^\star$. Then it has to satisfy

$$E[\mathbf{y}_1^\star \mathbf{y}_2^{\star t}] = \mathbf{K}_1 E[\mathbf{y}\mathbf{y}']\mathbf{K}_2' = \mathbf{K}_1\mathbf{V}\mathbf{K}_2' = \mathbf{0}.$$

Rows $\mathbf{k}'$ of matrix $\mathbf{K}_1$ are called *contrasts*, as they fulfill $\mathbf{k}'\mathbf{X} = \mathbf{0}$. The maximum number of contrasts linearly independent is $n - \mathrm{r}(\mathbf{X})$. We suppose that $\mathbf{X}$ has full rank p, so that rank of $\mathbf{K}_1$ is $n - p$. Matrix $\mathbf{K}_2$ is selected with rank p.

To introduce matrix $\mathbf{K}_1$, we consider the model without random effects

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \qquad \text{with} \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_\varepsilon). \tag{1.23}$$

The maximum likelihood estimator of β in (1.23) is

$$\widetilde{\beta} = \left(\mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{y}.$$

We define the transformed vector (normalized residual)

$$\mathbf{y}_1^\star = \Sigma_\varepsilon^{-1}(\mathbf{y} - \mathbf{X}\widetilde{\beta}) = \Sigma_\varepsilon^{-1}\left(\mathbf{y} - \mathbf{X}(\mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{y}\right) = \mathbf{K}_1\mathbf{y},$$

where $\mathbf{K}_1 = \Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}\mathbf{X}(\mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_\varepsilon^{-1}$. Further we select $\mathbf{K}_2 = \mathbf{X}'\mathbf{V}^{-1}$.

Since $\mathbf{K}_1 = \mathbf{K}_1'$, it holds that

$$\begin{aligned}
E[\mathbf{y}_1^\star] &= E[\mathbf{K}_1\mathbf{y}] = \left(\Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}\mathbf{X}(\mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_\varepsilon^{-1}\right)\mathbf{X}\beta = \mathbf{0},\\
E[\mathbf{y}_2^\star] &= E[\mathbf{K}_2\mathbf{y}] = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta,\\
V[\mathbf{y}_1^\star] &= E[\mathbf{y}_1^\star\mathbf{y}_1^{\star t}] = \mathbf{K}_1\mathbf{V}\mathbf{K}_1,\\
V[\mathbf{y}_2^\star] &= \mathbf{K}_2\mathbf{V}\mathbf{K}_2' = \mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X},\\
E[\mathbf{y}_1^\star\mathbf{y}_2^{\star t}] &= \mathbf{K}_1 E[\mathbf{y}\mathbf{y}']\mathbf{K}_2' = \mathbf{K}_1\mathbf{V}\mathbf{K}_2' = \mathbf{K}_1\mathbf{V}\mathbf{V}^{-1}\mathbf{X} = \mathbf{K}_1\mathbf{X} = \mathbf{0}.
\end{aligned}$$

As the maximum number of columns linearly independent of $\mathbf{K}_1$ is $n - \mathrm{r}(\mathbf{X})$, after the selection of $n - \mathrm{r}(\mathbf{X})$ of these columns we can construct a sub-matrix $\mathbf{K}$ with the order $n \times (n - \mathrm{r}(\mathbf{X}))$ and satisfying $\mathbf{K}'\mathbf{X} = \mathbf{0}$. We define the vectors $\mathbf{y}_1 = \mathbf{K}'\mathbf{y}$ and $\mathbf{y}_2 = \mathbf{y}_2^\star$. Since $\mathrm{r}(\mathbf{X}) = p$ we have that

$$\mathbf{y}_1 \sim \mathcal{N}_{n-p}(\mathbf{0}, \mathbf{K}'\mathbf{V}\mathbf{K}), \quad \mathbf{y}_2 \sim \mathcal{N}_p(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta, \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) \quad \text{are independent.}$$

We define $\sigma = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_m^2)'$ and $\mathbf{P} = \mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'$. The likelihood function of $\mathbf{y}_1$ is

$$l(\sigma) = -\frac{1}{2}(n-p)\log 2\pi - \frac{1}{2}\log|\mathbf{K}'\mathbf{V}\mathbf{K}| - \frac{1}{2}\mathbf{y}_1'(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{y}_1,$$

where $\mathbf{V} = \sum_{i=0}^{m} \sigma_i^2 \mathbf{G}_i$ and $\mathbf{y}_1 = \mathbf{K}'\mathbf{y}$. By taking partial derivatives with respect to σ_i^2, we obtain

$$\begin{aligned} S_{\sigma_i^2} &= \frac{\partial l(\sigma)}{\partial \sigma_i^2} = -\frac{1}{2}\frac{\partial}{\partial \sigma_i^2}\left\{\log|\mathbf{K}'\mathbf{V}\mathbf{K}|\right\} - \frac{1}{2}\frac{\partial}{\partial \sigma_i^2}\left\{\mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{y}\right\} \\ &= -\frac{1}{2}\operatorname{tr}\left((\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{G}_i\mathbf{K}\right) + \frac{1}{2}\mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}(\mathbf{K}'\mathbf{G}_i\mathbf{K})(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{y} \\ &= -\frac{1}{2}\operatorname{tr}(\mathbf{P}\mathbf{G}_i) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{y}. \end{aligned}$$

As

$$\frac{\partial \mathbf{P}}{\partial \sigma_j^2} = \frac{\partial[\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}']}{\partial \sigma_j^2} = -\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{G}_j\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}' = -\mathbf{P}\mathbf{G}_j\mathbf{P},$$

the second order partial derivatives are

$$\frac{\partial l(\sigma)}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{1}{2}\operatorname{tr}(\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i) - \mathbf{y}'\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{y}.$$

If we take expectations and change the sign, we obtain the Fisher information matrix. To calculate this matrix we use the relations $\mathbf{P}\mathbf{X} = \mathbf{0}$ and $\mathbf{P}\mathbf{V}\mathbf{P} = \mathbf{P}$, and the following result.

$$\text{If } E[\mathbf{y}] = \mu \text{ and } \operatorname{var}[\mathbf{y}] = \mathbf{V}, \text{ then } E[\mathbf{y}'\mathbf{A}\mathbf{y}] = \operatorname{tr}(\mathbf{A}\mathbf{V}) + \mu'\mathbf{A}\mu. \tag{1.24}$$

The elements of the Fisher information matrix are

$$\begin{aligned} F_{\sigma_j^2\sigma_i^2} &= -E\left[\frac{\partial l(\sigma)}{\partial \sigma_i^2 \partial \sigma_j^2}\right] = -\frac{1}{2}\operatorname{tr}(\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i) + \operatorname{tr}(\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{V}) + \beta'\mathbf{X}'\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{X}\beta \\ &= -\frac{1}{2}\operatorname{tr}(\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i) + \operatorname{tr}(\mathbf{G}_j\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{V}\mathbf{P}) = \frac{1}{2}\operatorname{tr}(\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i). \end{aligned}$$

To calculate the residual maximum likelihood estimators, the Fisher-scoring method uses the following updating formula

$$\sigma_{k+1} = \sigma_k + F(\sigma_k)^{-1}S(\sigma_k),$$

where $\mathbf{F}(\sigma_k)$ is the Fisher information matrix calculated in σ_k. We observe that $\mathbf{F}(\sigma)$ is a matrix $(m+1)\times(m+1)$; however the Fisher information matrix needed to calculate maximum likelihood estimators, $\mathbf{F}(\theta)$, is $(p+m+1)\times(p+m+1)$.

Fisher-scoring algorithm gives the estimate of σ. If we plug that estimate in the likelihood function of $\mathbf{y}_2$, we consider it as a constant, and we maximize on β, we get the REML estimators of β. The likelihood function of $\mathbf{y}_2$ is

$$l(\beta) = -\frac{p}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{y}_2 - \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta)'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}(\mathbf{y}_2 - \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta).$$

By taking partial derivatives with respect to β, and equating to zero, we obtain

$$0 = \frac{\partial l(\beta)}{\partial \beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}(\mathbf{y}_2 - \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta) = \mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta).$$

Therefore

$$\widehat{\beta}_{REML} = \left(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{y}_2 = \left(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

where $\widehat{\mathbf{V}} = \sum_{i=0}^{m} \widehat{\sigma}_i^2 \mathbf{G}_i$ and $\widehat{\sigma}_0^2, \widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_m^2$ are the REML estimators of $\sigma_0^2, \sigma_1^2, \ldots, \sigma_m^2$.

By taking again derivatives with respect to β, we get

$$\mathbf{F}_{\beta\beta} = -E\left[\partial^2 l(\beta)/\partial\beta\partial\beta'\right] = \mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X},$$

that is the same value of $\mathbf{F}_{\beta\beta}$ obtained with the maximum likelihood procedure.

Theorem 1.4.1 implies that residual maximum likelihood method does not depend on the selected matrix $\mathbf{K}$ (with $\mathbf{K}'\mathbf{X} = \mathbf{0}$).

Theorem 1.4.1. Let $\mathbf{K}'$ be a full rank $(n-r)\times n$ matrix. Let $\mathbf{V}$ be a symmetric and positive definite $n\times n$ matrix. Let $\mathbf{X}$ an $n\times p$ matrix with rank $r \leq p$. If $\mathbf{K}'\mathbf{X} = \mathbf{0}$, then

$$\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}' = \mathbf{P}, \qquad \text{with} \quad \mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}.$$

1.4.2 Residual maximum likelihood with alternative parametrization

In the model (1.5), we consider the parameters

$$\sigma^2 = \sigma_0^2, \quad \varphi_i = \sigma_i^2/\sigma_0^2, \quad i = 1, \ldots, m.$$

Let $\varphi' = (\sigma^2, \varphi_1, \ldots, \varphi_m)$, $\theta' = (\beta', \varphi')$ and $\mathbf{V} = \sigma^2\left(\Sigma_e + \sum_{i=1}^{m}\varphi_i \mathbf{G}_i\right) = \sigma^2\Sigma$. For the REML method, the log-likelihood associated to the this parametrization is

$$l(\varphi) = -\frac{1}{2}(n-p)\log 2\pi - \frac{1}{2}(n-p)\log\sigma^2 - \frac{1}{2}\log|\mathbf{K}'\Sigma\mathbf{K}| - \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}' = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$. The components of the vector of scores are

$$\begin{aligned} S_{\sigma^2} &= -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{y}, \\ S_{\varphi_i} &= -\frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{G}_i) + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{y}, \quad i = 1, \ldots, m. \end{aligned}$$

Second partial derivatives of the log-likelihood are

$$\begin{aligned} &H_{\sigma^2\sigma^2} = \tfrac{n-p}{2\sigma^4} - \tfrac{1}{\sigma^6}\mathbf{y}'\mathbf{P}\mathbf{y}, \quad H_{\sigma^2\varphi_i} = -\tfrac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{y}, \\ &H_{\varphi_i\varphi_j} = \tfrac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i) - \tfrac{1}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i\mathbf{P}\mathbf{y}. \end{aligned}$$

By taking expectations, changing the sign and applying $\mathbf{P}\mathbf{X} = \mathbf{0}$ and $\mathbf{P}\Sigma\mathbf{P} = \mathbf{P}$, we obtain the elements of the Fisher information matrix

$$F_{\sigma^2\sigma^2} = -\frac{n-p}{2\sigma^4} + \frac{1}{\sigma^4}\mathrm{tr}(\mathbf{P}\Sigma) = \frac{n-p}{2\sigma^4}, \quad F_{\sigma^2\varphi_i} = \frac{1}{2\sigma^2}\mathrm{tr}(\mathbf{P}\mathbf{G}_i), \quad F_{\varphi_i\varphi_j} = \frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{G}_j\mathbf{P}\mathbf{G}_i).$$

Observation 1.4.1. From equation $S_{\sigma^2} = \mathbf{0}$, we get

$$\widehat{\sigma}^2 = \frac{1}{n-p}\mathbf{y}'\mathbf{P}\mathbf{y} \tag{1.25}$$

which allows to introduce an algorithm that updates σ^2 with (1.25) and the remaining componentes of φ with

$$\varphi^{i+1} = \varphi^i + F(\varphi^i)^{-1}S(\varphi^i).$$

1.5 The Henderson 3 method

1.5.1 Description of the method

The maximum likelihood method gives at the same time the estimates of models coefficients β and components of variance $\sigma_1^2,\ldots,\sigma_m^2$. In this section we present the *method of fitting constants* to estimate the components of variance. The regression parameter β is estimated by the least squared method and random effects are predicted by using the BLUP theory, but replacing the components of variance by its obtained estimates. The predictor of $\mathbf{u}$ is called EBLUP (empirical BLUP). The method of fitting constants is also known as *Henderson 3 method* since its introduction by Henderson (1953). We write the general linear mixed model, $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, in the form

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{e}, \tag{1.26}$$

where $\mathbf{e} \sim N(\mathbf{0}, \sigma_0^2\mathbf{W}^{-1})$ and $\mathbf{W}$ is a known symmetric and positive definite matrix. We assume that $\mathbf{X}'\mathbf{W}\mathbf{X}$ and $\mathbf{X}_1'\mathbf{W}\mathbf{X}_1$ are invertible. The partition simply divides β in two groups of effects β_1 and β_2, without taking into account if they represent fixed or random effects. This issue will be considered later.

We apply the transformation

$$\mathbf{W}^{1/2}\mathbf{y} = \mathbf{W}^{1/2}\mathbf{X}_1\beta_1 + \mathbf{W}^{1/2}\mathbf{X}_2\beta_2 + \mathbf{W}^{1/2}\mathbf{e}$$

and we denote $\mathbf{y}^* = \mathbf{W}^{1/2}\mathbf{y}$, $\mathbf{X}_1^* = \mathbf{W}^{1/2}\mathbf{X}_1$, $\mathbf{X}_2^* = \mathbf{W}^{1/2}\mathbf{X}_2$ and $\mathbf{e}^* = \mathbf{W}^{1/2}\mathbf{e}$. The new model is

$$\mathbf{y}^* = \mathbf{X}_1^*\beta_1 + \mathbf{X}_2^*\beta_2 + \mathbf{e}^*, \tag{1.27}$$

with $\mathbf{e}^* \sim N(0, \sigma_0^2\mathbf{I}_n)$.

If we fit the model (1.27) under the assumption that β_1 and β_2 are fixed effects, the total sum of squares is

$$SST = \mathbf{y}^{*\prime}\mathbf{y}^* = \mathbf{y}'\mathbf{W}\mathbf{y}. \tag{1.28}$$

The residual sum of squares is

$$SSE(\beta_1,\beta_2) = \mathbf{y}'\mathbf{M}\mathbf{y}, \tag{1.29}$$

where $\mathbf{M} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]'\mathbf{W}[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]$. The reduction of sum of squares (regression sum of squares) is

$$SSR(\beta_1,\beta_2) = SST - SSE(\beta_1,\beta_2) = \mathbf{y}'Q\mathbf{y},$$

where $Q = \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}$.

If we fit the submodel

$$\mathbf{y}^* = \mathbf{X}_1^*\beta_1 + \mathbf{e}^*,$$

under the assumption that β_1 is a fixed effect, the residual sum of squares is

$$SSE(\beta_1) = \mathbf{y}'\mathbf{M}_1\mathbf{y}, \tag{1.30}$$

where $\mathbf{M}_1 = [\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}]'\mathbf{W}[\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}]$. The reduction of the sum of squares (regression sum of squares) is

$$SSR(\beta_1) = SST - SSE(\beta_1) = \mathbf{y}'Q_1\mathbf{y},$$

where $Q_1 = \mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}$. The reduction of the sum of squares because of the introduction of $\mathbf{X}_2$ in the model, that only had $\mathbf{X}_1$, is

$$SSR(\beta_2|\beta_1) = SSR(\beta_1,\beta_2) - SSR(\beta_1) = SSE(\beta_1) - SSE(\beta_1,\beta_2).$$

To introduce the Henderson 3 method, we first calculate the expectation of $SSR(\beta_2|\beta_1)$ and $SSR(\beta_1,\beta_2)$. In a second step we modify these statistics to make them unbiased. Note that all the considered sums of squares are quadratic functions of $\mathbf{y}$, so that we will apply (1.24) systematically. For a general linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, where β may contain fixed or random effects, we have $E[\mathbf{y}] = \mathbf{X}E[\beta]$ and $\text{var}[\mathbf{y}] = \mathbf{X}\text{var}[\beta]\mathbf{X}' + \sigma_0^2\mathbf{W}^{-1}$. From (1.24), we obtain

$$\begin{aligned}
E[\mathbf{y}'Q\mathbf{y}] &= \text{tr}\left(Q\left[\mathbf{X}\text{var}[\beta]\mathbf{X}' + \sigma_0^2\mathbf{W}^{-1}\right]\right) + E[\beta]'\mathbf{X}'Q\mathbf{X}E[\beta] \\
&= \text{tr}\left(Q\mathbf{X}\text{var}[\beta]\mathbf{X}'\right) + \sigma_0^2\text{tr}\left(Q\mathbf{W}^{-1}\right) + \text{tr}\left(Q\mathbf{X}E[\beta]E[\beta]'\mathbf{X}'\right) \\
&= \text{tr}\left(Q\mathbf{X}E[\beta\beta']\mathbf{X}'\right) + \sigma_0^2\text{tr}\left(Q\mathbf{W}^{-1}\right) \\
&= \text{tr}\left(\mathbf{X}'Q\mathbf{X}E[\beta\beta']\right) + \sigma_0^2\text{tr}\left(Q\mathbf{W}^{-1}\right).
\end{aligned}$$

The expectation of the total sum of squares appearing in (1.28) is

$$E[SST] = E[\mathbf{y}'\mathbf{W}\mathbf{y}] = \text{tr}\left(\mathbf{X}'\mathbf{W}\mathbf{X}E[\beta\beta']\right) + \sigma_0^2\text{tr}(\mathbf{I}_n) = \text{tr}\left(\mathbf{X}'\mathbf{W}\mathbf{X}E[\beta\beta']\right) + n\sigma_0^2 \tag{1.31}$$

The expectation of the sum of residual squares in (1.29) is

$$E[SSE(\beta_1,\beta_2)] = E[\mathbf{y}'\mathbf{M}\mathbf{y}] = \text{tr}\left(\mathbf{X}'\mathbf{M}\mathbf{X}E[\beta\beta']\right) + \sigma_0^2\text{tr}\left(\mathbf{M}\mathbf{W}^{-1}\right).$$

This expression can be simplified if we take into account that

$$\begin{aligned}
\mathbf{X}'\mathbf{M}\mathbf{X} &= \mathbf{X}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]'\mathbf{W}[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]\mathbf{X} = \mathbf{X}'\mathbf{W}\mathbf{X} \\
&- 2\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X} \\
&= \mathbf{0}
\end{aligned}$$

and

$$\begin{aligned}\mathbf{M}\mathbf{W}^{-1} &= [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]'\mathbf{W}[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]\mathbf{W}^{-1}\\ &= [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]'[\mathbf{I}_n - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}']\\ &= \mathbf{I}_n - 2\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}' + \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\\ &= \mathbf{I}_n - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}',\end{aligned}$$

Since $\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$ is equal to the identity, we obtain that

$$\begin{aligned}\mathrm{tr}\left(\mathbf{M}\mathbf{W}^{-1}\right) &= \mathrm{tr}\left(\mathbf{I}_n - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\right) = n - \mathrm{tr}\left(\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\right)\\ &= n - p = n - \mathrm{r}(\mathbf{X}),\end{aligned}$$

where $\mathrm{r}(\mathbf{X})$ denotes the rank of $\mathbf{X}$. This result can be proved in the case $\mathrm{r}(\mathbf{X}) < p$ too. Therefore

$$E[SSE(\beta_1,\beta_2)] = \sigma_0^2[n - \mathrm{r}(\mathbf{X})] \tag{1.32}$$

and also with (1.31) and (1.32) we obtain that

$$E[SSR(\beta_1,\beta_2)] = E[SST] - E[SSE(\beta_1,\beta_2)] = \mathrm{tr}\left(\mathbf{X}'\mathbf{W}\mathbf{X}E[\beta\beta']\right) + \sigma_0^2\mathrm{r}(\mathbf{X}).$$

From the model (1.26) it follows that

$$\mathbf{X}'\mathbf{W}\mathbf{X} = \begin{pmatrix}\mathbf{X}_1'\\ \mathbf{X}_2'\end{pmatrix}\mathbf{W}(\mathbf{X}_1,\ \mathbf{X}_2) = \begin{pmatrix}\mathbf{X}_1'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{W}\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{W}\mathbf{X}_2\end{pmatrix},$$

consequently

$$E[SSR(\beta_1,\beta_2)] = \mathrm{tr}\left\{\begin{pmatrix}\mathbf{X}_1'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{W}\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{W}\mathbf{X}_2\end{pmatrix}E[\beta\beta']\right\} + \sigma_0^2\mathrm{r}(\mathbf{X}). \tag{1.33}$$

From (1.30) and (1.24) we obtain

$$\begin{aligned}E[SSE(\beta_1)] &= \mathrm{tr}\left\{\mathbf{X}'\mathbf{M}_1\mathbf{X}E[\beta\beta']\right\} + \sigma_0^2\mathrm{tr}\left\{\mathbf{M}_1\mathbf{W}^{-1}\right\}\\ &= \mathrm{tr}\left\{\mathbf{X}'\mathbf{M}_1\mathbf{X}E[\beta\beta']\right\} + \sigma_0^2[n - \mathrm{r}\{\mathbf{X}_1\}].\end{aligned} \tag{1.34}$$

From (1.31) and (1.34), we have that

$$E[SSR(\beta_1)] = E[SST] - E[SSE(\beta_1)] = \mathrm{tr}\left\{\mathbf{X}'Q_1\mathbf{X}E[\beta\beta']\right\} + \sigma_0^2\mathrm{r}\{\mathbf{X}_1\},$$

where $Q_1 = \mathbf{W} - \mathbf{M}_1 = \mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}$. If $\mathbf{X}_1'\mathbf{W}\mathbf{X}_1$ is invertible, then

$$\begin{aligned}\mathbf{X}'Q_1\mathbf{X} &= \begin{pmatrix}\mathbf{X}_1'\\ \mathbf{X}_2'\end{pmatrix}\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}(\mathbf{X}_1\ \mathbf{X}_2)\\ &= \begin{pmatrix}\mathbf{X}_1'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}\mathbf{X}_2\end{pmatrix}\\ &= \begin{pmatrix}\mathbf{X}_1'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{W}\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}\mathbf{X}_2\end{pmatrix}\end{aligned}$$

and

$$E[SSR(\beta_1)] = \text{tr}\left\{\begin{pmatrix} \mathbf{X}_1'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{W}\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{W}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}\mathbf{X}_2 \end{pmatrix} E[\beta\beta']\right\} + \sigma_0^2 \text{r}(\mathbf{X}_1). \tag{1.35}$$

Therefore, applying (1.33) and (1.35), we obtain

$$\begin{aligned} & E[SSR(\beta_2|\beta_1)] = E[SSR(\beta_1,\beta_2)] - E[SSR(\beta_1)] \\ = \;& \text{tr}\left\{\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1']\mathbf{W}\mathbf{X}_2 \end{pmatrix} E[\beta\beta']\right\} + \sigma_0^2[\text{r}(\mathbf{X}) - \text{r}(\mathbf{X}_1)] \\ = \;& \text{tr}\left\{\mathbf{X}_2'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1']\mathbf{W}\mathbf{X}_2 E[\beta_2\beta_2{}']\right\} + \sigma_0^2[\text{r}(\mathbf{X}) - \text{r}(\mathbf{X}_1)]. \end{aligned} \tag{1.36}$$

We observe that $E[SSR(\beta_2|\beta_1)]$ is simply a function of $E[\beta_2\beta_2']$ and σ_0^2. It does not depend on $E[\beta_1\beta_1']$ and $E[\beta_1\beta_2']$. We also observe that (1.36) has been obtained without doing assumptions about the form of $E[\beta\beta']$. Therefore (1.36) holds for any structure of covariance matrix of β.

Let us consider again the model (1.5)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \ldots + \mathbf{Z}_m\mathbf{u}_m + \mathbf{e},$$

with $\mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, \sigma_0^2\mathbf{W}^{-1})$, and $\mathbf{u}_i \sim \mathcal{N}_{q_i}(\mathbf{0}, \sigma_i^2\mathbf{I}_{q_i})$, $i = 1, \ldots, m$. We define

$$\beta^{(i)} = (\beta', \mathbf{u}_1', \ldots, \mathbf{u}_{i-1}')' \quad \text{y} \quad \mathbf{u}^{(i)} = (\mathbf{u}_i', \ldots, \mathbf{u}_m')'.$$

For $i = 1, \ldots, m$ we consider the case

$$\begin{aligned} \mathbf{X}_1 = \mathbf{X}_1^{(i)} = (\mathbf{X}, \mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1}), \quad & \beta_1 = \beta^{(i)}, \\ \mathbf{X}_2 = \mathbf{X}_2^{(i)} = (\mathbf{Z}_i, \ldots, \mathbf{Z}_m), \qquad\quad & \beta_2 = \mathbf{u}^{(i)} \end{aligned}$$

and define

$$\begin{aligned} \mathbf{M}_i &= \mathbf{W} - \mathbf{W}\mathbf{X}_1^{(i)}(\mathbf{X}_1^{(i)t}\mathbf{W}\mathbf{X}_1^{(i)})^{-1}\mathbf{X}_1^{(i)t}\mathbf{W}, \\ \mathbf{L}_i &= \mathbf{Z}_i'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}_1^{(i)}(\mathbf{X}_1^{(i)t}\mathbf{W}\mathbf{X}_1^{(i)})^{-1}\mathbf{X}_1^{(i)t}]\mathbf{W}\mathbf{Z}_i. \end{aligned}$$

Then (1.32) and (1.36) becomes

$$E[SSE(\beta^{(i)}, \mathbf{u}^{(i)})] = E[SSE(\beta, \mathbf{u})] = \sigma_0^2[n - \text{r}(\mathbf{X}\,\mathbf{Z})] \tag{1.37}$$

$$E[SSR(\mathbf{u}^{(i)}|\beta^{(i)})] = \sum_{k=i}^{m} \text{tr}\{\mathbf{L}_k\}\sigma_k^2 + \sigma_0^2[\text{r}(\mathbf{X}\,\mathbf{Z}) - \text{r}(\mathbf{X}, \mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1})] \tag{1.38}$$

From (1.37) and (1.38), and applying the method of moments, we get the following linear and triangular system of equations.

$$\begin{aligned} SSE(\beta, \mathbf{u}) &= \sigma_0^2[n - \text{r}(\mathbf{X}, \mathbf{Z}_1, \ldots, \mathbf{Z}_m)] \\ SSR(\mathbf{u}^{(m)}|\beta^{(m)}) &= \sigma_0^2[\text{r}(\mathbf{X}\,\mathbf{Z}) - \text{r}(\mathbf{X}, \mathbf{Z}_1, \ldots, \mathbf{Z}_{m-1})] + \sigma_m^2\text{tr}\{\mathbf{L}_m\} \\ SSR(\mathbf{u}^{(m-1)}|\beta^{(m-1)}) &= \sigma_0^2[\text{r}(\mathbf{X}\,\mathbf{Z}) - \text{r}(\mathbf{X}, \mathbf{Z}_1, \ldots, \mathbf{Z}_{m-2})] + \sigma_m^2\text{tr}\{\mathbf{L}_m\} + \sigma_{m-1}^2\text{tr}\{\mathbf{L}_{m-1}\} \\ &\vdots \\ SSR(\mathbf{u}^{(1)}|\beta^{(1)}) &= \sigma_0^2[\text{r}(\mathbf{X}\,\mathbf{Z}) - \text{r}(\mathbf{X})] + \sum_{i=1}^{m} \sigma_i^2\text{tr}\{\mathbf{L}_i\} \end{aligned}$$

From the first equation we obtain an unbiased estimator of σ_0^2,

$$\widehat{\sigma}_0^2 = \frac{SSE(\beta,\mathbf{u})}{n - \mathrm{r}(\mathbf{X}\,\mathbf{Z})} = MSE(\beta,\mathbf{u}). \tag{1.39}$$

From the second equation we get an unbiased estimator of σ_m^2,

$$\widehat{\sigma}_m^2 = \frac{SSR(\mathbf{u}^{(m)}|\beta^{(m)}) - \widehat{\sigma}_0^2[\mathrm{r}(\mathbf{X}\,\mathbf{Z}) - \mathrm{r}(\mathbf{X},\mathbf{Z}_1,\ldots,\mathbf{Z}_{m-1})]}{\mathrm{tr}\{\mathbf{L}_m\}}. \tag{1.40}$$

From the third equation we get an unbiased estimator of σ_{m-1}^2,

$$\widehat{\sigma}_{m-1}^2 = \frac{SSR(\mathbf{u}^{(m-1)}|\beta^{(m-1)}) - \widehat{\sigma}_0^2[\mathrm{r}(\mathbf{X}\,\mathbf{Z}) - \mathrm{r}(\mathbf{X},\mathbf{Z}_1,\ldots,\mathbf{Z}_{m-2})] - \widehat{\sigma}_m^2\mathrm{tr}\{\mathbf{L}_m\}}{\mathrm{tr}\{\mathbf{L}_{m-1}\}},$$

and so on.

As $SSR(\mathbf{u}^{(i)}|\beta^{(i)}) = SSE(\beta^{(i)}) - SSE(\beta^{(i)},\mathbf{u}^{(i)}) = SSE(\beta^{(i)}) - SSE(\beta,\mathbf{u})$, then the previous formula can be expressed as a function of residual sum of squares. That is,

$$\begin{aligned}
\widehat{\sigma}_0^2 &= \frac{\mathbf{y}'\mathbf{M}_{m+1}\mathbf{y}}{n - \mathrm{r}(\mathbf{X}_1^{(m+1)})} \\
\widehat{\sigma}_m^2 &= \frac{\mathbf{y}'\mathbf{M}_m\mathbf{y} - \mathbf{y}'\mathbf{M}_{m+1}\mathbf{y} - \widehat{\sigma}_0^2\left[\mathrm{r}(X_1^{(m+1)}) - \mathrm{r}(X_1^{(m)})\right]}{\mathrm{tr}(\mathbf{L}_m)} \\
&\vdots \\
\widehat{\sigma}_i^2 &= \frac{\mathbf{y}'\mathbf{M}_i\mathbf{y} - \mathbf{y}'\mathbf{M}_{m+1}\mathbf{y} - \widehat{\sigma}_0^2\left[\mathrm{r}(X_1^{(m+1)}) - \mathrm{r}(X_1^{(i)})\right] - \sum_{j=i+1}^m \widehat{\sigma}_j^2\mathrm{tr}(\mathbf{L}_j)}{\mathrm{tr}(\mathbf{L}_i)} \\
&\vdots \\
\widehat{\sigma}_1^2 &= \frac{\mathbf{y}'\mathbf{M}_1\mathbf{y} - \mathbf{y}'\mathbf{M}_{m+1}\mathbf{y} - \widehat{\sigma}_0^2\left[\mathrm{r}(X_1^{(m+1)}) - \mathrm{r}(X_1^{(1)})\right] - \sum_{j=2}^m \widehat{\sigma}_j^2\mathrm{tr}(\mathbf{L}_j)}{\mathrm{tr}(\mathbf{L}_1)}
\end{aligned}$$

For more details see the Searle at al. (1992), 202-208, or Searle (1971), 443-445. If we replace the variance components $\sigma_0^2,\sigma_1^2,\ldots,\sigma_m^2$ by their estimators $\widehat{\sigma}_0^2,\widehat{\sigma}_1^2,\ldots,\widehat{\sigma}_m^2$ in (1.3) and (1.4), we obtain the estimator of β and the predictors $\mathbf{u}_1,\ldots,\mathbf{u}_m$.

Observation 1.5.1. If we use the alternative parametrization the system of equations is not linear any more. Consequently, by solving the transformed system one does not obtain unbiased estimators.

1.6 The area-level Fay-Herriot model

1.6.1 The model

Let us introduce the following notations and assumptions:

1. Let $\mathbf{x}_d = (x_{d1}, \ldots, x_{dp})$ be known vectors containing explanatory variables for the target variable $\mu_d = \overline{Y}_{d\cdot}$, $d = 1, \ldots, D$, where $\overline{Y}_{d\cdot}$ is the domain mean of variable y.
2. Assume that the μ_d's are independent with distribution $N(\mathbf{x}_d \beta, \sigma_u^2)$, where β is a vector of dimension p containing the regression parameters, i.e. $\mu = (\mu_1, \ldots, \mu_D)' \sim N(\mathbf{X}\beta, \Sigma_u)$ with $\Sigma_u = \sigma_u^2 \mathbf{I}_D$.
3. Let $\overline{\mathbf{y}} = (\overline{y}_{1\cdot}, \ldots, \overline{y}_{D\cdot})'$ be a vector of direct estimators of μ with distribution $N(\mu, \mathbf{V}_e)$, where $\mathbf{V}_e = \text{diag}(\sigma_1^2, \ldots, \sigma_D^2)$ and the diagonal elements σ_d^2 are assumed to be known.

The area-level Fay-Herriot model is

$$\overline{y}_{d\cdot} = \mu_d + e_d \quad \text{y} \quad \mu_d = \mathbf{x}_d \beta + u_d, \quad d = 1, \ldots, D, \tag{1.41}$$

where $\mathbf{e} = (e_1, \ldots, e_D)$ and $\mathbf{u} = (u_1, \ldots, u_D)$ are independent with distribution $N(\mathbf{0}, \mathbf{V}_e)$ and $N(\mathbf{0}, \Sigma_u)$ respectively. If we write (1.41) in the form $\mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$, we get

$$\begin{pmatrix} \overline{y}_{1\cdot} \\ \vdots \\ \overline{y}_{D\cdot} \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{D1} & \ldots & x_{Dp} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_D \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_D \end{pmatrix}.$$

It holds that $\mathbf{Z} = \mathbf{I}_D$, $\text{tr}(\mathbf{Z}'\mathbf{Z}) = D$, $r(X, Z) = D$, $Cov[\mathbf{y}, \mathbf{u}] = \mathbf{Z}\Sigma_u$,

$$\mathbf{V} = \text{var}(\mathbf{y}) = \mathbf{Z}\Sigma_u \mathbf{Z}' + \mathbf{V}_e = \Sigma_u + \mathbf{V}_e = \text{diag}(\sigma_u^2 + \sigma_1^2, \ldots, \sigma_u^2 + \sigma_D^2),$$

and

$$\mathbf{V}^{-1} = \text{diag}((\sigma_u^2 + \sigma_1^2)^{-1}, \ldots, (\sigma_u^2 + \sigma_D^2)^{-1}).$$

If σ_u^2 is known, then the best linear unbiased estimator (BLUE) and predictor (BLUP) of $\beta = (\beta_1, \ldots, \beta_p)'$ and $\mathbf{u} = (u_1, \ldots, u_D)'$ are

$$\widetilde{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\overline{\mathbf{y}} \qquad \text{and} \qquad \widetilde{\mathbf{u}} = \Sigma_u \mathbf{Z}'\mathbf{V}^{-1}\left(\overline{\mathbf{y}} - \mathbf{X}\widetilde{\beta}\right).$$

It is easy to check that the components of $\widetilde{\mathbf{u}}$ are

$$\widetilde{u}_d = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2}\left(\overline{y}_{d\cdot} - \mathbf{x}_d \widetilde{\beta}\right), \qquad d = 1, \ldots, D,$$

where $\mathbf{x}_d$ is the row d of matrix $\mathbf{X}$.

The BLUP of $\mu_d = \mathbf{x}_d \beta + u_d$ is

$$\widehat{\overline{Y}}_d^{blup} = \widetilde{\mu}_d = \mathbf{x}_d \widetilde{\beta} + \widetilde{u}_d = \mathbf{x}_d \widetilde{\beta} + \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2}\left(\overline{y}_{d\cdot} - \mathbf{x}_d \widetilde{\beta}\right) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2}\overline{y}_{d\cdot} + \frac{\sigma_d^2}{\sigma_u^2 + \sigma_d^2}\mathbf{x}_d \widetilde{\beta} \tag{1.42}$$

Proposition 1.6.1. The best predictor of μ_d is

$$E[\mu_d | \overline{y}_{d\cdot}] = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2}\overline{y}_{d\cdot} + \frac{\sigma_d^2}{\sigma_u^2 + \sigma_d^2}\mathbf{x}_d \beta,$$

so that the BLUP can be obtained from the BP substituting β by $\widetilde{\beta}$.

Proof. As $\bar{y}_{d\cdot} \sim N(\mathbf{x}_d\beta, \sigma_u^2+\sigma_d^2)$, $\bar{y}_{d\cdot}|u_d \sim N(\mathbf{x}_d\beta+u_d, \sigma_d^2)$ and $u_d \sim N(0,\sigma_u^2)$, then

$$\begin{aligned} f(u_d|\bar{y}_{d\cdot}) &\propto f(\bar{y}_{d\cdot}|u_d)f(u_d) = \frac{1}{\sigma_d^2\sqrt{2\pi}}\exp\Big\{-\frac{1}{2\sigma_d^2}(y_d-\mathbf{x}_d\beta-u_d)^2\Big\}\frac{1}{\sigma_u^2\sqrt{2\pi}}\exp\Big\{-\frac{1}{2\sigma_u^2}u_d^2\Big\} \\ &\propto \exp\left\{-\frac{1}{2\frac{\sigma_d^2\sigma_u^2}{\sigma_d^2+\sigma_u^2}}\left[u_d^2-2\frac{\sigma_u^2}{\sigma_d^2+\sigma_u^2}(\bar{y}_{d\cdot}-\mathbf{x}_d\beta)u_d\right]\right\}, \end{aligned}$$

which corresponds to a normal distribution with mean $E[u_d|\bar{y}_{d\cdot}] = \frac{\sigma_u^2}{\sigma_d^2+\sigma_u^2}(\bar{y}_{d\cdot}-\mathbf{x}_d\beta)$ and variance $\mathrm{var}[u_d|\bar{y}_{d\cdot}] = \frac{\sigma_d^2\sigma_u^2}{\sigma_d^2+\sigma_u^2}$. Therefore

$$E[\mu_d|\bar{y}_{d\cdot}] = \mathbf{x}_d\beta + E[u_d|\bar{y}_{d\cdot}] = \mathbf{x}_d\beta + \frac{\sigma_u^2}{\sigma_d^2+\sigma_u^2}(\bar{y}_{d\cdot}-\mathbf{x}_d\beta) = \frac{\sigma_u^2}{\sigma_u^2+\sigma_d^2}\bar{y}_{d\cdot} + \frac{\sigma_d^2}{\sigma_u^2+\sigma_d^2}\mathbf{x}_d\beta.$$

Definition 1.6.1. The empirical BLUP (EBLUP) of the domain mean $\overline{Y}_d$, under the model (1.41) is obtained plugging an estimator $\hat{\sigma}_u^2$ in the place of σ_u^2 por un estimador $\hat{\sigma}_u^2$, i.e.

$$\widehat{\overline{Y}}_d^{FH} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2+\sigma_d^2}\bar{y}_{d\cdot} + \frac{\sigma_d^2}{\hat{\sigma}_u^2+\sigma_d^2}\mathbf{x}_d\widehat{\beta} \tag{1.43}$$

in the case that the σ_d^2's are known, or

$$\widehat{\overline{Y}}_d^{FH} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2+\hat{\sigma}_d^2}\bar{y}_{d\cdot} + \frac{\hat{\sigma}_d^2}{\hat{\sigma}_u^2+\hat{\sigma}_d^2}\mathbf{x}_d\widehat{\beta}, \tag{1.44}$$

with $\hat{\sigma}_d^2 = \widehat{V}(\bar{y}_{d\cdot})$, $d=1,\ldots,D$, otherwise.

1.6.2 Random effect variance estimation

We consider three procedures for estimating σ_u^2: (1) Moments, (2) Maximum likelihood, and (3) residual maximum likelihood.

The method of moments

An unbiased estimator of σ_u^2 1s

$$\widehat{\sigma}_u^2 = \frac{1}{D-p}\left[\sum_{d=1}^D \tilde{u}_d^2 - \sum_{d=1}^D \sigma_d^2\left(1-\mathbf{x}_d\left(\sum_{d=1}^D \mathbf{x}_d'\mathbf{x}_d\right)^{-1}\mathbf{x}_d'\right)\right],$$

where $\tilde{u}_d = \bar{y}_d - \mathbf{x}_d\tilde{\beta}$ and $\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{y}} = \left(\sum_{d=1}^D \mathbf{x}_d'\mathbf{x}_d\right)^{-1}\left(\sum_{d=1}^D \mathbf{x}_d'\bar{y}_d\right)$.

It may occur that $\widehat{\sigma}_u^2$ takes negative values, but $Pr(\widehat{\sigma}_u^2 \leq 0)$ tends to 0 when $a\to\infty$. If $\widehat{\sigma}_u^2$ is negative, we equate it to zero and we define

$$\tilde{\sigma}_u^2 = \max\left\{\widehat{\sigma}_u^2, 0\right\} \tag{1.45}$$

Maximum likelihood method

In what follows we particularize the results of Section 1.3 to the case $m=1$, $q_1=D$, $\sigma_1^2=\sigma_u^2$, $\Omega_1=I_D$. It holds that $\mathbf{y}\sim N(\mathbf{X}\beta,\mathbf{V})$, with covariance matrix $\mathbf{V}=\text{diag}_{1\leq d\leq D}(\sigma_u^2+\sigma_d^2)$. The log-likelihood is

$$\ell(\sigma_u^2,\beta;\mathbf{y})=-\frac{D}{2}\ln 2\pi-\frac{1}{2}\ln|\mathbf{V}|-\frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta).$$

The partial derivatives of the log-likelihood are

$$\begin{aligned}
\mathbf{S}_\beta &= \mathbf{X}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta)=\sum_{d=1}^D \mathbf{x}_d'\frac{1}{\sigma_u^2+\sigma_d^2}(y_d-\mathbf{x}_d\beta),\\
S_{\sigma_u^2} &= -\frac{1}{2}\text{tr}(\mathbf{V}^{-1}\mathbf{G}_u)+\frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}\mathbf{G}_u\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta)\\
&= -\frac{1}{2}\sum_{d=1}^D\frac{1}{\sigma_u^2+\sigma_d^2}+\frac{1}{2}\sum_{d=1}^D\frac{1}{(\sigma_u^2+\sigma_d^2)^2}(y_d-\mathbf{x}_d\beta)^2,
\end{aligned}$$

where $\mathbf{G}_u=\partial\mathbf{V}/\partial\sigma_u^2=\mathbf{I}_D$. To calculate the second order partial derivatives we use the formulas (1.14)-(1.16) to obtain

$$\begin{aligned}
\mathbf{H}_{\beta\beta} &= -\mathbf{X}'\mathbf{V}^{-1}\mathbf{X},\quad \mathbf{H}_{\beta\sigma_u^2}=-\mathbf{X}'\mathbf{V}^{-2}(\mathbf{y}-\mathbf{X}\beta),\\
H_{\sigma_u^2\sigma_u^2} &= \frac{1}{2}\text{tr}(\mathbf{V}^{-2})-(\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-3}(\mathbf{y}-\mathbf{X}\beta).
\end{aligned}$$

The components of the Fisher information matrix are

$$\begin{aligned}
\mathbf{F}_{\beta\beta} &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}=\sum_{d=1}^D\frac{1}{\sigma_u^2+\sigma_d^2}\mathbf{x}_d'\mathbf{x}_d,\quad \mathbf{F}_{\beta\sigma_u^2}=\mathbf{F}_{\sigma_u^2\beta}=\mathbf{0},\\
F_{\sigma_u^2\sigma_u^2} &= -\frac{1}{2}\text{tr}(\mathbf{V}^{-2})+\text{tr}(\mathbf{V}^{-3}\mathbf{V})=\frac{1}{2}\text{tr}(\mathbf{V}^{-2})=\frac{1}{2}\sum_{d=1}^D\frac{1}{(\sigma_u^2+\sigma_d^2)^2}.
\end{aligned}$$

Observation 1.6.1. Let

$$\mathbf{T}=(\mathbf{V}_e^{-1}+\sigma_u^{-2}\mathbf{I}_D)^{-1}=\sigma_u^2\mathbf{I}_D-\sigma_u^4\mathbf{V}^{-1}.$$

Applying the formula

$$(A+BCD)^{-1}=A^{-1}-A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}$$

with $A=\sigma_u^{-2}\mathbf{I}_D$, $B=\mathbf{I}_D$, $C=\mathbf{V}_e^{-1}=\text{diag}_{1\leq d\leq D}(\sigma_d^{-2})$ and $D=\mathbf{I}_D$, we get

$$\mathbf{T}=\sigma_u^2\mathbf{I}_D-\sigma_u^4\mathbf{V}^{-1}\quad\text{y}\quad \mathbf{V}^{-1}=\frac{\sigma_u^2\mathbf{I}_D-\mathbf{T}}{\sigma_u^4}.$$

Therefore

$$F_{\sigma_u^2\sigma_u^2}=\frac{1}{2\sigma_u^8}\text{tr}\left((\sigma_u^2\mathbf{I}_d-\mathbf{T})^2\right)=\frac{1}{2\sigma_u^4}\left(D-\frac{2}{\sigma_u^2}\text{tr}(\mathbf{T})+\frac{1}{\sigma_u^4}\text{tr}(\mathbf{T}^2)\right).$$

The updating formulas of the Fisher-scoring algorithm are

$$\sigma_u^{2(k+1)} = \sigma_u^{2(k)} + F^{-1}_{\sigma_u^{2(k)}\sigma_u^{2(k)}} S_{\sigma_u^{2(k)}}, \quad \beta^{(k+1)} = \beta^{(k)} + F^{-1}_{\beta^{(k)}\beta^{(k)}} S_{\beta^{(k)}}.$$

Residual maximum likelihood method

In what follows we particularize the results of Section 1.4 to the case $m = 1$, $q_1 = D$, $\varphi_1 = \sigma_u^2$, $\sigma^2 = 1$, $\Omega_1 = I_D$. The REML log-likelihood is

$$\ell_R(\sigma_u^2; \mathbf{y}) = -\frac{D-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}'\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. It holds that

$$\mathbf{y}'\frac{\partial \mathbf{P}}{\partial \sigma_u^2}\mathbf{y} = -(\mathbf{y} - \mathbf{X}\widehat{\beta})'\mathbf{V}^{-1}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}) = -\sum_{d=1}^{D}\frac{1}{(\sigma_u^2+\sigma_d^2)^2}(y_d - \mathbf{x}_d\widehat{\beta})^2,$$

and

$$\mathbf{P} = \frac{1}{\sigma_u^2}\left(\mathbf{I}_D - \frac{1}{\sigma_u^2}\mathbf{R}\right), \quad \text{tr}(\mathbf{P}) = \frac{1}{\sigma_u^2}\left[D - \frac{1}{\sigma_u^2}\text{tr}(\mathbf{R})\right],$$

where

$$\begin{aligned} \mathbf{R} &= \mathbf{T} + \mathbf{M}, \quad \mathbf{M} = \mathbf{T}\mathbf{V}_e^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_e^{-1}\mathbf{T}, \\ \mathbf{T} &= \left(\mathbf{V}_e^{-1} + \sigma_u^{-2}\mathbf{I}_D\right)^{-1} = \text{diag}_{1\le d\le D}\left(\frac{\sigma_u^2\sigma_d^2}{\sigma_u^2+\sigma_d^2}\right). \end{aligned}$$

First order derivative of the log-likelihood is

$$\frac{\partial \ell_R}{\partial \sigma_u^2} = -\frac{1}{2}\text{tr}(\mathbf{P}) - \frac{1}{2}\mathbf{y}'\frac{\partial \mathbf{P}}{\partial \sigma_u^2}\mathbf{y} = -\frac{1}{2\sigma_u^2}\left[D - \frac{1}{\sigma_u^2}\text{tr}(\mathbf{R})\right] + \frac{1}{2}\sum_{d=1}^{D}\frac{1}{(\sigma_u^2+\sigma_d^2)^2}(y_d - \mathbf{x}_d\widehat{\beta})^2.$$

Second order derivative of the log-likelihood is

$$\frac{\partial^2 \ell_R}{\partial \sigma_u^2 \partial \sigma_u^2} = \frac{1}{2}\text{tr}(\mathbf{P}^2) - \mathbf{y}'\mathbf{P}^3\mathbf{y}$$

As $\mathbf{PVP} = \mathbf{P}$, the Fisher amount of information associated to σ_u^2 is

$$F_{\sigma_u^2} = -\frac{1}{2}\text{tr}(\mathbf{P}^2) + \text{tr}(\mathbf{P}^3\mathbf{V}) = \frac{1}{2}\text{tr}(\mathbf{P}^2) = \frac{1}{2\sigma_u^4}\left[D - \frac{2}{\sigma_u^2}\text{tr}(\mathbf{R}) + \frac{1}{\sigma_u^4}\text{tr}(\mathbf{R}^2)\right].$$

The REML estimators may be obtained by applying the following Fisher-scoring algorithm.

1. Set the seeds $\widehat{\sigma}_{u,0}^2 = \widetilde{\sigma}_u^2 = \max\{\widehat{\sigma}_u^2, 0\}$ and $\widehat{\beta}_0 = \widetilde{\beta}$, where $\widehat{\sigma}_u^2$ and $\widetilde{\beta}$ are the moment estimators given by (1.45).

2. For $k = 1, 2, \dots,$ do

$$\widehat{\beta}_k = \left(\sum_{d=1}^{D} \frac{\mathbf{x}_d' \mathbf{x}_d}{\widehat{\sigma}_{u,k-1}^2 + \sigma_d^2} \right)^{-1} \left(\sum_{d=1}^{D} \frac{\mathbf{x}_d' y_d}{\widehat{\sigma}_{u,k-1}^2 + \sigma_d^2} \right), \quad \widehat{\sigma}_{u,k}^2 = \widehat{\sigma}_{u,k-1}^2 + F_{k-1}^{-1} \mathbf{S}_{k-1},$$

whre

$$\begin{aligned}
\mathbf{S}_k &= -\frac{1}{2\widehat{\sigma}_{u,k}^2} \left(D - \frac{\mathrm{tr}(\widehat{R}_k)}{\widehat{\sigma}_{u,k}^2} \right) + \frac{1}{2} \sum_{d=1}^{D} \frac{1}{(\widehat{\sigma}_{u,k}^2 + \sigma_d^2)^2} (y_d - \mathbf{x}_d \widehat{\beta}_k)^2, \\
F_k &= \frac{1}{2\widehat{\sigma}_{u,k}^4} \left(D - \frac{2}{\widehat{\sigma}_{u,k}^2} tr\{\widehat{\mathbf{R}}_k\} + \frac{1}{\widehat{\sigma}_{u,k}^4} tr\{\widehat{\mathbf{R}}_k^2\} \right), \\
tr\{\widehat{\mathbf{R}}_k\} &= \mathrm{tr}(\widehat{\mathbf{T}}_k) + \mathrm{tr}(\widehat{\mathbf{M}}_k), \quad tr\{\widehat{\mathbf{R}}_k^2\} = \mathrm{tr}(\widehat{\mathbf{T}}_k^2) + 2\mathrm{tr}(\widehat{\mathbf{T}}_k \widehat{\mathbf{M}}_k) + \mathrm{tr}(\widehat{\mathbf{M}}_k^2), \\
\mathrm{tr}(\widehat{\mathbf{T}}_k) &= \sum_{d=1}^{D} \frac{\widehat{\sigma}_{u,k}^2 \sigma_d^2}{\widehat{\sigma}_{u,k}^2 + \sigma_d^2}, \quad \mathrm{tr}(\widehat{\mathbf{T}}_k^2) = \sum_{d=1}^{D} \frac{\widehat{\sigma}_{u,k}^4 \sigma_d^4}{(\widehat{\sigma}_{u,k}^2 + \sigma_d^2)^2},
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}(\widehat{\mathbf{M}}_k) &= \mathrm{tr}\left[\left(\sum_{d=1}^{D} \frac{\widehat{\sigma}_{u,k}^4 \mathbf{x}_d' \mathbf{x}_d}{(\widehat{\sigma}_{u,k}^2 + \sigma_d^2)^2} \right) \left(\sum_{d=1}^{D} \frac{\mathbf{x}_d' \mathbf{x}_d}{\widehat{\sigma}_{u,k}^2 + \sigma_d^2} \right)^{-1} \right], \\
\mathrm{tr}(\widehat{\mathbf{T}}_k \widehat{\mathbf{M}}_k) &= \mathrm{tr}\left[\left(\sum_{d=1}^{D} \frac{\widehat{\sigma}_{u,k}^6 \sigma_d^2 \mathbf{x}_d' \mathbf{x}_d}{(\widehat{\sigma}_{u,k}^2 + \sigma_d^2)^3} \right) \left(\sum_{d=1}^{D} \frac{\mathbf{x}_d' \mathbf{x}_d}{\widehat{\sigma}_{u,k}^2 + \sigma_d^2} \right)^{-1} \right], \\
\mathrm{tr}(\widehat{\mathbf{M}}_k^2) &= \mathrm{tr}\left[\left\{ \left(\sum_{d=1}^{D} \frac{\widehat{\sigma}_{u,k}^4 \mathbf{x}_d' \mathbf{x}_d}{(\widehat{\sigma}_{u,k}^2 + \sigma_d^2)^2} \right) \left(\sum_{d=1}^{D} \frac{\mathbf{x}_d' \mathbf{x}_d}{\widehat{\sigma}_{u,k}^2 + \sigma_d^2} \right)^{-1} \right\}^2 \right].
\end{aligned}$$

3. Stop if $|\widehat{\sigma}_{u,k}^2 - \widehat{\sigma}_{u,k-1}^2| < \varepsilon$ and $\left[(\widehat{\beta}_k - \widehat{\beta}_{k-1})'(\widehat{\beta}_k - \widehat{\beta}_{k-1}) \right]^{1/2} < \varepsilon$. Output: $\widehat{\beta}_{ML} = \widehat{\beta}_k$, $\widehat{u}_d = \widehat{u}_{d,k}$ and $\widehat{\sigma}_{u,ML}^2 = \widehat{\sigma}_{u,k}^2$

Alternatively the following algorithm can be used.

1. Set the seeds $\widehat{\sigma}_{u,0}^2 = \widetilde{\sigma}_u^2 = \max\{\widehat{\sigma}_u^2, 0\}$ and $\widehat{\beta}_0 = \widetilde{\beta}$, where $\widehat{\sigma}_u^2$ y $\widetilde{\beta}$ are the moment estimators given by (1.45).

2. For $k = 1, 2, \dots,$ do

$$\begin{aligned}
\widehat{\beta}_k &= \left(\sum_{d=1}^{D} \frac{\mathbf{x}_d' \mathbf{x}_d}{\widehat{\sigma}_{u,k-1}^2 + \sigma_d^2} \right)^{-1} \left(\sum_{d=1}^{D} \frac{\mathbf{x}_d' y_d}{\widehat{\sigma}_{u,k-1}^2 + \sigma_d^2} \right), \\
\widehat{u}_{d,k} &= \frac{\widehat{\sigma}_{u,k-1}^2}{(\widehat{\sigma}_{u,k-1}^2 + \sigma_d^2)} (y_d - \mathbf{x}_d \widehat{\beta}_{k-1}), \quad \mathrm{tr}(\widehat{\mathbf{T}}_k) = \sum_{d=1}^{D} \frac{\widehat{\sigma}_{u,k}^2 \sigma_d^2}{\widehat{\sigma}_{u,k}^2 + \sigma_d^2}, \\
\widehat{\sigma}_{u,k}^2 &= \frac{\sum_{d=1}^{D} \widehat{u}_{d,k}^2}{D - \frac{1}{\widehat{\sigma}_{u,k-1}^2} \mathrm{tr}(T_{k-1}))}.
\end{aligned}$$

3. Stop when $|\widehat{\sigma}^2_{u,k} - \widehat{\sigma}^2_{u,k-1}| < \varepsilon$ and $\left[(\widehat{\beta}_k - \widehat{\beta}_{k-1})'(\widehat{\beta}_k - \widehat{\beta}_{k-1})\right]^{1/2} < \varepsilon$. Output: $\widehat{\beta}_{ML} = \widehat{\beta}_k$, $\widehat{u}_d = \widehat{u}_{d,k}$ y $\widehat{\sigma}^2_{u,ML} = \widehat{\sigma}^2_{u,k}$

1.7 The EBLUP and its mean squared error

1.7.1 Introducción

Let us consider model (1.1) with N in the place of n. Let s and r denote subsets of $\{1,\dots,N\}$ with sizes n and k respectively. Subset s contains the indexes of observed componentes of vector $\mathbf{y}$ and subset r is used to define a linear combination of fixed and random effects. Note that we do not assume that $n+k=N$ holds. let us define $\tau = \mathbf{a}_r'(\mathbf{X}_r\beta + \mathbf{Z}_r\mathbf{u})$, where $\mathbf{a}_r$ is a vector containing known constants. We are interested in predicting τ by using the EBLUP.

We consider 3 cases:

1. β, $\theta_0,\theta_1,\dots,\theta_m$ are known,
2. $\theta_0,\theta_1,\dots,\theta_m$ are known, β is unknown,
3. All the model parameters are unknown.

All the model parameters are known

Assume that β and $\theta_0,\theta_1,\dots,\theta_m$ are known. The BLUP of τ is

$$\widetilde{\tau} = \mathbf{a}_r'(\mathbf{X}_r\beta + \mathbf{Z}_r\widetilde{\mathbf{u}}), \quad \text{with} \quad \widetilde{\mathbf{u}} = \mathbf{C}_s'\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\beta)$$

where $\mathbf{C}_s = Cov(\mathbf{y}_s,\mathbf{u}) = \mathbf{Z}_s\mathbf{V}_u$. The prediction error is thus $\widetilde{\tau} - \tau = \mathbf{a}_r'\mathbf{Z}_r(\widetilde{\mathbf{u}} - \mathbf{u})$. The mean squared error is

$$MSE(\widetilde{\tau}) = E[(\widetilde{\tau} - \tau)^2] = V(\widetilde{\tau} - \tau) = \mathbf{a}_r'\mathbf{Z}_r Var(\widetilde{\mathbf{u}} - \mathbf{u})\mathbf{Z}_r'\mathbf{a}_r$$

It holds that

$$\begin{aligned} Var(\widetilde{\mathbf{u}} - \mathbf{u}) &= Var(\widetilde{\mathbf{u}}) + Var(\mathbf{u}) - 2Cov(\widetilde{\mathbf{u}},\mathbf{u}) = \mathbf{C}_s'\mathbf{V}_s^{-1}\mathbf{V}_s\mathbf{V}_s^{-1}\mathbf{C}_s + \mathbf{V}_u - 2\mathbf{C}_s'\mathbf{V}_s^{-1}\mathbf{C}_s \\ &= \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{V}_u. \end{aligned}$$

We know that $\mathbf{V}_s^{-1} = (\mathbf{V}_{es} + \mathbf{Z}_s\mathbf{V}_u\mathbf{Z}_s')^{-1}$. By using the inversion formula

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \tag{1.46}$$

we get

$$\mathbf{V}_s^{-1} = \mathbf{V}_{es}^{-1} - \mathbf{V}_{es}^{-1}\mathbf{Z}_s(\mathbf{V}_u^{-1} + \mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s)^{-1}\mathbf{Z}_s'\mathbf{V}_{es}^{-1}.$$

We can write $\mathbf{V}_s$ as a function of $\mathbf{T}_s = (\mathbf{V}_u^{-1} + \mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s)^{-1}$ in the following manner

$$\mathbf{V}_s^{-1} = \mathbf{V}_{es}^{-1} - \mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}.$$

Similarly, by applying (1.46) to $\mathbf{T}_s$ we get

$$\mathbf{T}_s = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}_s'(\mathbf{V}_{es} + \mathbf{Z}_s\mathbf{V}_u\mathbf{Z}_s')^{-1}\mathbf{Z}_s\mathbf{V}_u = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{V}_u$$

Therefore

$$Var(\widetilde{\mathbf{u}} - \mathbf{u}) = \mathbf{T}_s.$$

and

$$MSE(\widetilde{\tau}) = \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r \triangleq g_1(\theta).$$

The variance components are known but the regression parameters are unknown

In this case we assume that $\theta_0, \theta_1, \ldots, \theta_m$ are known, but β is unknown. Let us define $\mathbf{Q}_s = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}$ and $\mathbf{C}_s = Cov(\mathbf{y}_s, \mathbf{u}) = \mathbf{Z}_s\mathbf{V}_u$. The BLUP of τ is

$$\widehat{\tau}_{blup} = \mathbf{a}_r'(\mathbf{X}_r\widehat{\beta} + \mathbf{Z}_r\widehat{\mathbf{u}}),$$

where

$$\widehat{\mathbf{u}} = \mathbf{C}_s'\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta}) \quad \text{y} \quad \widehat{\beta} = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s = \mathbf{Q}_s\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s.$$

It holds that

$$\begin{aligned} MSE(\widehat{\tau}_{blup}) &= g_1(\theta) + g_2(\theta), \\ g_1(\theta) &= \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r, \\ g_2(\theta) &= [\mathbf{a}_r'\mathbf{X}_r - \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r - \mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r]. \end{aligned}$$

All the parameters are unknown

When the componentes of $\theta = (\theta_0, \theta_1, \ldots, \theta_m)$ are known, the BLUP of τ is $\widehat{\tau}_{blup} = \tau(\theta)$. If θ is unknown, then it is replaced by an estimator to obtain the EBLUP of τ, i.e.

$$\widehat{\tau}_{eblup} = \tau(\widehat{\theta}).$$

The mean squared error of $\widehat{\tau}_{eblup}$ is

$$\begin{aligned} MSE(\widehat{\tau}_{eblup}) &= E\left[(\widehat{\tau}_{eblup} - \widehat{\tau}_{blup} + \widehat{\tau}_{blup} - \tau)^2\right] \\ &= MSE(\widehat{\tau}_{blup}) + E\left[(\widehat{\tau}_{eblup} - \widehat{\tau}_{blup})^2\right] + 2E\left[(\widehat{\tau}_{eblup} - \widehat{\tau}_{blup})(\widehat{\tau}_{blup} - \tau)\right]. \end{aligned}$$

Kackar and Harville (1981) showed that if $E[\tau(\theta)]$ is finite and $\widehat{\theta}$ is an even and translation invariante (as the Henderson 3, Ml and REML estimators are), then $\widehat{\tau}_{eblup} = \tau(\widehat{\theta})$ is unbiased. Further, under these assumptions, Kackar and Harville (1984) proved that

$$E\left[(\widehat{\tau}_{eblup} - \widehat{\tau}_{blup})(\widehat{\tau}_{blup} - \tau)\right] = 0. \tag{1.47}$$

Here we assume that (1.47) holds, so that

$$MSE(\widehat{\tau}_{eblup}) = MSE(\widehat{\tau}_{blup}) + E\left[(\widehat{\tau}_{eblup} - \widehat{\tau}_{blup})^2\right]. \tag{1.48}$$

In what follows an approximation to

$$E\left[(\widehat{\tau}_{eblup} - \widehat{\tau}_{blup})^2\right].$$

is given. For this sake, consider an admisible value $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_m)$ of θ and define $\mathbf{d}(\theta) = (d_0(\theta), d_1(\theta), \ldots, d_m(\theta))'$, where

$$d_j(\theta) = \left.\frac{\partial \tau(\gamma)}{\partial \gamma_j}\right|_{\theta}, \quad j = 0, 1, \ldots, m.$$

A first order Taylor series expansion of $\tau(\gamma)$ around θ yields to

$$\tau(\gamma) \approx \tau(\theta) + \sum_{j=0}^{m} d_j(\theta)(\gamma_j - \theta_j).$$

By doing the substitution $\gamma = \widehat{\theta}$, we get

$$\widehat{\tau}_{eblup} \approx \widehat{\tau}_{blup} + \sum_{j=0}^{m} d_j(\theta)(\widehat{\theta}_j - \theta_j) = \widehat{\tau}_{blup} + \mathbf{d}'(\theta)(\widehat{\theta} - \theta).$$

Let us now assume that $\widehat{\theta}$ is asymptotically unbiased, i.e.

$$E\left[\widehat{\theta}_j - \theta_j\right] \underset{n\to\infty}{\longrightarrow} 0, \quad j = 0, 1, \ldots, m.$$

Then

$$E\left[(\widehat{\tau}_{eblup} - \widehat{\tau}_{blup})^2\right] \approx E\left[(\mathbf{d}'(\theta)(\widehat{\theta} - \theta))^2\right] = \sum_{i=0}^{m}\sum_{j=0}^{m} E\left[d_i(\theta)(\widehat{\theta}_i - \theta_i)d_j(\theta)(\widehat{\theta}_j - \theta_j)\right]. \tag{1.49}$$

Further, it holds

$$E\left[d_j(\theta)\right] = 0, \quad j = 0, 1, \ldots, m.$$

As $\mathbf{d}(\theta) = \mathbf{d}(\theta, \mathbf{u})$ is a random vector, the summand (i, j) in (1.49) is

$$E\left[d_i(\theta)d_j(\theta)(\widehat{\theta}_i - \theta_i)(\widehat{\theta}_j - \theta_j)\right] = E_{\widehat{\theta}}\left[(\widehat{\theta}_i - \theta_i)(\widehat{\theta}_j - \theta_j)E_{\mathbf{d}}\left[d_i(\theta)d_j(\theta)\,|\,\widehat{\theta}\right]\right].$$

Now we have

$$E_{\mathbf{d}}\left[d_i(\theta)d_j(\theta)\,|\,\widehat{\theta}\right] = Cov\left(d_i(\theta), d_j(\theta)\,|\,\widehat{\theta}\right).$$

In the case that $\widehat{\theta}$ is obtained from data independent of the data used to calculate $\widehat{\tau}_{blup} = \widehat{\tau}(\theta)$, we have that

$$Cov\left(d_i(\theta), d_j(\theta)\,|\,\widehat{\theta}\right) = Cov\left(d_i(\theta), d_j(\theta)\right)$$

and therefore

$$E\left[d_i(\theta)d_j(\theta)(\widehat{\theta}_i-\theta_i)(\widehat{\theta}_j-\theta_j)\right] = Cov(d_i(\theta),d_j(\theta))E\left[(\widehat{\theta}_i-\theta_i)(\widehat{\theta}_j-\theta_j)\right] = Cov(d_i(\theta),d_j(\theta))Cov(\widehat{\theta}_i,\widehat{\theta}_j)$$

The second summand in (1.48) can be written as

$$E\left[(\widehat{\tau}_{eblup}-\widehat{\tau}_{blup})^2\right] = \sum_{j=0}^{m}\sum_{i=0}^{m} Cov(d_i(\theta),d_j(\theta))Cov(\widehat{\theta}_i,\widehat{\theta}_j) = \operatorname{tr}\{\mathbf{G}(\theta)\mathbf{B}(\theta)\},$$

where $\mathbf{G}(\theta)$ and $\mathbf{B}(\theta)$ are the covariance matrices of $\mathbf{d}(\theta)$ and $\widehat{\theta}$ respectively.

In the case that $\widehat{\theta}$ and $\widehat{\tau}_{blup}=\widehat{\tau}(\theta)$ are calculated from the same data, Kackar and Harville (1984) propose the approximation

$$E\left[(\widehat{\tau}_{eblup}-\widehat{\tau}_{blup})^2\right] \approx \operatorname{tr}\{\mathbf{G}(\theta)\mathbf{B}(\theta)\}.$$

Therefore an approximation of the MSE of $\widehat{\tau}_{eblup}$ is

$$MSE(\widehat{\tau}_{eblup}) \approx MSE(\widehat{\tau}_{blup}) + \operatorname{tr}\{\mathbf{G}(\theta)\mathbf{B}(\theta)\}.$$

Prasad and Rao (1990) gave the new approximation

$$\operatorname{tr}\{\mathbf{G}(\theta)\mathbf{B}(\theta)\} \approx \operatorname{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}_s(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}, \tag{1.50}$$

where $\mathbf{b}'=(b_1,\ldots,b_n)=\mathbf{a}_r'\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}$,

$$\frac{\partial\mathbf{b}'}{\partial\theta_j} = \left(\frac{\partial b_1}{\partial\theta_j},\ldots,\frac{\partial b_n}{\partial\theta_j}\right) \quad \text{and} \quad \nabla\mathbf{b}' = \begin{pmatrix} \frac{\partial\mathbf{b}'}{\partial\theta_0} \\ \frac{\partial\mathbf{b}'}{\partial\theta_1} \\ \vdots \\ \frac{\partial\mathbf{b}'}{\partial\theta_m} \end{pmatrix} = \begin{pmatrix} \frac{\partial b_1}{\partial\theta_0} & \cdots & \frac{\partial b_n}{\partial\theta_0} \\ \frac{\partial b_1}{\partial\theta_1} & \cdots & \frac{\partial b_n}{\partial\theta_1} \\ \vdots & \cdots & \vdots \\ \frac{\partial b_1}{\partial\theta_m} & \cdots & \frac{\partial b_n}{\partial\theta_m} \end{pmatrix}_{(m+1)\times n}.$$

Finally, if the componentes of the vector of variances $\theta=(\theta_0,\theta_1,\ldots,\theta_m)$ are know, we have the approximation

$$\begin{aligned} MSE(\widehat{\tau}_{eblup}) &= g_1(\theta)+g_2(\theta)+g_3(\theta), \\ g_1(\theta) &= \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r. \\ g_2(\theta) &= [\mathbf{a}_r'\mathbf{X}_r-\mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r-\mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r], \\ g_3(\theta) &\approx \operatorname{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}_s(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}. \end{aligned}$$

1.7.2 Mean squared error estimation

A simple estimator of $MSE(\widehat{\tau})$ is obtained by plugging $\widehat{\theta}$ in the place θ to obtain

$$mse_1(\widehat{\tau}_{eblup}) = g_1(\widehat{\theta}) + g_2(\widehat{\theta}) + g_3(\widehat{\theta}). \tag{1.51}$$

If consistent estimators $\widehat{\theta}$ of θ are used, then $E[g_2(\widehat{\theta})] \cong g_2(\theta)$, $E[g_3(\widehat{\theta})] \cong g_3(\theta)$. However this property does not hold for for g_1.

To evaluate the bias of $g_1(\widehat{\theta})$, we expand $g_1(\widehat{\theta})$ in Taylor series aroundθ. We get

$$g_1(\widehat{\theta}) \approx g_1(\theta) + (\widehat{\theta}-\theta)'\nabla g_1(\theta) + \frac{1}{2}(\widehat{\theta}-\theta)'\nabla^2 g_1(\theta)(\widehat{\theta}-\theta) \triangleq g_1(\theta) + \Delta_1 + \Delta_2,$$

where $\nabla g_1(\theta)$ is the vector of first order derivatives of $g_1(\theta)$ with respect to θ and $\nabla^2 g_1(\theta)$ is the matrix of second order derivatives. If $\widehat{\theta}$ is unbiased for θ, then $E[\Delta_1] = 0$. In general, if the term $E[\Delta_1] \approx \mathbf{b}'_{\widehat{\theta}}(\theta)\nabla g_1(\theta)$ is of inferior order than $E[\Delta_2]$, where $\mathbf{b}_{\widehat{\theta}}(\theta)$ is an approximation to the bias $E[\widehat{\theta}-\theta]$, then the following approximation to $E[g_1(\widehat{\theta})]$ is obtained

$$E[g_1(\widehat{\theta})] \approx g_1(\theta) + \frac{1}{2}\operatorname{tr}\left(\nabla^2 g_1(\theta)\overline{\mathbf{V}}[\widehat{\theta}]\right), \tag{1.52}$$

where $\overline{\mathbf{V}}[\widehat{\theta}]$ is the asymptotic variance covariance matrix of $\widehat{\theta}$. Further, if $\mathbf{V}$ has a linear structure in θ, then (1.52) becomes

$$E[g_1(\widehat{\theta})] \approx g_1(\theta) - g_3(\theta). \tag{1.53}$$

From (1.51) and (1.53) we have that the bias of $mse_1(\widehat{\tau}_{eblup})$ is

$$E[mse_1(\widehat{\tau}_{eblup})] - MSE(\widehat{\tau}_{eblup}) \approx (g_1(\theta) - g_3(\theta) + g_2(\theta) + g_3(\theta)) - (g_1(\theta) + g_2(\theta) + g_3(\theta)) = -g_3(\theta).$$

Therefore $MSE(\widehat{\tau}_{eblup})$ can be estimated with

$$mse(\widehat{\tau}_{eblup}) = g_1(\widehat{\theta}) + g_2(\widehat{\theta}) + 2g_3(\widehat{\theta}). \tag{1.54}$$

Formula (1.54) is valid if $\widehat{\theta}$ is estimated by using the Henderson 3 or the REML method, which produces unbiased or quasi-unbiased estimators $\widehat{\theta}$ of θ. However for MLE estimators $\widehat{\theta}$ we have that $E[\Delta_1] \approx \mathbf{b}'_{\widehat{\theta}}(\theta)\nabla g_1(\theta) \neq 0$. In this case $MSE(\widehat{\tau}_{eblup})$ is estimated with

$$mse(\widehat{\tau}_{eblup}) = g_1(\widehat{\theta}) + g_2(\widehat{\theta}) + 2g_3(\widehat{\theta}) - \mathbf{b}'_{\widehat{\theta}}(\theta)\nabla g_1(\theta). \tag{1.55}$$

The term $\mathbf{b}_{\widehat{\theta}}(\theta)$ can be calculated more easily if $\mathbf{V}$ is a block diagonal matrix

$$\mathbf{V} = \operatorname{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_m)$$

with

$$\mathbf{V}_i = \mathbf{Z}_i\mathbf{V}_{ui}\mathbf{Z}'_i + \mathbf{V}_{ei}, \quad i = 1, \ldots, m.$$

In this case the components of model (1.1) can be written in the form $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_m')'$, $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_m')'$, $\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)'$, $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_m')'$, $\mathbf{e} = (\mathbf{e}_1', \ldots, \mathbf{e}_m')'$, where $\mathbf{X}_i$ es $n_i \times p$, $\mathbf{Z}_i$ es $n_i \times q_i$, $\mathbf{y}_i$ es $n_i \times 1$, $n = \sum_{i=1}^m n_i$ y $q = \sum_{i=1}^m q_i$. A model of this type can be decomposed in m submodels

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i, \quad i = 1, \ldots, m. \tag{1.56}$$

Under the model (1.56), if $\widehat{\theta}$ is the MLE of θ, an approximation to the bias is (see e.g. Rao (2003))

$$\mathbf{b}_{\widehat{\theta}}(\theta) = \frac{1}{2m}\left\{ I^{-1}(\theta) \operatorname*{col}_{1\le j\le m}\left[\mathrm{tr}\left[\sum_{i=1}^m (\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\left(\sum_{i=1}^m \mathbf{X}_i'\mathbf{V}_i^{(j)}\mathbf{X}_i\right)\right]\right]\right\},$$

where $\operatorname*{col}_{1\le j\le m}[a_j]$ is a column vector with elements a_j, $j = 1, \ldots, m$,

$$\mathbf{V}_i^{(j)} = \frac{\partial \mathbf{V}_i^{-1}}{\partial \theta_j} = -\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1} \quad \text{and} \quad I_{jk}(\theta) = \frac{1}{2}\sum_{i=1}^m tr\left[\left(\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\right)\left(\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\right)\right].$$

Prasad and Rao (1990) obtained the estimator of ECM given in (1.54) for moments estimators and special cases of the general linear mixed model with block diagonal covariance matrix. Harville and Jeske (1992) proposed (1.54) for a more general linear mixed model (1.1), under the hypothesis $E[\widehat{\theta} - \theta] = 0$. Das, Jiang and Rao (2001) gave rigorous proofs of approximations (1.54) and (1.55) for ML and REML estimators. Finally Lahiri and Rao (1995) have studied the robustness of the above cited approximations.

Chapter 2

EB prediction of non-linear domain parameters with unit level models

This chapter describes a methodology for obtaining empirical best predictors of general, possibly non-linear, domain parameters using unit level linear regression models. The proposed method is particularized to FGT poverty measures (Foster et al., 1984) as particular cases of non-linear parameters. The mean squared error of the proposed estimators is obtained by a parametric bootstrap for finite populations. This chapter is based on the results of Molina and Rao (2010). The chapter is organizes as follows. Section 2.1 describes the empirical best predictor of a non-linear population parameter. Section 2.2 is devoted to the estimation of domain parameters. This is done under normality and using a Monte Carlo approximation of the empirical best predictor. Section 2.3 introduces the nested-error model and gives a fast way for generating multivariate normal vectors for the domains. This method makes feasible the application of the proposed empirical best prediction method in real situations with large domains. Section 2.4 describes the parametric bootstrap for mean squared error estimation. Section 2.5 particularizes de proposed method to the estimation of domain FGT poverty measures. Section 2.6 describes the method of Elbers et al. (2003) for the estimation of domain parameters, and it discusses its properties when estimating domain means in comparison with the method proposed here. Sections 2.7 and 2.8 describe the results of model-based and design-based simulation experiments respectively, conducted to analyze and compare the performance of empirical best predictors, direct estimators and estimators obtained by the method of Elbers et al. (2003) for the FGT poverty measures.

2.1 Empirical best predictor under a finite population

Let $\mathbf{y}$ be a random vector containing the values of a random variable in the units of a finite population. Let $\mathbf{y}_s$ be the sub-vector of $\mathbf{y}$ corresponding to sample elements and $\mathbf{y}_r$ the sub-vector of out-of-sample elements and consider without loss of generality that the elements of $\mathbf{y}$ are sorted as $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$. Now consider a real measurable function $\delta = h(\mathbf{y})$ of the random vector $\mathbf{y}$. The target is to predict $\delta = h(\mathbf{y})$

using the sample data $\mathbf{y}_s$. Let $\hat{\delta}$ denote a predictor of δ. The mean squared error (MSE) of $\hat{\delta}$ is defined as

$$MSE(\hat{\delta}) = E_{\mathbf{y}}[(\hat{\delta}-\delta)^2], \tag{2.1}$$

where $E_{\mathbf{y}}$ denotes expectation with respect to the joint distribution of the population vector $\mathbf{y}$. The BP of δ is the function of $\mathbf{y}_s$ that minimizes (2.1). Consider the conditional expectation $\delta^0 = E_{\mathbf{y}_r}(\delta|\mathbf{y}_s)$, where the expectation is taken with respect to the joint distribution of $\mathbf{y}_r$ given $\mathbf{y}_s$ and the result is a function of sample data $\mathbf{y}_s$. Subtracting and adding δ^0 in the MSE, we obtain

$$\begin{aligned} MSE(\hat{\delta}) &= E_{\mathbf{y}}[(\hat{\delta}-\delta^0+\delta^0-\delta)^2] \\ &= E_{\mathbf{y}}[(\hat{\delta}-\delta^0)^2]+2E_{\mathbf{y}}[(\hat{\delta}-\delta^0)(\delta^0-\delta)]+E_{\mathbf{y}}[\delta^0-\delta)^2] \end{aligned}$$

In this expression, the last term does not depend on $\hat{\delta}$. For the second term, observe that

$$\begin{aligned} E_{\mathbf{y}}[(\hat{\delta}-\delta^0)(\delta^0-\delta)] &= E_{\mathbf{y}_s}\left\{E_{\mathbf{y}_r}\left[(\hat{\delta}-\delta^0)(\delta^0-\delta)|\mathbf{y}_s\right]\right\} \\ &= E_{\mathbf{y}_s}\left\{(\hat{\delta}-\delta^0)\left[\delta^0-E_{\mathbf{y}_r}(\delta|\mathbf{y}_s)\right]\right\} \\ &= 0. \end{aligned}$$

Thus, the BP of δ is the predictor $\hat{\delta}$ that minimizes $E_{\mathbf{y}}[(\hat{\delta}-\delta^0)^2]$. Since this quantity is non-negative and its minimum value is zero, the BP of δ is

$$\hat{\delta}^B = \delta^0 = E_{\mathbf{y}_r}(\delta|\mathbf{y}_s). \tag{2.2}$$

Note that the BP is unbiased in the sense that $E_{\mathbf{y}}(\hat{\delta}^B-\delta)=0$ because

$$E_{\mathbf{y}_s}(\hat{\delta}^B) = E_{\mathbf{y}_s}\{E_{\mathbf{y}_r}(\delta|\mathbf{y}_s)\} = E_{\mathbf{y}}(\delta).$$

Typically, $\mathbf{y}$ follows a distribution depending on an unknown parameter vector θ. This parameter is previously estimated using the sample data $\mathbf{y}_s$. Then, the empirical best predictor (EBP) of δ, denoted $\hat{\delta}^{EB}$, is equal to (2.2), with the expectation taken with respect to the distribution of $\mathbf{y}_r|\mathbf{y}_s$ with θ replaced by an estimator $\hat{\theta}$. The EBP is not exactly unbiased, but the bias coming from the estimation of the parameter θ is typically negligible.

Observation 2.1.1. *Assume that $\mathbf{y}=(\mathbf{y}_s',\mathbf{y}_r')'$ follows a Normal distribution with mean vector $\mu=\mathbf{X}\beta$, for a known matrix $\mathbf{X}$, with sample and out-of-sample decomposition $\mathbf{X}=(\mathbf{X}_s',\mathbf{X}_r')'$, and positive definite covariance matrix $\mathbf{V}$ decomposed accordingly as*

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{pmatrix}.$$

Assume also that the target parameter δ is a linear function of $\mathbf{y}$, that is, $\delta=\mathbf{a}'\mathbf{y}$, where $\mathbf{a}=(\mathbf{a}_s',\mathbf{a}_r')'$. Then, the BP of $\delta=\mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\mathbf{y}_r$ is given by

$$\hat{\delta}^B = \mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\left[\mathbf{X}_r\beta+\mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}(\mathbf{y}_s-\mathbf{X}_s\beta)\right]. \tag{2.3}$$

Replacing β by the weighted least squares estimator $\hat{\beta}=(\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s$ in in (2.3), we obtain the best linear unbiased predictor (BLUP) of $\delta=\mathbf{a}'\mathbf{y}$ as defined by Royall (1976).

2.2 Empirical best predictors of small domain non-linear parameters

The BP of a non-linear measurable function $\delta = h(\mathbf{y})$ can be obtained as soon as the population vector $\mathbf{y}$ follows a distribution such that the distribution of $\mathbf{y}_r|\mathbf{y}_s$ is known. Under this condition, the EB method allows the estimation of practically any characteristic of a finite population. Here we concentrate on the estimation of domain characteristics. For this, let $\mathbf{y}_d = (\mathbf{y}'_{ds}, \mathbf{y}'_{dr})'$ be the subvector of $\mathbf{y}$ for d-th domain and let $\delta_d = h(\mathbf{y}_d)$ be the target parameter, for a real measurable function h. Then the BP of δ is given by

$$\hat{\delta}_d^B = E_{\mathbf{y}_{dr}}(\delta_d|\mathbf{y}_{ds}). \tag{2.4}$$

When the domain vectors $\mathbf{y}_d$, $d = 1,\ldots,D$, are independent following a Normal distribution, the distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$ is also Normal and then the expectation in (2.4) can be easily derived. Thus, we consider that

$$\mathbf{y}_d \sim \text{ind } N(\mu_d, \mathbf{V}_d), \quad d = 1,\ldots,D, \tag{2.5}$$

where the mean vector μ_d and the covariance matrix $\mathbf{V}_d$ can be partitioned in submatrices corresponding to sample and out-of-sample elements

$$\mu_d = \begin{pmatrix} \mu_{ds} \\ \mu_{dr} \end{pmatrix}, \quad \mathbf{V}_d = \begin{pmatrix} \mathbf{V}_{ds} & \mathbf{V}_{dsr} \\ \mathbf{V}_{drs} & \mathbf{V}_{dr} \end{pmatrix}. \tag{2.6}$$

Then, the distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$ is

$$\mathbf{y}_{dr}|\mathbf{y}_{ds} \sim N(\mu_{dr|s}, \mathbf{V}_{dr|s}), \tag{2.7}$$

where

$$\mu_{dr|s} = \mu_{dr} - \mathbf{V}_{drs}\mathbf{V}_{ds}^{-1}(\mathbf{y}_{ds} - \mu_{ds}) \quad \text{and} \quad \mathbf{V}_{dr|s} = \mathbf{V}_{dr} - \mathbf{V}_{drs}\mathbf{V}_{ds}^{-1}\mathbf{V}_{dsr}.$$

For complex non-linear domain parameters $\delta_d = h(\mathbf{y}_d)$, the expectation in (2.19) cannot be calculated analytically, but an empirical Monte Carlo approximation is easy to obtain. For this, generate a large number L of vectors $\mathbf{y}_{dr}$ from (2.7). Let $\mathbf{y}_{dr}^{(\ell)}$ be the vector generated in the ℓ-th replication. Attach this vector to the sample vector $\mathbf{y}_{ds}$ to obtain the population vector for d-th domain, $\mathbf{y}_d^{(\ell)} = (\mathbf{y}'_{ds}, (\mathbf{y}_{dr}^{(\ell)})')'$. Let $\delta_d^{(\ell)} = h(\mathbf{y}_d^{(\ell)})$ be the target parameter for the corresponding domain obtained from $\mathbf{y}_d^{(\ell)}$. A Monte Carlo approximation to the BP of δ_d is simply the average of $\delta_d^{(\ell)} = h(\mathbf{y}_d^{(\ell)})$, $\ell = 1,\ldots,L$, that is,

$$\hat{\delta}_d^B = E_{\mathbf{y}_r}[h(\mathbf{y}_d)|\mathbf{y}_{ds}] \approx \frac{1}{L}\sum_{\ell=1}^{L} h(\mathbf{y}_d^{(\ell)}). \tag{2.8}$$

Typically, the mean vectors and covariance matrices in (2.5) involve an unknown parameter vector θ, that is, $\mu_d = \mu_d(\theta)$ and $\mathbf{V}_d = \mathbf{V}_d(\theta)$. An estimator $\hat{\theta}$ of θ is replaced in (2.7). Then the EBP of δ_d, denoted $\hat{\delta}_d^{EB}$, is obtained by generating out-of-sample vectors $\mathbf{y}_{dr}^{(\ell)}$ from the distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$, with θ replaced by $\hat{\theta}$, and applying (2.8).

2.3 Empirical best predictor under a nested error model

A possible model for the elements of the population vector $\mathbf{y}$ that can be used to evaluate the EBP is the nested error regression model, introduced by Battese, Harter and Fuller (1988). This model relates the population variables Y_{dj} (e.g., log-earnings) to a vector of p explanatory variables $\mathbf{x}_{dj}$ for all domains, and includes random domain-specific effects u_d along with the usual individual errors e_{dj}:

$$\begin{aligned} Y_{dj} &= \mathbf{x}_{dj}\beta + u_d + e_{dj}, \quad j = 1,\dots,N_d, \quad d = 1,\dots,D, \\ u_d &\sim \text{iid } N(0,\sigma_u^2), \quad e_{dj} \sim \text{iid } N(0,\sigma_e^2). \end{aligned} \tag{2.9}$$

where the domain effects u_d and the errors e_{dj} are independent. Let us define vectors and matrices obtained by stacking the elements for domain d

$$\mathbf{y}_d = \underset{1\le j\le N_d}{\text{col}}(Y_{dj}), \quad \mathbf{X}_d = \underset{1\le j\le N_d}{\text{col}}(\mathbf{x}_{dj}), \quad \mathbf{e}_d = \underset{1\le j\le N_d}{\text{col}}(e_{dj}).$$

Then, the domain vectors $\mathbf{y}_d$ are independent and follow the model

$$\mathbf{y}_d = \mathbf{X}_d\beta + u_d\mathbf{1}_{N_d} + \mathbf{e}_d, \quad \mathbf{e}_d \sim \text{ind } N(\mathbf{0},\sigma_e^2\mathbf{I}_{N_d}), \quad d = 1,\dots,D,$$

where u_d is independent of $\mathbf{e}_d$. Under this model, the mean vector and the covariance matrix of $\mathbf{y}_d$ are given by

$$\mu_d = \mathbf{X}_d\beta \quad \text{and} \quad \mathbf{V}_d = \sigma_u^2\mathbf{1}_{N_d}\mathbf{1}'_{N_d} + \sigma_e^2\mathbf{I}_N.$$

Consider the decomposition of $\mathbf{y}_d$ into sample and out-of-sample elements $\mathbf{y}_d = (\mathbf{y}'_{dr},\mathbf{y}'_{ds})'$, and the corresponding decomposition of $\mu_d = E(\mathbf{y}_d)$ and $\mathbf{V}_d = \text{Var}(\mathbf{y}_d)$ as in (2.6). The distribution of the out-of-sample vector $\mathbf{y}_{dr}$ given the sample data $\mathbf{y}_{ds}$ is given by (2.7) where, for this particular model, the conditional mean vector and covariance matrix are given by

$$\mu_{dr|s} = \mathbf{X}_{dr}\beta + \sigma_u^2\mathbf{1}_{N_d-n_d}\mathbf{1}'_{n_d}\mathbf{V}_{ds}^{-1}(\mathbf{y}_{ds} - \mathbf{X}_{ds}\beta), \tag{2.10}$$

$$\mathbf{V}_{dr|s} = \sigma_u^2(1-\gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}'_{N_d-n_d} + \sigma_e^2\mathbf{I}_{N_d-n_d}, \tag{2.11}$$

with $\gamma_d = \sigma_u^2(\sigma_u^2 + \sigma_e^2/n_d)^{-1}$. Observe that the application of the Monte Carlo approximation (2.8) involves simulation of D multivariate Normal vectors of sizes $N_d - n_d$, $d = 1,\dots,D$, from (2.7). Then this process has to be repeated L times, something computationally unfeasible. This can be avoided by noting that the conditional covariance matrix $\mathbf{V}_{dr|s}$, given by (2.7), corresponds to the covariance matrix of a vector $\mathbf{y}_{dr}$ generated by the model

$$\mathbf{y}_{dr} = \mu_{dr|s} + v_d\mathbf{1}_{N_d-n_d} + \varepsilon_{dr}, \tag{2.12}$$

with new random effects v_d and errors ε_{dr} that are independent and satisfy

$$v_d \sim N(0,\sigma_u^2(1-\gamma_d)) \quad \text{and} \quad \varepsilon_{dr} \sim N(\mathbf{0}_{N_d-n_d},\sigma_\varepsilon^2\mathbf{I}_{N_d-n_d}).$$

Using model (2.12), instead of generating a multivariate normal vector of size $N_d - n_d$, we need to generate only univariate normal variables $v_d \sim N(0,\sigma_u^2(1-\gamma_d))$ and $\varepsilon_{dj} \sim N(0,\sigma_\varepsilon^2)$ independently, for

$j \in r_d$, and then obtain the corresponding out-of-sample elements Y_{dj} from (2.12) using as means the corresponding elements of $\mu_{dr|s}$ given by (2.10). As mentioned before, in practice the model parameters $\theta = (\beta', \sigma_u^2, \sigma_e^2)'$ are replaced by suitable estimators $\hat{\theta} = (\hat{\beta}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)'$, and then the variables Y_{dj} are generated from (2.12) with θ replaced by $\hat{\theta}$.

2.4 Parametric bootstrap for MSE estimation

The MSE of the EB estimator $\hat{\delta}_d^{EB}$ with respect to the model is given by

$$\text{MSE}(\hat{\delta}_d^{EB}) = E\left[(\hat{\delta}_d^{EB} - \delta_d)^2\right], \tag{2.13}$$

Note that here the target parameter δ_d is a random variable, so the usual decomposition of the MSE in terms of squared bias and variance of $\hat{\delta}_d^{EB}$ does not hold. However, (2.13) can be decomposed as

$$\text{MSE}(\hat{\delta}_d^{EB}) = \left[E(\hat{\delta}_d^{EB} - \delta_d)\right]^2 + V(\hat{\delta}_d^{EB} - \delta_d). \tag{2.14}$$

Thus, the MSE is equal to the sum of the squared model bias and the variance of the prediction error. Since the model bias of the "best" estimator $\hat{\delta}_d^{B}$ is exactly zero, the squared bias of the "empirical best" estimator $\hat{\delta}_d^{EB}$ in (2.14) is typically very small relative to the variance of the prediction error $\hat{\delta}_d^{EB} - \delta_d$ when m is large. In this case, the MSE is dominated by the variance term in (2.14).

Analytical approximations to the MSE are difficult to derive in the case of complex parameters such as the FGT poverty measures. We therefore obtain a parametric bootstrap MSE estimator by following the bootstrap method for finite populations of González-Manteiga et al. (2008). This bootstrap method can be readily applied to other complex parameters. This parametric bootstrap method works as follows:

1. Fit model (2.9) to sample data $\mathbf{y}_s$ and obtain model parameter estimates $\hat{\beta}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$.

2. Generate bootstrap random domain effects as $u_d^* \sim \text{iid } N(0, \hat{\sigma}_u^2)$, $d = 1, \ldots, D$.

3. Generate, independently of the random effects u_d^*, bootstrap random errors $e_{dj}^* \sim \text{iid } N(0, \hat{\sigma}_e^2)$, $j = 1, \ldots, N_d$, $d = 1, \ldots, D$. ,

4. Construct a bootstrap population vector $\mathbf{y}^* = ((\mathbf{y}_1^*)', \ldots, (\mathbf{y}_D^*)')'$ using the estimated model,

$$Y_{dj}^* = \mathbf{x}_{dj}\hat{\beta} + u_d^* + e_{dj}^*, \quad j = 1, \ldots, N_d, \quad d = 1, \ldots, D, \tag{2.15}$$

   and calculate the true domain quantities for this bootstrap population, $\delta_d^* = h(\mathbf{y}_d^*)$, $d = 1, \ldots, D$.

5. Take the elements Y_{dj}^* of the population vector $\mathbf{y}^*$ with indices contained in the sample s, denoted $\mathbf{y}_s^*$. Fit model (2.9) again to bootstrap data $\mathbf{y}_s^*$, obtaining new model parameter estimates $\hat{\beta}^*$, $\hat{\sigma}_u^{2*}$ and $\hat{\sigma}_e^{2*}$.

6. Using the bootstrap sample data $\mathbf{y}_s^*$ and the known matrix $\mathbf{X}$, apply the EB method as described in Section 2.2 and calculate bootstrap EBPs, $\hat{\delta}_d^{EB*}$, $d = 1, \ldots, D$.

Observe that the bootstrap elements Y^*_{dj}, given the original sample data $\mathbf{y}_s$, preserve properties of the original population model. Let E_* and Var_* denote expectation and variance with respect to the distribution defined by the bootstrap model (2.15) given sample data $\mathbf{y}_s$. Then bootstrap random effects u^*_d and errors e^*_{dj} are iid with

$$E_*(u^*_d) = 0, \quad Var_*(u^*_d) = \hat{\sigma}^2_u, \quad E_*(e^*_{dj}) = 0, \quad Var_*(e^*_{dj}) = \hat{\sigma}^2_e. \tag{2.16}$$

Observe also that the mean vectors and covariance matrices of the bootstrap domain vectors $\mathbf{y}^*_d$ are given by

$$E_*(\mathbf{y}^*_d) = \mathbf{X}_d\hat{\beta} \quad \text{and} \quad Var_*(\mathbf{y}^*_d) = \hat{\sigma}^2_u \mathbf{1}_{N_d}\mathbf{1}'_{N_d} + \hat{\sigma}^2_e \mathbf{I}_N.$$

Thus, the distribution of the bootstrap population $\mathbf{y}^*$ (given the sample data $\mathbf{y}_s$) imitates that of the original population $\mathbf{y}$. Then an estimator of $MSE(\hat{\delta}^{EB}_d)$ is the bootstrap MSE of the bootstrap EBP, that is

$$MSE_*(\hat{\delta}^{EB*}_d) = E_*\left[(\hat{\delta}^{EB*}_d - \delta^*_d)^2\right].$$

In practice, this quantity is approximated through a Monte Carlo procedure. For this, repeat steps 2–6 a large number of times, B. Then we have generated B bootstrap populations with their corresponding true values of parameters and EBPs. An approximation for the bootstrap MSE is obtained then by averaging the squared errors over the B replicates. More specifically, let $\delta^{*(b)}_d$ and $\hat{\delta}^{EB*(b)}_d$ be the true domain parameter and its corresponding EBP for the bootstrap replicate b, for $b = 1,\dots,B$. Then the final bootstrap estimator of the MSE is

$$mse(\hat{\delta}^{EB}_d) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\delta}^{EB*(b)}_d - \delta^{*(b)}_d\right)^2. \tag{2.17}$$

It is possible to obtain a better MSE estimator, in terms of relative bias, by using a double bootstrap method (Hall and Maiti, 2006). However, under the finite population setup, in which full populations are generated in each bootstrap replication, the double bootstrap may be computationally infeasible.

2.5 Empirical best estimators of small domain FGT poverty measures

Consider the FGT family of poverty measures for domain d

$$F_{\alpha d} = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z - E_{dj}}{z}\right)^{\alpha} I(E_{dj} < z), \quad \alpha = 0,1,2, \tag{2.18}$$

where E_{dj} is the value of a quantitative welfare measure for j-th individual within d-th domain and z is the given poverty line. For $\alpha = 0$ we obtain the proportion of individuals under the poverty line, which is called poverty incidence. For $\alpha = 1$ we obtain the domain mean of relative distances to the poverty line, which is called poverty gap. While the poverty incidence accounts for the quantity of people under the poverty line, the poverty gap measures the degree of poverty of the people under the poverty line.

The BP of the FGT poverty measure $\delta_d = F_{\alpha d}$ is given by

$$\hat{F}^B_{\alpha d} = E_{\mathbf{y}_{dr}}(F_{\alpha d}|\mathbf{y}_{ds}).$$

Thus, in order to obtain the BP of $F_{\alpha d}$, we need to express $F_{\alpha d}$ in terms of a domain vector $\mathbf{y}_d$, for which the conditional distribution of the out-of-sample vector $\mathbf{y}_{dr}$ given sample data $\mathbf{y}_{ds}$ is known. The distribution of the welfare variables E_{dj} is seldom Normal due to the typical strong right-skewness of these kind of economical variables. However, many times it is possible to find a transformation of the E_{dj}'s whose distribution is approximately Normal. This transformation can be chosen from a suitable family such that the Box-Cox power family of transformations.

Thus, here we suppose that there exists a one-to-one transformation $Y_{dj} = T(E_{dj})$ of the welfare variables E_{dj}, which follows a Normal distribution. In particular, we will assume that the Y_{dj}'s follow the nested error model (2.9). Let $\mathbf{y}_d = (\mathbf{y}'_{ds}, \mathbf{y}'_{dr})'$ be the vector containing the values of the transformed variables Y_{dj} for the sample and out-of-sample units within domain d. Then $F_{\alpha d}$ is function of $\mathbf{y}_d$, that is

$$F_{\alpha d} = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z - T^{-1}(Y_{dj})}{z}\right)^{\alpha} I(T^{-1}(Y_{dj}) < z) =: h_\alpha(\mathbf{y}_d), \quad \alpha = 0,1,2.$$

Thus, the FGT poverty measure of order α is a non-linear function $h_\alpha(\mathbf{y}_d)$ of $\mathbf{y}_d$. Then the BP of $F_{\alpha d}$ is given by

$$\hat{F}^B_{dj} = E_{\mathbf{y}_{dr}}\left[h_\alpha(\mathbf{y}_d)|\mathbf{y}_{ds}\right] = \int_{I\!R} h_\alpha(\mathbf{y}_d) f(\mathbf{y}_{dr}|\mathbf{y}_{ds})\, d\mathbf{y}_{dr}, \tag{2.19}$$

where $f(\mathbf{y}_{dr}|\mathbf{y}_{ds})$ is the joint density of $\mathbf{y}_{dr}$ given the observed data vector $\mathbf{y}_{ds}$ obtained from (2.7). Due to the complexity of the function $h_\alpha(\cdot)$, there is not explicit expression for the expectation in (2.19), but this expectation can be approximated by Monte Carlo as explained in Section 2.2. Then, an approximation to the best predictor of $F_{\alpha d}$ is

$$\hat{F}^B_{\alpha d} \approx \frac{1}{L}\sum_{\ell=1}^{L} h_\alpha(\mathbf{y}^{(\ell)}_d).$$

Typically, the mean vector μ_d and the covariance matrix $\mathbf{V}_d$ depend on an unknown vector of parameters θ. Then the conditional density $f(\mathbf{y}_{dr}|\mathbf{y}_{ds})$ depends on θ, and we make this explicit by writing $f(\mathbf{y}_{dr}|\mathbf{y}_{ds},\theta)$. We take an estimator $\hat{\theta}$ of θ such as the maximum likelihood (ML) or restricted ML (REML)estimator. Then the expectation can be approximated by generating values from the estimated density $f(\mathbf{y}_{dr}|\mathbf{y}_{ds},\hat{\theta})$. The result is the EBP, denoted $\hat{F}^{EB}_{\alpha d}$.

2.6 ELL estimators of small domain non-linear parameters

The method of Elbers et al. (2003), called ELL or World Bank (WB) method, assumes a nested error model on the transformed population values, Y_{dj}, similar to (2.9) but using random cluster effects, where the clusters may be different from the small areas. In fact, the small areas are not specified in advance. They compute estimators of domain parameters δ_d by applying a method similar to the bootstrap procedure described in Section 2.4. More concretely, the ELL method follows the steps below:

1. With the original sample data $\mathbf{y}_s$, fit a linear model with cluster random effects,

$$Y_{dj} = \mathbf{x}_{dj}\beta + u_c + e_{dj}, \quad j = 1,\ldots,N_d, \quad d = 1,\ldots,D, \quad c = 1,\ldots,C,$$
$$u_c \sim \text{iid } N(0,\sigma_c^2), \quad e_{dj} \sim \text{iid } N(0,\sigma_e^2). \tag{2.20}$$

   where u_c is the random effect of cluster c. Let $\hat{\beta}$, $\hat{\sigma}_c^2$ and $\hat{\sigma}_e^2$ be the estimators of β, σ_c^2 and σ_e^2 in this model.

2. Generate bootstrap cluster effects $u_c^* \sim \text{iid } N(0,\hat{\sigma}_c^2)$, $c = 1,\ldots,C$.

3. Independently of the cluster effects, generate bootstrap model errors $e_{dj}^* \sim \text{iid } N(0,\hat{\sigma}_e^2)$, $j = 1,\ldots,N_d$, $d = 1,\ldots,D$.

4. Construct a population vector $\mathbf{y}^*$ from the bootstrap model

$$Y_{dj}^* = \mathbf{x}_{dj}\beta + u_c^* + e_{dj}^*, \quad j = 1,\ldots,N_d, \quad d = 1,\ldots,D, \quad c = 1,\ldots,C. \tag{2.21}$$

5. Calculate the true bootstrap domain parameters $\delta_d^* = h(\mathbf{y}_d^*)$, $d = 1,\ldots,D$.

6. The ELL estimator of δ_d is then the bootstrap mean

$$\hat{\delta}_d^{ELL} = E_*(\delta_d^*),$$

   and the bootstrap variance is used as an estimator of the MSE of the ELL estimator $\hat{\delta}_d^{ELL}$, that is, the ELL method uses

$$mse(\hat{\delta}_d^{ELL}) = Var_*(\delta_d^*) = E_*[\delta_d^* - E_*(\delta_d^*)]^2,$$

Note that $E_*(\delta_d^*)$ is tracking $E(\delta_d)$ and $Var_*(\delta_d^*)$ is tracking $V(\delta_d) = E[\delta_d - E(\delta_d)]^2$. In practice, ELL estimators are obtained from a Monte Carlo approximation by generating a large number, A, of population vectors $\mathbf{y}^{*(a)} = ((\mathbf{y}_1^{*(a)})',\ldots,(\mathbf{y}_D^{*(a)})')'$, $a = 1,\ldots,A$, calculating the bootstrap domain parameters for each population a in the form $\delta_d^{*(a)} = h(\mathbf{y}_d^{*(a)})$, $d = 1,\ldots,D$, and later averaging over the A populations; that is, taking

$$\hat{\delta}_d^{ELL} \approx \frac{1}{A}\sum_{a=1}^{A}\delta_d^{*(a)} =: \delta_d^{*(\cdot)} \quad \text{and} \quad mse(\hat{\delta}_d^{ELL}) \approx \frac{1}{A}\sum_{a=1}^{A}\left(\delta_d^{*(a)} - \delta_d^{*(\cdot)}\right)^2.$$

Note that ELL population vectors $\mathbf{y}^{*(a)}$ do not contain the observed sample data in contrast to the EB method described in Section 2.2.

To illustrate the ELL method and compare it with the EB method, consider the special case of estimating the domain means, that is, $\delta_d = \overline{Y}_d$, where

$$\overline{Y}_d = N_d^{-1}\sum_{d=1}^{N_d} Y_{dj}, \quad d = 1,\ldots,D.$$

The ELL estimator of the domain mean $\bar{Y}_d$ is the bootstrap mean

$$\hat{\bar{Y}}_d^{ELL} = E_*(\bar{Y}_d^*), \tag{2.22}$$

and the ELL estimator of the MSE of $\hat{\bar{Y}}_d^{ELL}$ is the bootstrap variance

$$mse(\hat{\bar{Y}}_d^{ELL}) = Var_*(\bar{Y}_d^*).$$

In many cases, as in some establishment surveys, there are no clusters. Then, the ELL method fits the linear model

$$Y_{dj} = \mathbf{x}_{dj}\beta + e_{dj}, \quad e_{dj} \sim \text{iid } N(0,\sigma_e^2), \quad j = 1,\dots,N_d, \quad d = 1,\dots,D, \tag{2.23}$$

and uses this model to construct bootstrap populations. Let us consider, for simplicity of exposition, that all the parameters involved in the model are known. The bootstrap mean for d-th domain is given by

$$\bar{Y}_d^* = N_d^{-1}\sum_{j=1}^{N_d} Y_{dj}^* = \frac{1}{N_d}\sum_{j=1}^{N_d}(\mathbf{x}_{dj}\beta + e_{dj}^*) = \hat{\bar{Y}}_d^{SYN} + \bar{E}_d^*,$$

where $\bar{E}_d^* = N_d^{-1}\sum_{j=1}^{N_d} e_{dj}^*$ and $\hat{\bar{Y}}_d^{SYN}$ is used to denote the synthetic estimator $\bar{\mathbf{X}}_d\beta$, where $\bar{\mathbf{X}}_d = N_d^{-1}\sum_{j=1}^{N_d}\mathbf{x}_{dj}$. The synthetic estimator is obtained by predicting all population elements Y_{dj} through the linear model (2.23) by $\hat{Y}_{dj} = \mathbf{x}_{dj}\beta$ and then taking the mean over the d-th domain, that is,

$$\hat{\bar{Y}}_d^{SYN} = \frac{1}{N_d}\sum_{j=1}^{N_d}\hat{Y}_{dj}.$$

By (2.22), the ELL estimator is given by

$$\hat{\bar{Y}}_d^{ELL} = E_*(\bar{Y}_d^*) = E_*(\hat{\bar{Y}}_d^{SYN} + \bar{E}_d^*) = \hat{\bar{Y}}_d^{SYN} + E_*(\bar{E}_d^*) = \hat{\bar{Y}}_d^{SYN},$$

due to property (2.16) of the bootstrap method. On the other hand, the EB estimator of $\bar{Y}_d$ under the linear model (2.23) is obtained by predicting only the out-of-sample observations and keeping the sample data, that is,

$$\hat{\bar{Y}}_d^{EB} = \frac{1}{N_d}\left\{\sum_{j\in s_d} Y_{dj} + \sum_{j\in r_d}\hat{Y}_{dj}\right\}.$$

Let us compare the MSEs of ELL and EB estimators. Taking the average of (2.23) over the elements in d-th domain, we can express the true mean as

$$\bar{Y}_d = \bar{\mathbf{X}}_d\beta + \bar{E}_d,$$

where $\bar{\mathbf{X}}_d = N_d^{-1}\sum_{j=1}^{N_d}\mathbf{x}_{dj}$ and $\bar{E}_d = N_d^{-1}\sum_{j=1}^{N_d} e_{dj}$. Now let us express the ELL estimator as $\hat{\bar{Y}}_d^{ELL} = \bar{\mathbf{X}}_d\beta$. Then, it holds that

$$\hat{\bar{Y}}_d^{ELL} - \bar{Y}_d = \bar{\mathbf{X}}_d\beta - \left(\bar{\mathbf{X}}_d\beta + \bar{E}_d\right) = \bar{E}_d,$$

and then the MSE of ELL estimator is

$$MSE(\hat{\bar{Y}}_d^{ELL}) = E\{(\hat{\bar{Y}}_d^{ELL} - \bar{Y}_d)^2\} = E(\bar{E}_d^2) = \frac{Var(e_{dj})}{N_d} = \frac{\sigma_e^2}{N_d}.$$

On the other hand, for the MSE of $\hat{\bar{Y}}_d^{EB}$, observe that the difference between the EB estimator and the true mean is equal to

$$\hat{\bar{Y}}_d^{EB} - \bar{Y}_d = \frac{1}{N_d}\sum_{j\in r_d} e_{dj},$$

which implies that the MSE of $\hat{\bar{Y}}_d^{EB}$ is given by

$$MSE(\hat{\bar{Y}}_d^{EB}) = E[(\hat{\bar{Y}}_d^{EB} - \bar{Y}_d)^2] = \frac{\sigma_e^2}{N_d}\left(1 - \frac{n_d}{N_d}\right) < \frac{\sigma_e^2}{N_d} = MSE(\hat{\bar{Y}}_d^{ELL}).$$

Thus, under model (2.23) with known model parameters, if $n_d \geq 1$, the EB estimator has always smaller MSE than the ELL estimator due to the more efficient use of the available information, namely the sample data. When the sampling fraction n_d/N_d is negligible, both estimators have a similar MSE.

Moreover, the ELL estimator of the MSE is

$$mse(\hat{\bar{Y}}_d^{ELL}) = E_*[(\bar{Y}_d^* - E_*(\bar{Y}_d^*))^2] = E_*[(\bar{E}_d^*)^2] = \frac{Var_*(e_{dj}^*)}{N_d} = \frac{\sigma_e^2}{N_d}, \tag{2.24}$$

which is the true MSE of the ELL estimator under model (2.23). Thus, when fitting a linear model without cluster effects, the ELL estimator of a small area mean is essentially the synthetic estimator, which is a good estimator when there are not domain effects and the true model is (2.23). In this case, the ELL estimator of the MSE tracks the true MSE.

However, many times there is extra domain variation that is not fully explained by the auxiliary variables; that is, the true model is (2.9). However, when there are no clusters, the ELL method fits model (2.23). In this case, the true mean for d-th domain is given by

$$\bar{Y}_d = \bar{\mathbf{X}}_d\beta + u_d + \bar{E}_d.$$

This means that the MSE of the ELL estimator under the true model is

$$MSE(\hat{\bar{Y}}_d^{ELL}) = E[(u_d + \bar{E}_d)^2] = \sigma_u^2 + \frac{\sigma_e^2}{N_d}. \tag{2.25}$$

Summarizing, when the true model is (2.9), the ELL estimator, equal to the synthetic estimator, is not accounting for the domain effects, and the ELL estimator of the MSE has a bias equal to σ_u^2, compare (2.24) with (2.25). Thus, this MSE estimator can lead to serious underestimation when the domain effects have a substantial variance σ_u^2.

Now, if we take the clusters in the ELL method equal to the small domains, then due to (2.16), the ELL estimator under the correct model is again the synthetic estimator, that is,

$$\hat{\bar{Y}}_d^{ELL} = E_*(\hat{\bar{Y}}_d^{SYN} + u_d^* + \bar{E}_d^*) = \hat{\bar{Y}}_d^{SYN}.$$

Moreover, the ELL estimator of the MSE is

$$mse(\hat{\bar{Y}}_d^{ELL}) = Var_*(\bar{Y}_d^*) = E_*[(\bar{Y}_d^* - E_*(\bar{Y}_d^*))^2] = E_*[(u_d^* + \bar{E}_d^*)^2] = \sigma_u^2 + \frac{\sigma_e^2}{N_d},$$

which is equal to the true MSE given in (2.25). This indicates that when the clusters are equal to the small areas, the ELL estimator remains essentially equal to a synthetic estimator, but in this case the ELL variance estimator is unbiased. Actually, when the true model is the nested-error model (2.9), the difference between ELL and EB methods is that the target quantities are not the same. The EB method tries to estimate (or better predict) the actual domain means $\bar{Y}_d$, while the ELL method is estimating instead the marginal expectations $E(\bar{Y}_d)$ along with the marginal variances $Var(\bar{Y}_d)$.

2.7 Model-based simulation experiment

We consider in this section a simulation study to check the EBP model in terms of measures (2.18) with a poverty incidence and a poverty gap ($\alpha = 0$ and $\alpha = 1$ respectively). We simulated populations of size $N = 20000$, composed of $D = 80$ areas with $N_d = 250$ elements in each area $d = 1, \dots, D$. The response variables Y_{dj} we generated from (2.9) using two binary (auxiliary) variables X_1 and X_2 plus an intercept. The binary variables were simulated from Bernoulli distributions with

$$\begin{aligned} p_{1d} &= 0.3 + 0.5d/80; \\ p_{2d} &= 0.2, \end{aligned}$$

$d = 1, \dots, D$, namely where p_{1d} is directly proportional to the area index for X_1 and p_{2d} is constant. We consider sample indices s_d with $n_d = 50$ drawn independently in each area d by simple random sampling without replacement. Variables X_1 and X_2 for population units and sample indices were the same for all Monte Carlo simulations.

The transformation $T(\cdot)$ defined in Section 2.5 is $T(x) = \log(x)$, in this way, the welfare variables E_{dj} are the exponential of the model responses Y_{dj}

The intercept and the regression coefficients of X_1 and X_2 were $\beta = (3, 0.03, -0.04)'$. By using this values , In this way, the mean welfare E_{dj} is larger from case $(X_1 = 0, X_2 = 0)$ to $(X_1 = 1, X_2 = 0)$, but decreases from $(X_1 = 0, X_2 = 0)$ to $(X_1 = 0, X_2 = 1)$. It can be interpreted as the higher income level is reached when $X_1 = 1$ and $X_2 = 0$.

Note that p_{1d} of $X_1 = 1$ increases with the area index but p_{2d} of $X_2 = 1$ is constant, then the last areas will have more individuals with larger Y_{dj} and then the FGT poverty measures will decrease with the area index.

We fixed the following values in the model:

(i) Random area effects variance, $\sigma_u^2 = (0.15)^2$.

(ii) Error variance, $\sigma_e^2 = (0.5)^2$.

(iii) The poverty line, $z = 12$ (roughly 0.6 times the median of the welfare variables E_{dj} for a population generated as mentioned above). Hence, the poverty incidence for the simulated populations is approximately 16%.

Therefore, we generated $I = 10000$ population vectors $\mathbf{y}^{(i)}$ from the true model and for each population i, we considered these steps:

(a) The FGT measures for $\alpha = 0$ and $\alpha = 1$ (true area poverty incidences and gaps) for each area $d = 1,\ldots,D$ and each population i are obtained as

$$F_{\alpha d}^{(i)} = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z-E_{dj}^{(i)}}{z}\right)^{\alpha} I(E_{dj}^{(i)} < z), \quad E_{dj}^{(i)} = \exp(Y_{dj}^{(i)}).$$

(b) Using the sample part of the i-th population vector, $\mathbf{y}_s^{(i)}$, direct estimators of $F_{\alpha d}^{(i)}$ were calculated as

$$\hat{F}_{\alpha d}^{(i)} = \frac{1}{n_d}\sum_{j\in s_d}\left(\frac{z-E_{dj}^{(i)}}{z}\right)^{\alpha} I(E_{dj}^{(i)} < z).$$

(c) The nested-error model given in (2.9) was fitted to sample data $(\mathbf{y}_s^{(i)}, \mathbf{X}_s)$. Then, substituting the estimated model parameters in (2.10) and (2.11), $L = 50$ out-of-sample vectors $\mathbf{y}_r^{(i\ell)}$, $\ell = 1,\ldots,L$ were generated from the conditional distribution (2.7) using (2.12) for $d = 1,\ldots,D$. The sample data $\mathbf{y}_s^{(i)}$ was attached to the generated out-of-sample data $\mathbf{y}_r^{(i\ell)}$ to form a population vector $\mathbf{y}^{(i\ell)}$. The domain poverty measures for $\alpha = 0,1$ and $d = 1,\ldots,D$ were obtained for each population vector $\mathbf{y}^{(i\ell)}$ as

$$F_{\alpha d}^{(i\ell)} = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z-E_{dj}^{(i\ell)}}{z}\right)^{\alpha} I(E_{dj}^{(i\ell)} < z), \quad E_{dj}^{(i\ell)} = \exp(Y_{dj}^{(i\ell)}), \quad d = 1,\ldots,D.$$

Then the Monte Carlo approximations to the EBPs of poverty measures for $\alpha = 0,1$ and $d = 1,\ldots,D$ were calculated as

$$\hat{F}_{\alpha d}^{EB(i)} = \frac{1}{L}\sum_{\ell=1}^{L} F_{\alpha d}^{(i\ell)}.$$

(d) Finally, we compute the ELL estimators of the poverty measures. Therefore, we applied the model (2.9) with the sample data $\mathbf{y}_s$ and generate $A = 50$ populations by means of a parametric bootstrap (see Section 2.4). In each population the poverty measures were computed and then averaged over the $A = 50$ populations, in order to find the ELL estimators $\hat{F}_{\alpha d}^{ELL(i)}$ for each i (see Section 2.6).

Observation 2.7.1. *Note that we used $L = A = 50$ for the EB and ELL methods in the simulation studies. A limited comparison of EB estimators for $L = 50$ with the corresponding values for $L = 1000$ showed that the choice $L = 50$ gives fairly accurate results. In practice, however, when dealing with a given sample data set, it is advisable to use larger values of L such as $L \geq 200$.*

Means over Monte Carlo populations $i = 1,\dots,I$ of the true values of the FGT measures of order $\alpha = 0,1$ were computed as

$$E(F_{\alpha d}) = \frac{1}{I}\sum_{i=1}^{I} F_{\alpha d}^{(i)}, \quad d = 1,\dots,D.$$

Similarly, biases $E(\hat{F}_{\alpha d}^{EB}) - E(F_{\alpha d})$, $E(\hat{F}_{\alpha d}) - E(F_{\alpha d})$ and $E(\hat{F}_{\alpha d}^{ELL}) - E(F_{\alpha d})$, and MSEs over Monte Carlo populations $E(\hat{F}_{\alpha d}^{EB} - F_{\alpha d})^2$, $E(\hat{F}_{\alpha d} - F_{\alpha d})^2$ and $E(\hat{F}_{\alpha d}^{ELL} - F_{\alpha d})^2$ of the three estimators were computed.

Figures 2.1 a) and b) report respectively the biases and the MSEs of the estimators for the poverty gap ($\alpha = 1$). Figure 2.1 a) shows that the EB estimator has the smallest absolute bias followed by ELL estimator, but compared to the corresponding values of MSE (Figure 2.1 b)), the square of the model bias is negligible for all the three estimators. Hence, the MSE of these estimators is dominated by the model variance of the prediction error, as explained at the beginning of Section 2.4. It is clear from Figure 2.1 b) that the EB estimator is significantly more efficient than ELL and direct estimators. Surprisingly, Figure 2.1 b) also reveals that, in these simulations, the ELL estimator is less efficient than the direct estimator, showing that the prediction error variance is larger for the ELL method. Results for the poverty incidence ($\alpha = 0$) were similar and are not reported here.

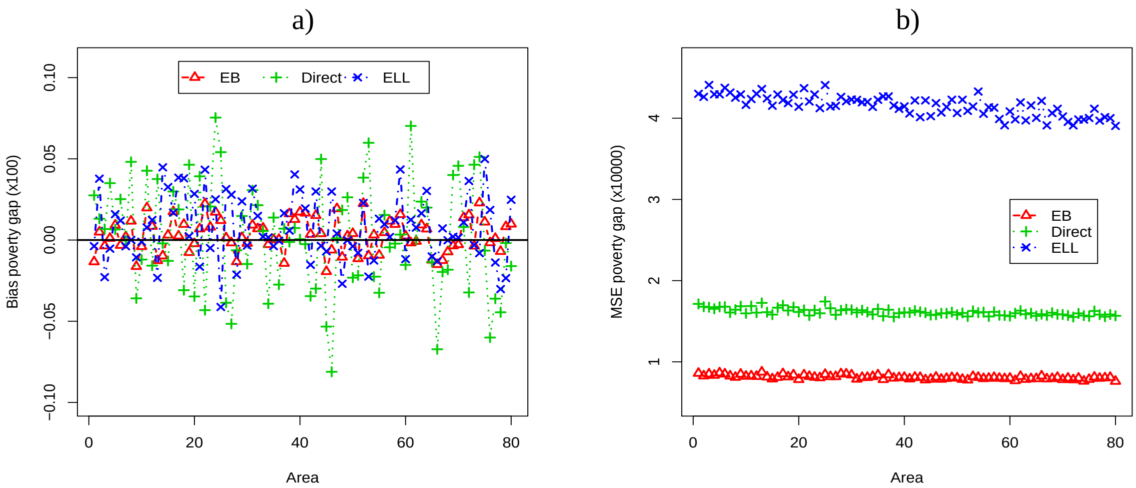


Figure 2.1: a) Bias ($\times 100$) and b) MSE ($\times 10^4$) over simulated populations of EB, direct and ELL estimators of the poverty gap F_{1d} for each area d.

Turning to MSE estimation, the parametric bootstrap procedure described in Section 2.4 was implemented with $B = 500$ replicates and the results are plotted in Figure 2.2 for the poverty gap ($\alpha = 1$). The number of Monte Carlo simulations was $I = 500$ and the true values of the MSE were independently computed with $I = 50000$ Monte Carlo simulations. Figure 2.4 shows that the bootstrap MSE estimator tracks the pattern of the true MSE values. Similar results were obtained for the poverty incidence ($\alpha = 0$).

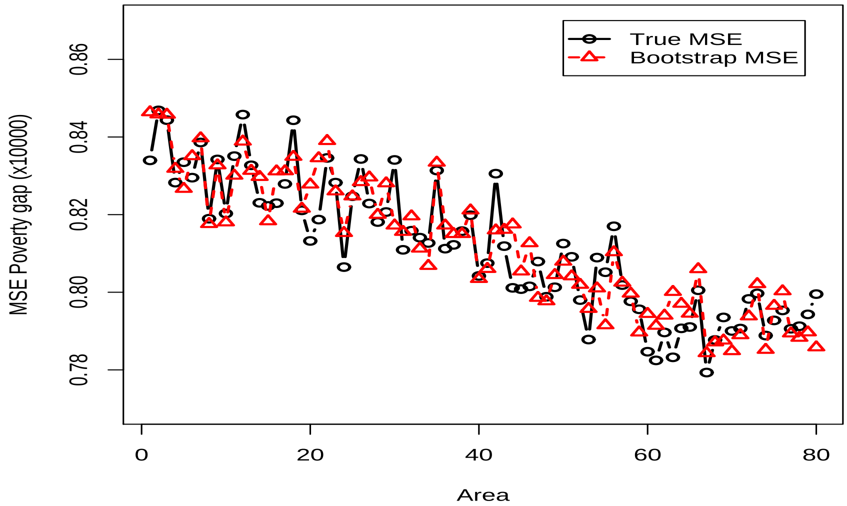


Figure 2.2: True MSE ($\times 10^4$) of EB predictor of poverty gap ($\alpha = 1$) and bootstrap MSE estimate with $B = 500$ for each area d.

2.8 Design-based simulation experiment

Now, we deal with the performance of the estimators when obtaining repeated samples drawn from a given population. In this way, we generate a population with the same parameters as showed in Section 2.7, and draw $I = 1000$ replicates. In each replicate another sample is taken based on a simple random procedure without replacement within each area. We computed in each sample, the three types of estimators of poverty measures: EBP, direct and ELL.

In Figures 2.3 a) and b) the design bias and design MSE of the estimators for poverty gap $(\alpha = 1)$ are shown. As expected in Figure 2.3 it represented the almost zero value of the Monte Carlo design bias of the direct estimator and a greater value of the EB estimator.

In Figure 2.3 b) it is shown that ELL estimators have small MSEs for some of the areas and large for the other areas, while the MSE of EB and direct estimators are small for all areas. Morteover, for most areas, the MSE of the EB estimator is smaller than the corresponding one of the direct estimator.

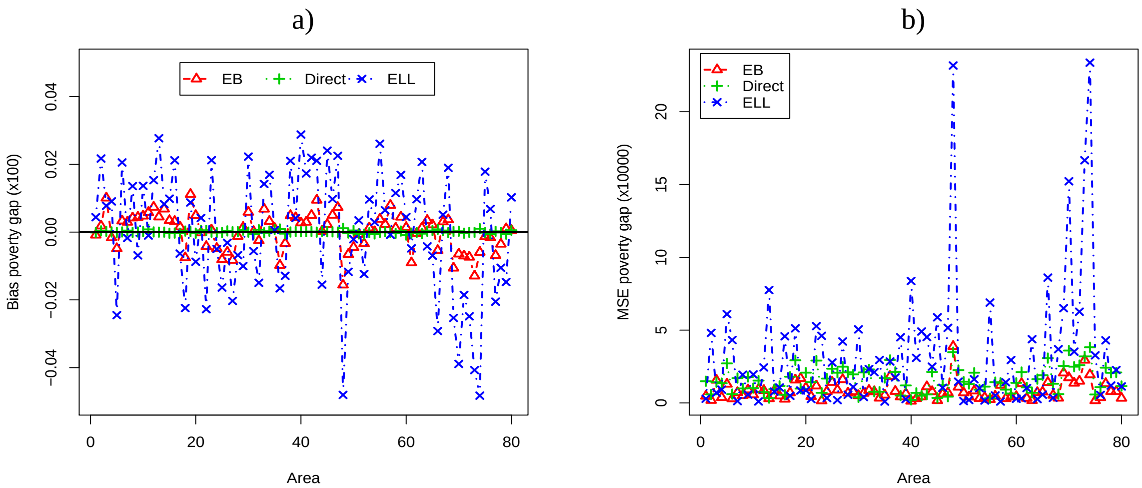


Figure 2.3: a) Bias ($\times 100$) and b) MSE ($\times 10^4$) of EB, direct and ELL estimators of the poverty gap F_{1d} for each area d under the design-based setup.

Chapter 3

Fast EB method for estimation of *fuzzy* poverty measures

3.1 Introduction

The traditional poverty measures for a population are obtained doing a simple dichotomization of the population into poor and non poor. This is done in relation to some chosen poverty line that represents a percentage (generally 50%, 60% or 70%) of the median of the equivalised income distribution, see e.g. Foster et al. (1984). On the one hand, this poverty line is arbitrary, and on the other, a more appropriate measure based on some degree of poverty, would be desirable. Moreover, this approach is unidimensional, that is, it refers to only one proxy of poverty, namely low income or consumption expenditure. Here we consider the estimation of fuzzy monetary and fuzzy supplementary indicators. The former are based on assigning a degree of poverty to the individuals using a ranking of the welfare variable used as proxy. The latter does a ranking on a score variable that is obtained after applying a multidimensional approach that takes into account a variety of non-monetary indicators of deprivation.

The EB method proposed by Molina and Rao (2009) for estimating small domain non-linear poverty indicators requires generation of full populations. For very large populations or for complex indicators, like those whose computation require sorting the data, the EB method might be unfeasible. Here we propose a modification of the EB method, called fast EB method, which reduces drastically the computing time, making feasible the estimation of complex non-linear quantities under large populations, whereas loosing little efficiency.

In simulations we compare the results of different small area estimation methods, including the original and the fast EB method, of complex domain poverty indicators using a unit level linear regression model. The indicators considered are the head count ratio (HCR) also called poverty incidence, the fuzzy monetary (FM) indicator and the fuzzy supplementary (FS) index. Moreover, the proposed approach is applied to the estimation of HCR, FM and FS indexes in Tuscany provinces.

3.2 Fuzzy monetary and supplementary indicators

Let $U = \{1,\ldots,N\}$ be a finite population of size N, where E_i is the value of a welfare variable (e.g. equivalised income) for individual i. Let us consider the empirical distribution function of $\{E_1,\ldots,E_N\}$, defined as

$$F_E(x) = \frac{1}{N}\sum_{j=1}^{N} I\{E_j \leq x\}, \quad x \in \mathbf{R},$$

where $I\left\{E_j \leq x\right\} = 1$ if $E_j \leq x$ and 0 otherwise. Consider also the (empirical) Lorenz curve, given by

$$L_E(x) = \frac{\sum_{j=1}^{N} E_j I\left\{E_j \leq x\right\}}{\sum_{j=1}^{N} E_j}, \quad x \in \mathbf{R}.$$

Following the Integrated Fuzzy and Relative (IFR) approach of Betti et al. (2006), the Fuzzy Monetary Index (FMI) for individual i is defined as

$$\begin{aligned} FM_i &= \left\{\frac{N}{N-1}(1-F_E(E_i))\right\}^{\alpha-1}\{1-L_E(E_i)\} \\ &= \left\{\frac{1}{N-1}\sum_{j=1}^{N} I\left\{E_j > E_i\right\}\right\}^{\alpha-1}\left\{\frac{\sum_{j=1}^{N} E_j I\left\{E_j > E_i\right\}}{\sum_{j=1}^{N} E_j}\right\}, \quad i \in U. \end{aligned}$$

Here, $1-F_E(E_i)$ is the proportion of individuals that are less poor than individual i. This gives a degree of poverty of individual i and it was proposed by Cheli e Lemmi (1995) as a poverty indicator. Observe that $N(1-F_E(E_i))/(N-1)$ is equal to 1 when individual i is the poorest. Moreover, $1-L_E(E_i)$ is the share of the total welfare of all individuals that are less poor than this individual, indicator that was proposed by Betti and Verma (1999). The average FMI for the population is given by

$$FM = \frac{1}{N}\sum_{i=1}^{N} FM_i \tag{3.1}$$

Observe that the FMI for individual i depends on the whole population of welfare values, $\{E_1, \ldots, E_N\}$.

Consider now a score variable S_i for i-th individual defined using the IFR approach, instead of a welfare variable E_i. These scores S_i are obtained by applying a multidimensional approach that takes into account a variety of non-monetary indicators of deprivation. Then the Fuzzy Supplementary Index (FSI) for individual i is defined analogously to the FMI, but in terms of the scores $\{S_1, \ldots, S_N\}$, as

$$\begin{aligned} FS_i &= \left\{\frac{N}{N-1}(1-F_S(S_i))\right\}^{\alpha-1}\{1-L_S(S_i)\} \\ &= \left\{\frac{1}{N-1}\sum_{j=1}^{N} I\left\{S_j > S_i\right\}\right\}^{\alpha-1}\left\{\frac{\sum_{j=1}^{N} S_j I\left\{S_j > S_i\right\}}{\sum_{j=1}^{N} S_j}\right\}, \quad i \in U. \end{aligned}$$

Here, $F_S(x)$ is the empirical distribution function and $L_S(x)$ the Lorenz curve of the score variables $\{S_1, \ldots, S_N\}$. Similarly, $1-F_S(S_i)$ is the proportion of individuals who are less deprived than individual

i and $1 - L_S(S_i)$ is the share of the total lack of deprivation score assigned to all individuals less deprived than individual i. The average FSI for the population is given by

$$FS = \frac{1}{N}\sum_{i=1}^{N} FS_i \tag{3.2}$$

Now consider that the population U is partitioned into D domains or areas $U_1, \ldots, U_D$ of sizes $N_1, \ldots, N_D$. Let E_{dj} be the welfare for individual j within domain d. The average fuzzy monetary index for domain d is

$$FM_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FM_{dj}, \quad d = 1, \ldots, D, \tag{3.3}$$

where FM_{dj} is the FMI for j-th individual from d-th domain.

A random sample $s \subseteq U$ of size $n \leq N$ is drawn from the population. Let s_d be the subsample from domain d, $d = 1, \ldots, D$. A design-based estimator of the average FMI for domain d, FM_d, is

$$\widehat{FM}_d^{DB} = \frac{\sum_{j \in s_d} w_{dj} \widehat{FM}_{dj}^{DB}}{\sum_{j \in s_d} w_{dj}}, \quad d = 1, \ldots, D, \tag{3.4}$$

where w_{dj} is the sampling weight for individual j within domain d and

$$\widehat{FM}_{dj}^{DB} = \left\{ \frac{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i}} \right\}^{\alpha - 1} \left\{ \frac{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i} E_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i} E_{\ell i}} \right\}. \tag{3.5}$$

Observe that $\widehat{FM}_{dj}^{DB}$ is not a direct estimators because it uses the sample data from the whole population and not only from domain d. The average FSI for domain d is given by

$$FS_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FS_{dj}, \quad d = 1, \ldots, D. \tag{3.6}$$

Finally, a design-based estimator of FS_d would be

$$\widehat{FS}_d^{DB} = \frac{\sum_{j \in s_d} w_{dj} \widehat{FS}_{dj}^{DB}}{\sum_{j \in s_d} w_{dj}}, \quad d = 1, \ldots, D. \tag{3.7}$$

where

$$\widehat{FS}_{dj}^{DB} = \left\{ \frac{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i} I\{S_{\ell i} > S_{dj}\}}{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i}} \right\}^{\alpha - 1} \left\{ \frac{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i} S_{\ell i} I\{S_{\ell i} > S_{dj}\}}{\sum_{\ell=1}^{D} \sum_{i \in s_d} w_{\ell i} S_{\ell i}} \right\}. \tag{3.8}$$

In these poverty indicators, the parameter α can be fixed to the value such that the FM and FS indicators coincide with the head count ratio computed for the official poverty line (60% of the median).

3.3 Fast Empirical Best Prediction

In order to apply the EB method of Molina and Rao (2010) to estimate the domain average FMI, FM_d, we need to express this indicator in terms of a population vector $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$, for which the conditional distribution of the non-sampled part $\mathbf{y}_r$ given the sample data $\mathbf{y}_s$ is known. The distribution of the welfare variable E_{dj} is seldom Normal. However, many times it is possible to find a transformation whose distribution is approximately Normal. Suppose that there exists a one-to-one transformation $Y_{dj} = T(E_{dj})$ of the welfare variable E_{dj}, which follows a Normal distribution. Concretely, we assume that Y_{dj} follows the nested error linear regression model of Battese, Harter and Fuller (1988), defined as

$$\begin{aligned} &Y_{dj} = \mathbf{x}_{dj}\beta + u_d + e_{dj}, \qquad j = 1,\dots,N_d, \quad d = 1,\dots D, \\ &u_d \sim \text{ iid } N(0,\sigma_u^2), \quad e_{dj} \sim \text{ iid } N(0,\sigma_e^2) \end{aligned} \tag{3.9}$$

where $\mathbf{x}_{dj}$ is a row vector with the values of p explanatory variables, u_d is a random area-specific effect and e_{dj} are residual errors. Let $\mathbf{y}_d = (Y_{d1},\dots,Y_{dN_D})'$ be vector of responses for domain d and $\mathbf{y} = (\mathbf{y}_1',\dots,\mathbf{y}_D')'$ be the full population vector. Then, observe that the individual FMIs can be expressed as

$$\begin{aligned} FM_{dj} &= \left\{ \frac{1}{N-1} \sum_{\ell=1}^{D} \sum_{i=1}^{N_\ell} I\left\{T^{-1}(Y_{\ell i}) > T^{-1}(Y_{dj})\right\} \right\}^{\alpha-1} \\ &\times \left\{ \frac{\sum_{\ell=1}^{D}\sum_{i=1}^{N_\ell} T^{-1}(Y_{\ell i}) I\left\{T^{-1}(Y_{\ell i}) > T^{-1}(Y_{dj})\right\}}{\sum_{\ell=1}^{D}\sum_{i=1}^{N_\ell} T^{-1}(Y_{\ell i})} \right\}, \quad j = 1,\dots,N_d,\ d = 1,\dots,D. \end{aligned}$$

This means that the average FMI for domain d is a non-linear function of the population vector $\mathbf{y}$, that is,

$$FM_d = \frac{1}{N_d} \sum_{j=1}^{N_d} FM_{dj} = h_d(\mathbf{y}), \quad d = 1,\dots,D.$$

Let us separate the population vector of responses $\mathbf{y}$ in the sample and non-sample parts, that is, $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$, where $\mathbf{y}_s$ corresponds to the sample and $\mathbf{y}_r$ to the non-sample. Then the BP of FM_d is

$$\widehat{FM}_d^B = E_{\mathbf{y}_r}(FM_d|\mathbf{y}_s) = E_{\mathbf{y}_r}(h_d(\mathbf{y})|\mathbf{y}_s). \tag{3.10}$$

This expectation can be empirically approximated by Monte Carlo simulation. For this, first fit the nested-error model (3.9) to the sample data $\mathbf{y}_s$, to obtain estimates $\hat{\beta}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ of the model parameters β, σ_u^2 and σ_e^2 respectively. Obtain also the EB predictor $\hat{u}_d$ of u_d, given by $E(u_d|\mathbf{y}_s)$ with unknown parameters replaced by estimated values. Then, using those estimates, generate a large number L of vectors $\mathbf{y}_r$ from the estimated conditional distribution $\mathbf{y}_r|\mathbf{y}_s$. Let $\mathbf{y}_r^{(l)}$ be the vector generated in l-th generation. We attach this vector to the sample vector to obtain the full population vector $\mathbf{y}^{(l)} = (\mathbf{y}_s', (\mathbf{y}_r^{(l)})')'$. Using the elements of $\mathbf{y}^{(l)}$, we calculate the domain parameter of interest $FM_d^{(l)} = h_d(\mathbf{y}^{(l)})$, $d = 1,\dots,D$. Then, a Monte Carlo approximation to the EB predictor of FM_d is given by

$$\widehat{FM}_d^{EB} \approx \frac{1}{L} \sum_{l=1}^{L} FM_d^{(l)}, \quad d = 1,\dots,D. \tag{3.11}$$

Observe that for each population $l = 1,\ldots,L$, instead of generating a multivariate normal vector of size $N-n$, we just need to generate univariate values Y_{dj} from

$$Y_{dj} = \mathbf{x}_{dj}\hat{\beta} + \hat{u}_d + v_d + \varepsilon_{di},\ v_d \sim N(0, \hat{\sigma}_u^2(1-\hat{\gamma}_d)),\ \varepsilon_{dj} \sim N(0, \hat{\sigma}_e^2),\ j \in U_d - s_d,\ d = 1,\ldots,D, \quad (3.12)$$

where $\gamma_d = \sigma_u^2(\sigma_u^2 + \sigma_e^2/n_d)^{-1}$ and n_d is the sample size in domain d. Still, for large populations and/or complex indicators, the EB method can be unfeasible. FMIs require sorting of all population elements, and this needs to be repeated for $l = 1,\ldots,L$. This is too time consuming for large N and large L. Here we propose a faster version the EB estimator that is based on replacing the true value of the domain average FMI in population l, $FM_d^{(l)}$, by the design-based estimator given in (3.11). Since the design-based estimator is obtained from a sample drawn from l-th population, this avoids the task of generation of the full population of responses (we need to generate only the responses for the sample elements) and the sorting of all the population elements. Concretely, for each Monte Carlo replication l, we take a sample $s(l) \subseteq U$ using the same sampling scheme and the same sample size allocation as in the original sample s. We take the values of the auxiliary variables corresponding to the units in $s(l)$, that is, we take $\mathbf{x}_{dj}$, $j \in s_d(l)$, where $s_d(l)$ is the subsample from d-th domain. Then we generate the corresponding responses Y_{dj}, $j \in s_d(l)$, for $d = 1,\ldots,D$, as in (3.12). Let us denote the vector containing those values as $\mathbf{y}_{s(l)}$. With $\mathbf{y}_{s(l)}$, calculate the design-based estimator as in (3.4) and (3.5), that is, obtain

$$\widehat{FM}_d^{DB}(l) = \frac{\sum_{j\in s_d(l)} w_{dj}\widehat{FM}_{dj}^{DB}(l)}{\sum_{j\in s_d(l)} w_{dj}}, \quad d = 1,\ldots,D, \quad (3.13)$$

where

$$\widehat{FM}_{dj}^{DB}(l) = \left\{\frac{\sum_{\ell=1}^{D}\sum_{i\in s_d(l)} w_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D}\sum_{i\in s_d(l)} w_{\ell i}}\right\}^{\alpha-1} \left\{\frac{\sum_{\ell=1}^{D}\sum_{i\in s_d(l)} w_{\ell i} E_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D}\sum_{i\in s_d(l)} w_{\ell i} E_{\ell i}}\right\}.$$

Finally, the fast EB estimator of FM_d is given by

$$\widehat{FM}_d^{FEB} = \frac{1}{L}\sum_{l=1}^{L}\widehat{FM}_{dj}^{DB}(l), \quad d = 1,\ldots,D.$$

As showed in the next section, a model-based simulation study has been carried out to study the performance of the proposed method to estimate a traditional poverty measures, the HCR, and the average FMI in small domains. Results indicate that the new method keeps similar properties of the standard EB, but it allows to overcome computational problems due to large populations or to more complex poverty measures such as the average FMI.

3.4 Model-based simulation experiment

A model based simulation experiment has been carried out to study the efficiency of the fast EB estimator of the HCR in comparison with the EB estimator. On the other hand, we compared the behaviour of the fast EB estimator of the average FMI with that of design-based and ELL estimators (Elbers et al., 2003). For this, we considered a population with $N = 20000$ units, partitioned into $D = 80$ domains with $N_d = 250$ units in each domain d, for $d = 1,\dots,D$. The response variables for the population units Y_{dj} were generated from the nested-error model (3.9) using an intercept and two auxiliary variables, that is, $\mathbf{x}_{dj} = (1, x_{dj1}, x_{dj2})$, where the values of the two auxiliary variables were generated from $x_{dj1} \sim \text{Binom}(1, 0.2)$ and $x_{dj2} \sim \text{Binom}(1, p_d)$ and , where

$$p_d = 0.3 + 0.5d/D, \quad d = 1,\dots,D.$$

We assume that the model responses Y_{dj} are the logarithm of the welfare variables E_{dj}. Thus, $E_{dj} = \exp(Y_{dj})$. A set of sample indices s_d with $n_d = 50$ was drawn independently from each domain d using simple random sampling without replacement (SRSWR). The values of the auxiliary variables for the population units and the sample indices were kept fixed over all Monte Carlo simulations. The intercept and the regression coefficients associated with the two auxiliary variables were taken as $\beta = (3, 0.03, -0.04)'$. The random area effects variance was taken as $\sigma_u^2 = (0.15)^2$ and the error variance as $\sigma_e^2 = (0.5)^2$. The poverty line z was fixed as z = 12, which is equal to 0.6 times the median of the welfare variables for a given generated population. We generated $I = 1000$ Monte Carlo population vectors $\mathbf{y}^{(i)}$ from the true model. For each population i, for $i = 1,\dots,I$, the following quantities were computed:

1. The true domain HCRs,

$$HCR_d^{(i)} = \frac{1}{N_d} \sum_{j=1}^{N_d} I(E_{dj}^{(i)} < z), \ E_{dj}^{(i)} = \exp(Y_{dj}^{(i)}), \quad d = 1,\dots,D,$$

and the true domain average FMIs obtained for $\alpha = 2$, that is,

$$FM_d^{(i)} = \frac{1}{N_d} \sum_{j=1}^{N_d} FM_{dj}^{(i)}, \quad d = 1,\dots,D,$$

where

$$FM_{dj}^{(i)} = \left\{ \frac{1}{N-1} \sum_{\ell=1}^{D} \sum_{k=1}^{N_\ell} I(E_{\ell k}^{(i)} > E_{dj}^{(i)}) \right\} \left\{ \frac{\sum_{\ell=1}^{D} \sum_{i=1}^{N_\ell} E_{\ell k}^{(i)} I(E_{\ell k}^{(i)} > E_{dj}^{(i)})}{\sum_{\ell=1}^{D} \sum_{k=1}^{N_\ell} E_{\ell k}^{(i)}} \right\}.$$

2. Let $\mathbf{y}_s^{(i)}$ be the sample part of the i-th population vector $\mathbf{y}^{(i)}$, which is obtained taking the elements of $\mathbf{y}^{(i)}$ whose index is contained in the original sample s. Design-based estimators of the domain HCR and of the FMI were calculated using the data from $\mathbf{y}_s^{(i)}$.

3. The nested-error model (3.9) was fitted to sample data $\mathbf{y}_s^{(i)}$ and model parameters β, σ_u^2 and σ_e^2 were substituted by their estimates.

4. $L = 50$ non-sampled vectors $\mathbf{y}_r^{(il)}$, $l = 1,\dots,L$ were generated from the conditional distribution $\mathbf{y}_r|\mathbf{y}_s^{(i)}$ using (3.12). The population vector $\mathbf{y}^{(il)}$ was formed attaching the sample data $\mathbf{y}_s^{(i)}$ to the generated non-sample data $\mathbf{y}_r^{(il)}$. Then the Monte Carlo approximations to the EBPs of the domain HCRs were calculated.

5. $L = 50$ samples $s^{(il)}$ were drawn from the population using the same sampling scheme as with the original sample s, that is, for each $l = 1,\dots,L$, a set of indexes $s_d^{(il)}$ was drawn from d-th domain using SRSWR. The corresponding responses Y_{dj}, $j \in s_d^{(il)}$, $d = 1,\dots,D$, were generated from (3.12) and the fast EB estimators of the domain HCRs and of the average FMIs were calculated.

6. ELL estimators (Elbers *et al.*, 2003) of the domain HCRs were also calculated. For this, first model (3.9) was fitted to sample data $\mathbf{y}_s$ and then A = 50 censuses were generated using a parametric bootstrap method (for details see Molina and Rao, 2009). For each population, the domain HCRs were calculated and the results were averaged over the A populations.

7. Means over Monte Carlo populations $i = 1,\dots,I$ of true values and of design-based, EB, fast EB and ELL estimators of domain HCRs and FMIs, were calculated. For the estimators, biases and MSEs over Monte Carlo populations $i = 1,\dots,I$ were also computed.

Figures 3.1, 3.2 and 3.3 show respectively the mean values, biases and MSEs of the HCR for each area. Observe in Figure 3.1 that the mean values of the fast EB estimators (labelled “EBnew”) are very close to those of the EB estimators. However, the design-based estimators (labelled as “Sample”), are more variable across areas, whereas the ELL estimators are less variable across areas, not tracking the true values. Moreover, from Figure 3.2, we can appreciate that the biases of the fast EB estimators are very similar to those of the EB estimators. Biases of all estimators are not significantly different. However, observe in Figures 3.3 that the MSEs of the EB and fast EB estimators are considerably below the MSEs of the other estimators for all areas, while the MSEs of the fast EB estimators are only slightly greater than those of the EB estimators. These results suggest that the new fast EB estimators can gain a lot in computational workload, while loosing little efficiency as compared to the EB estimators.

Analogously, Figures 3.4, 3.5 and 3.6 show respectively the mean values, biases and MSEs of design-based and fast EB estimators for the domain average FMI. Again, these figures show that the bias of the fast EB estimator is preserved small, similar to that of the design-based estimator, while the MSE is uniformly smaller for all areas.

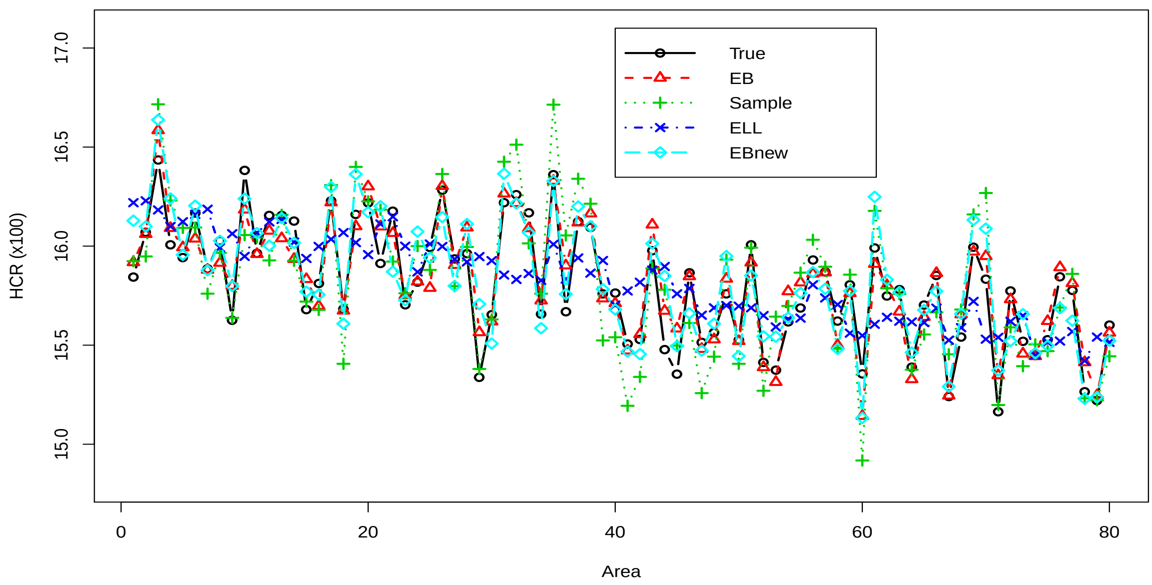


Figure 3.1: Mean over simulated populations of true values, EB, design-based (labelled “Sample”), ELL and fast EB estimators (labelled “EBnew”) of HCR for each area d.

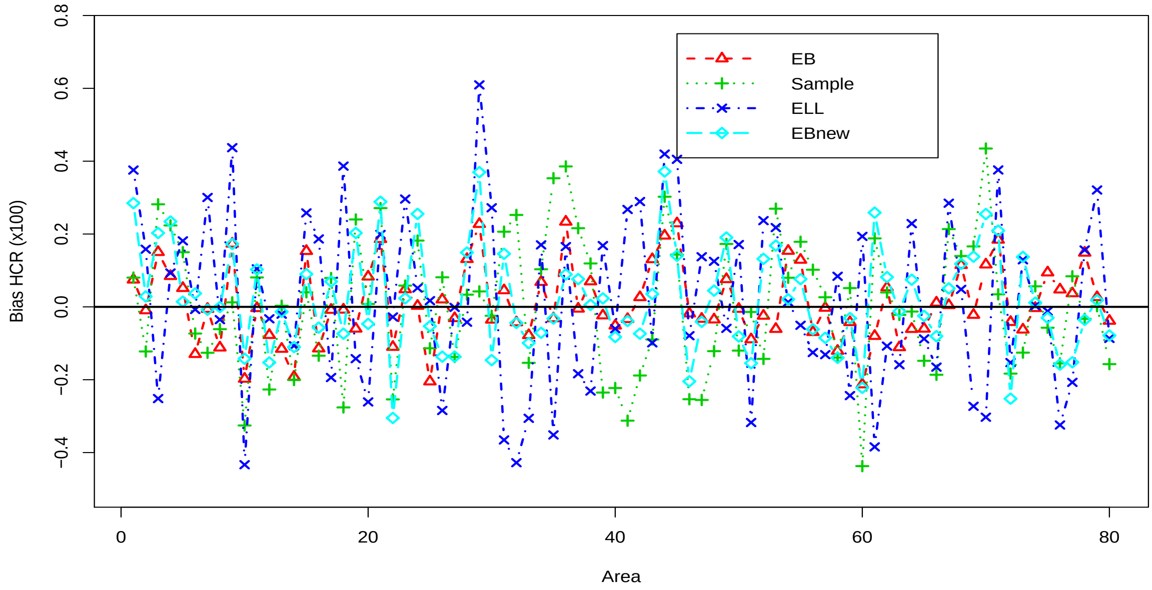


Figure 3.2: Bias ($\times$100) over simulated populations of EB, design-based (labelled “Sample”), ELL and fast EB estimators (labelled “EBnew”) of HCR for each area d.

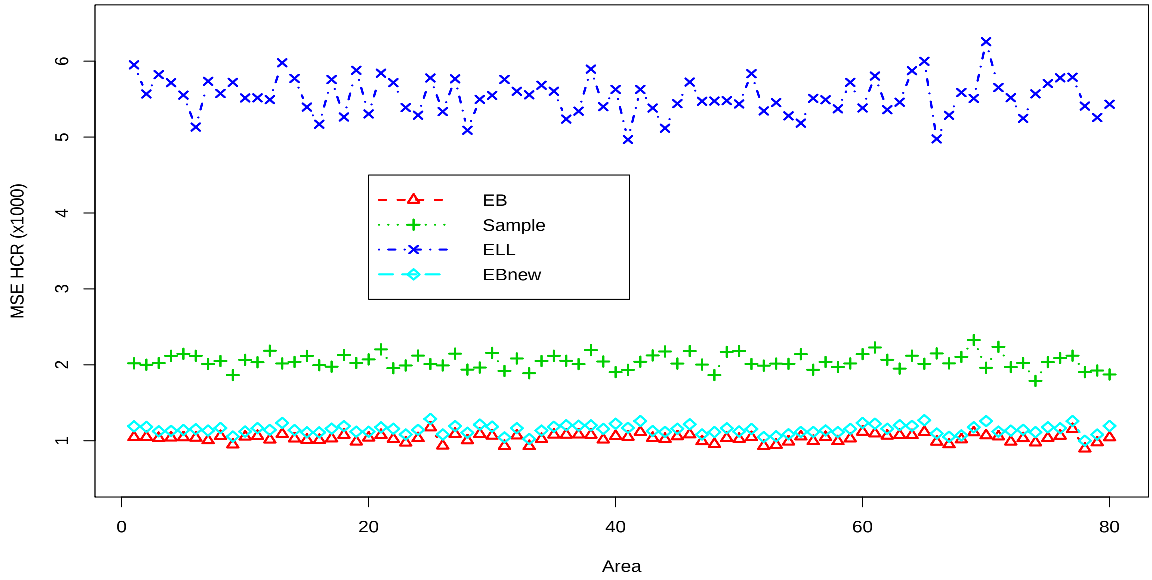


Figure 3.3: MSE (×1000) over simulated populations of EB, design-based (labelled “Sample”), ELL and fast EB estimators (labelled “EBnew”) of HCR for each area d.

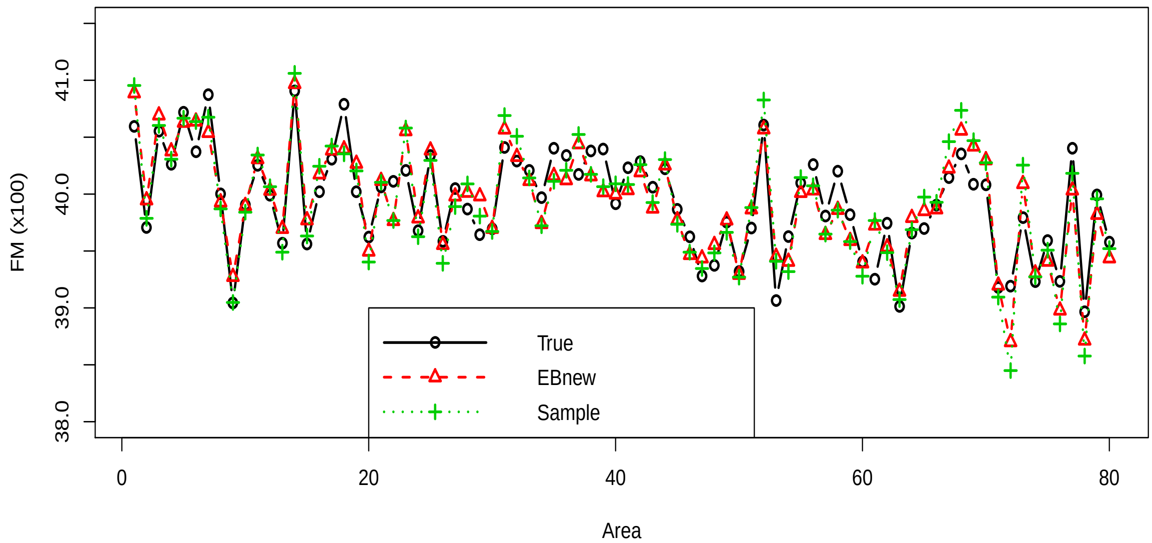


Figure 3.4: Mean over simulated populations of true values, fast EB (labelled “EBnew”) and design-based (labelled “Sample”) estimators of average FMI for each area d.

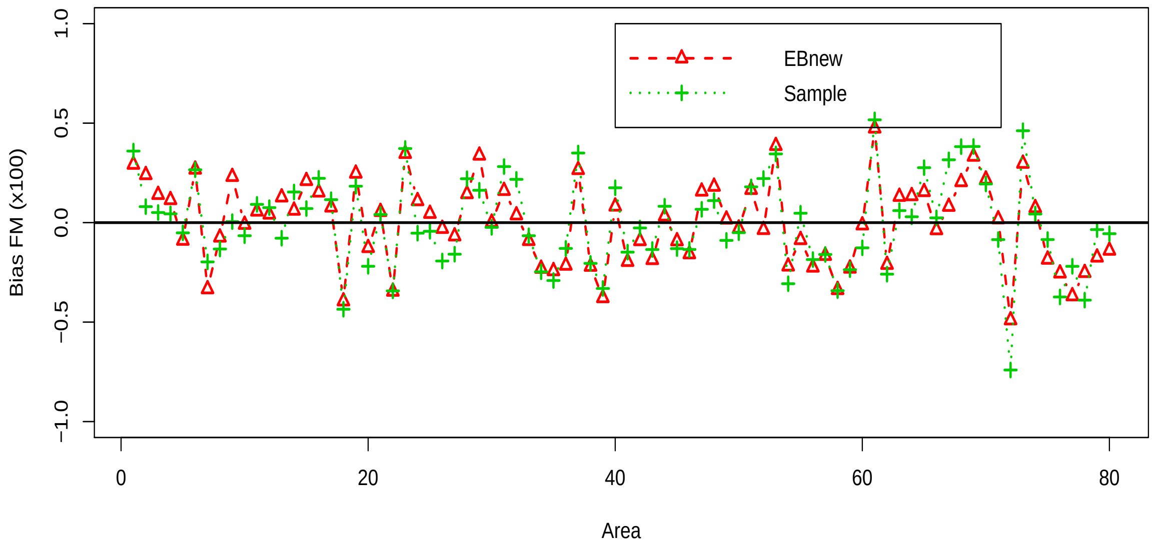


Figure 3.5: Bias (×100) over simulated populations of fast EB (labelled “EBnew”) and design-based (labelled “Sample”) estimators of average FMI for each area d.

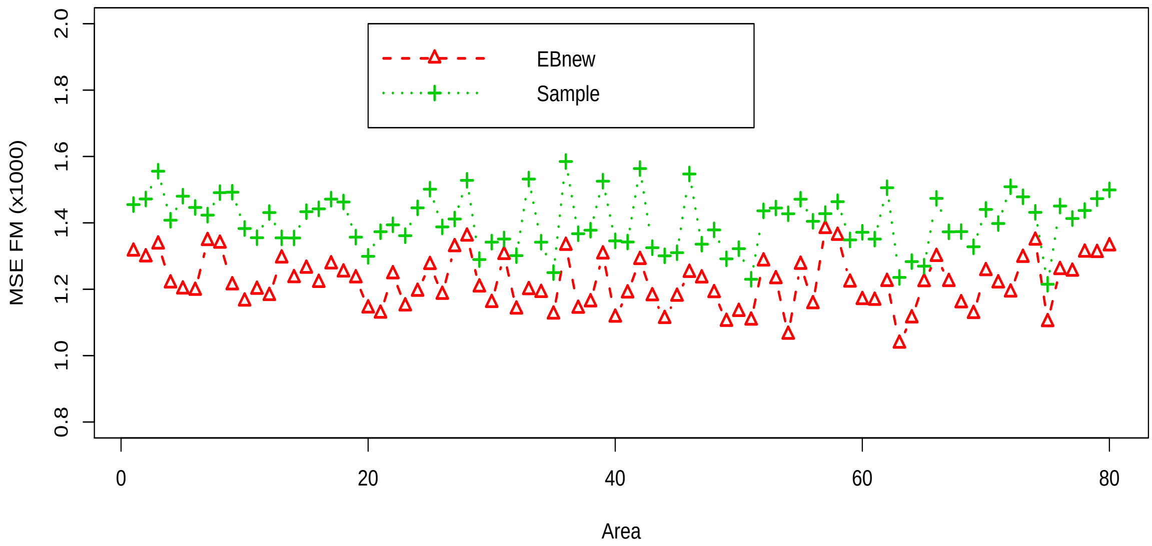


Figure 3.6: MSE (×1000) over simulated populations of fast EB (labelled “EBnew”) and design-based (labelled “Sample”) estimators of average FMI for each area d.

Chapter 4

Spatial Fay-Herriot models

In this chapter, the small area quantities of interest (e.g. the FGT poverty measures for Spanish provinces) are assumed to follow a Fay-Herriot model with spatial correlation. For this model, the EBLUP, called here Spatial EBLUP, is introduced and ML and REML model fitting methods are described. Analytical approximations of the mean squared error (MSE) of the Spatial EBLUP are discussed, and parametric and nonparametric bootstrap procedures for estimating the MSE are proposed. A simulation study based on the Italian Agriculture Census 2000 compares bootstrap and analytical estimates of the MSE and studies their robustness to non-normality. Results indicate good behavior of the non-parametric bootstrap under specific departures from normality. This chapter is based on the research of Molina and Rao (2009).

The contents of the chapter is the following. Section 4.2 presents the spatial Fay-Herriot model and it describes how the Spatial EBLUP is obtained from the model. Section 4.3 describes the available model fitting methods. Section 4.4 discusses the estimation of the MSE of the Spatial EBLUP, introducing some heuristic analytical approximations of this quantity together with an estimator. Section 4.5 introduces the mentioned parametric and nonparametric bootstrap methods for estimating the MSE. Section 4.6 describes the simulation study carried out for comparing the MSE estimators. The usefulness of the bootstrap techniques is demonstrated through a simulation study based on a real data set in Section 4.7, and finally, some conclusions are drawn in Section 4.8.

4.1 Introduction

Fay-Herriot (FH) models were introduced by Fay and Herriot (1979). to obtain small area estimators of median income in small places in the U.S. These models are well known in the literature of small area estimation (SAE) and are the basic tool when auxiliary data at the unit level are not available or there are confidentiality reasons preventing their use, and then only aggregated data at the small area level can be used. Even when unit-level auxiliary data are available, these models are still useful if the small area target parameter is not a linear function of the values of the response variable in the small area units. In the case of non-linear small area parameters, the BLUP and the EBLUP under a unit level model are not defined. However, in the FH models, this need for linearity is avoided by the fact that the model response is the direct estimator of the target parameter. Moreover, when the target parameter is obtained as an average over the area units of some quantities such as the FGT measures, then the Central Limit Theorem ensures that the distribution of the direct estimators (obtained also as averages) will not be too far from the Normal distribution.

In many practical applications, data from neighboring small areas display spatial correlation. In these cases, between-area correlation should be somehow represented in the covariance structure of the model unless sufficiently explaining covariates are available. However, the introduction of a dependence structure among small areas entails a serious conceptual difference with respect to the traditional framework of SAE, where the overall covariance matrix has a block-diagonal structure with block associated to the small areas Prasad and Rao. (1990).

In the context of SAE, Cressie (1991) introduced a model with spatially correlated random effects. More recently, an extension of the FH model through the Simultaneously Autoregressive (SAR) process has been considered by Singh et al. (2005), Petrucci and Salvati (2006) and Pratesi and Salvati (2008). When all parameters involved in the covariance matrix are known, Pratesi, M., Salvati, N. (2008) introduced the Spatial BLUP.

Usually, the model covariance matrix contains unknown parameters, called here variance components, which must be estimated from the available data. Replacing the derived estimates for the parameters in the Spatial BLUP leads to the so called Spatial EBLUP. Singh et al. (2005) proposed a second order approximation of the MSE of the Spatial EBLUP. However, this approximation does not tale into account the uncertainty due to estimation of the spatial autocorrelation parameter, and as shown by Pratesi and Salvati (2008), it might produce too optimistic or conservative confidence intervals depending on the strength of the spatial correlation and on the values of the sampling variances.

Resampling techniques are the alternative to heuristic analytical approximations. They are attractive for practitioners because of their conceptual simplicity and their easy application to complex statistical models. Furthermore, they usually require less assumptions and their performance relies less in the number of small areas. Some resampling procedures have been already proposed in the small area framework, see e.g. the jackknife method of Jiang and Lahiri (2002), the more recent parametric bootstrap approaches of González-Manteiga et al. (2007, 2008a, 2008b), Hall and Maiti (2006a) and Ugarte et al. (2008), and the nonparametric bootstrap of Hall and Maiti (2006b).

Here the parametric bootstrap of González-Manteiga et al. (2007) is extended to the spatial FH model. Moreover, a nonparametric approach is introduced that resamples both the random effects and the errors from the empirical distribution of their respective estimators. A simulation study compares the efficiency of the analytical and the bootstrap MSE estimators introduced in the paper for different levels of spatial autocorrelation, and analyzes the robustness of the bootstrap procedures to the absence of normality in the two random components of the model.

4.2 Spatial Fay-Herriot model

Consider a finite population partitioned into D small areas. The basic FH model relates linearly the quantity of inferential interest for d-th small area, θ_d, (e.g. the d-th area FGT poverty measure) to a vector of p area level auxiliary covariates $\mathbf{x}_d = (x_{d1}, x_{d2}, \ldots, x_{dp})$, and includes a random effect v_d associated to the area; that is,

$$\theta_d = \mathbf{x}_d \beta + v_d, \quad d = 1, \ldots, D. \tag{4.1}$$

Here β is the $p \times 1$ vector of regression parameters and the random effects $\{v_d; d = 1, \ldots, D\}$ are independent and identically distributed, each with mean 0 and variance σ_v^2. Model (4.1) is called linking model since all small areas are linked by the common β. Moreover, the FH model assumes that a design-unbiased direct estimator y_d of θ_d is available for each small area $d = 1,\dots,D$, and that these direct estimators can be expressed as

$$y_d = \theta_d + e_d, \quad d = 1,\dots,D, \tag{4.2}$$

where $\{e_d; d = 1,\dots,D\}$ are independent sampling errors, independent of the random effects v_d, and where e_d has mean 0 and variance ψ_d assumed to be known, $d = 1,\dots,D$. See Ghosh and Rao (1994). Model (4.2) is called sampling model. Combining both, the linking model (4.1) and the sampling model (4.2), we obtain the linear mixed model

$$y_d = \mathbf{x}_d\beta + v_d + e_d, \quad d = 1,\dots,D. \tag{4.3}$$

Let us define vectors $\mathbf{y} = (y_1,\dots,y_D)'$, $\mathbf{v} = (v_1,\dots,v_D)'$ and $\mathbf{e} = (e_1,\dots,e_D)'$, and matrices $\mathbf{X} = (\mathbf{x}_1',\dots,\mathbf{x}_D')'$ and $\Psi = \operatorname{diag}(\psi_1,\dots,\psi_D)$. Then the model in matrix notation is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{v} + \mathbf{e}. \tag{4.4}$$

Model (4.4) can be extended to allow for spatially correlated area effects as follows. Let $\mathbf{v}$ be the result of a SAR process with unknown autoregression parameter ρ and proximity matrix $\mathbf{W}$ (see Anselin (1988) and Cressie(1993)), i.e.,

$$\mathbf{v} = \rho\mathbf{W}\mathbf{v} + \mathbf{u}. \tag{4.5}$$

We assume that the matrix $(\mathbf{I}_D - \rho\mathbf{W})$ is non-singular. Then $\mathbf{v}$ can be expressed as

$$\mathbf{v} = (\mathbf{I}_D - \rho\mathbf{W})^{-1}\mathbf{u}. \tag{4.6}$$

Here, $\mathbf{u} = (u_1,\dots,u_D)'$ is a vector with mean $\mathbf{0}$ and covariance matrix $\sigma_u^2\mathbf{I}_D$, where $\mathbf{I}_D$ denotes the $D\times D$ identity matrix and σ_u^2 is an unknown parameter. We consider that the proximity matrix $\mathbf{W}$ is defined in row standardized form; that is, $\mathbf{W}$ is row stochastic. Then, $\rho \in (-1,1)$ is called spatial autocorrelation parameter Banerjee et al.(2004). Hereafter, the vector of variance components will be denoted $\omega = (\omega_1,\omega_2)' = (\sigma_u^2,\rho)'$. Equation (4.6) implies that $\mathbf{v}$ has mean vector $\mathbf{0}$ and covariance matrix equal to

$$\mathbf{G}(\omega) = \sigma_u^2[(\mathbf{I}_D - \rho\mathbf{W})'(\mathbf{I}_D - \rho\mathbf{W})]^{-1}. \tag{4.7}$$

Since $\mathbf{e}$ is independent of $\mathbf{v}$, the covariance matrix of $\mathbf{y}$ is equal to

$$\mathbf{V}(\omega) = \mathbf{G}(\omega) + \Psi.$$

Combining (4.4) and (4.6) the model is

$$\mathbf{y} = \mathbf{X}\beta + (\mathbf{I}_D - \rho\mathbf{W})^{-1}\mathbf{u} + \mathbf{e} \tag{4.8}$$

Under model (4.8), the Spatial BLUP of the quantity of interest $\theta_d = \mathbf{x}_d\beta + v_d$ is

$$\tilde{\theta}_d(\omega) = \mathbf{x}_d\tilde{\beta}(\omega) + \mathbf{b}_d'\mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)[\mathbf{y} - \mathbf{X}\tilde{\beta}(\omega)], \tag{4.9}$$

where $\tilde{\beta}(\omega) = [\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{y}$ is the generalized least squares estimator of the regression parameter β and $\mathbf{b}_d'$ is the $1\times D$ vector $(0,\dots,0,1,0,\dots,0)$ with 1 in the d-th position. The Spatial BLUP $\tilde{\theta}_d(\omega)$ depends on the unknown vector of variance components $\omega = (\sigma_u^2,\rho)'$. The two stage estimator $\tilde{\theta}_d(\hat{\omega})$ obtained by replacing ω in expression (4.9) by a consistent estimator $\hat{\omega} = (\hat{\sigma}_u^2,\hat{\rho})'$ is called Spatial EBLUP (see Singh et al.(2005) and Petrucci and Salvati(2006)).

4.3 Fitting methods based on the likelihood

Assuming normality of the random effects and the errors, the variance components $\omega = (\sigma_u^2, \rho)'$ can be estimated by ML or REML procedures. In fact, under regularity conditions, the estimators derived from these two methods (and using the Normal likelihood) remain consistent at order $O_p(D^{-1/2})$ even without the Normality assumption, for details see Jiang(1996).

A maximum likelihood estimator (MLE) of $\omega = (\sigma_u^2, \rho)'$ is obtained maximizing the log-likelihood of ω given the data vector $\mathbf{y}$,

$$\ell(\omega;\mathbf{y}) = c - \frac{1}{2}\log|\mathbf{V}(\omega)| - \frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}(\omega)(\mathbf{y}-\mathbf{X}\beta),$$

where c denotes a constant. In practice, an iterative algorithm such as the Fisher-scoring algorithm must be applied to maximize the likelihood. Let $\mathbf{S}(\omega) = (S_{\sigma_u^2}, S_\rho)'$ be the scores or derivatives of the log-likelihood with respect to σ_u^2 and ρ, and let $\mathcal{I}(\omega)$ be the Fisher information matrix obtained from $\ell(\omega;\mathbf{y})$, with elements

$$\mathcal{I}(\omega) = \begin{pmatrix} \mathcal{I}_{\sigma_u^2,\sigma_u^2} & \mathcal{I}_{\sigma_u^2,\rho^2} \\ \mathcal{I}_{\rho,\sigma_u^2} & \mathcal{I}_{\rho,\rho} \end{pmatrix}.$$

Then the Fisher-scoring algorithm starts with an initial estimate $\omega^{(0)} = (\sigma_u^{2(0)}, \rho^{(0)})'$ and then at each iteration k, this estimate are updated with the equation

$$\omega^{(k+1)} = \omega^{(k)} + \mathcal{I}^{-1}(\omega^{(k)})\mathbf{S}(\omega^{(k)}).$$

The ML equation for β obtained by equating the corresponding score to zero yields

$$\tilde{\beta}(\omega) = [\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{y}. \tag{4.10}$$

Let us denote

$$\mathbf{C}(\rho) = (\mathbf{I}_D - \rho\mathbf{W})'(\mathbf{I}_D - \rho\mathbf{W})$$

and

$$\mathbf{P}(\omega) = \mathbf{V}^{-1}(\omega) - \mathbf{V}^{-1}(\omega)\mathbf{X}\left[\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{V}^{-1}(\omega).$$

Then the derivative of $\mathbf{C}(\rho)$ with respect to ρ is

$$\frac{\partial\mathbf{C}(\rho)}{\partial\rho} = -\mathbf{W} - \mathbf{W}' + 2\rho\mathbf{W}'\mathbf{W}$$

and the derivatives of $\mathbf{V}(\omega)$ with respect to σ_u^2 and ρ are respectively given by

$$\frac{\partial\mathbf{V}(\omega)}{\partial\sigma_u^2} = \mathbf{C}^{-1}(\rho), \quad \frac{\partial\mathbf{V}(\omega)}{\partial\rho} = -\sigma_u^2\mathbf{C}^{-1}(\rho)\frac{\partial\mathbf{C}(\rho)}{\partial\rho}\mathbf{C}^{-1}(\rho) \triangleq \mathbf{A}(\omega).$$

The scores associated to σ_u^2 and ρ, after replacing (4.10), are given by

$$\begin{aligned} S_{\sigma_u^2} &= -\frac{1}{2}\operatorname{trace}\left\{\mathbf{V}^{-1}(\omega)\mathbf{C}^{-1}(\rho)\right\} + \frac{1}{2}\mathbf{y}'\mathbf{P}(\omega)\mathbf{C}^{-1}(\rho)\mathbf{P}(\omega)\mathbf{y}, \\ S_\rho &= -\frac{1}{2}\operatorname{trace}\left\{\mathbf{V}^{-1}(\omega)\mathbf{A}^{-1}(\omega)\right\} + \frac{1}{2}\mathbf{y}'\mathbf{P}(\omega)\mathbf{A}(\omega)\mathbf{P}(\omega)\mathbf{y}. \end{aligned}$$

The elements of the Fisher information matrix are

$$I_{\sigma_u^2,\sigma_u^2} = \frac{1}{2}\,\mathrm{trace}\left\{\mathbf{V}^{-1}(\omega)\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega)\mathbf{C}^{-1}(\rho)\right\},$$
$$I_{\sigma_u^2,\rho} = I_{\rho,\sigma_u^2} = \frac{1}{2}\,\mathrm{trace}\left\{\mathbf{V}^{-1}(\omega)\mathbf{A}(\omega)\mathbf{V}^{-1}(\omega)\mathbf{C}^{-1}(\rho)\right\},$$
$$I_{\rho,\rho} = \frac{1}{2}\,\mathrm{trace}\left\{\mathbf{V}^{-1}(\omega)\mathbf{A}(\omega)\mathbf{V}^{-1}(\omega)\mathbf{A}(\omega)\right\}.$$

A restricted maximum likelihood estimator (RMLE) of ω is obtained by maximizing the restricted likelihood, which is the likelihood of ω after eliminating the vector of coefficients β. Let $\mathbf{F}$ be an $D\times p$ matrix satisfying $\mathbf{F}'\mathbf{X} = \mathbf{0}$. Then, the restricted log-likelihood is the likelihood of the transformed data $\mathbf{F}'\mathbf{y}$ and is given by

$$\ell_R(\omega;\mathbf{y}) = c - \frac{1}{2}\log|\mathbf{F}'\mathbf{V}(\omega)\mathbf{F}| - \frac{1}{2}\mathbf{y}'\mathbf{F}(\mathbf{F}'\mathbf{V}(\omega)\mathbf{F})^{-1}\mathbf{F}'\mathbf{y}.$$

It can be shown that

$$\mathbf{F}\left[\mathbf{F}'\mathbf{V}(\omega)\mathbf{F}\right]^{-1}\mathbf{F}' = \mathbf{P}(\omega),$$

so that the restricted log-likelihood becomes

$$\ell_R(\omega;\mathbf{y}) = c - \frac{1}{2}\log|\mathbf{F}'\mathbf{V}(\omega)\mathbf{F}| - \frac{1}{2}\mathbf{y}'\mathbf{P}(\omega)\mathbf{y}.$$

Using the following properties of the matrix $\mathbf{P}(\omega)$,

$$\mathbf{P}(\omega)\mathbf{V}(\omega)\mathbf{P}(\omega) = \mathbf{P}(\omega), \quad \frac{\partial\mathbf{P}(\omega)}{\partial\omega_j} = -\mathbf{P}(\omega)\frac{\partial\mathbf{V}(\omega)}{\partial\omega_j}\mathbf{P}(\omega),$$

we obtain the scores corresponding to this restricted log-likelihood,

$$S^R_{\sigma_u^2} = -\frac{1}{2}\,\mathrm{trace}\left\{\mathbf{P}(\omega)\mathbf{C}^{-1}(\rho)\right\} + \frac{1}{2}\mathbf{y}'\mathbf{P}(\omega)\mathbf{C}^{-1}(\rho)\mathbf{P}(\omega)\mathbf{y},$$
$$S^R_{\rho} = -\frac{1}{2}\,\mathrm{trace}\left\{\mathbf{P}\omega)\mathbf{A}(\omega)\right\} + \frac{1}{2}\mathbf{y}'\mathbf{P}(\omega)\mathbf{A}(\omega)\mathbf{P}(\omega)\mathbf{y},$$

Finally, the elements of the Fisher information obtained from ℓ_R are

$$I^R_{\sigma_u^2,\sigma_u^2} = \frac{1}{2}\mathrm{tr}\{\mathbf{P}(\omega)\mathbf{C}^{-1}(\rho)\mathbf{P}(\omega)\mathbf{C}^{-1}(\rho)\},$$
$$I^R_{\sigma_u^2,\rho} = I^R_{\rho,\sigma_u^2} = \frac{1}{2}\mathrm{tr}\{\mathbf{P}(\omega)\mathbf{A}(\omega)\mathbf{P}(\omega)\mathbf{C}^{-1}(\rho)\},$$
$$I^R_{\rho,\rho} = \frac{1}{2}\mathrm{tr}\{\mathbf{P}(\omega)\mathbf{A}(\omega)\mathbf{P}(\omega)\mathbf{A}(\omega)\}.$$

4.4 Analytical approximation of the MSE

In practical applications, the Spatial EBLUP $\tilde{\theta}_d(\hat{\omega})$ should be accompanied with its estimated MSE. Under normality of random effects and errors, the MSE of the Spatial EBLUP can be decomposed as

$$\begin{aligned} MSE[\tilde{\theta}_d(\hat{\omega})] &= MSE[\tilde{\theta}_d(\omega)] + E\{[\tilde{\theta}_d(\hat{\omega}) - \tilde{\theta}_d(\omega)]^2\} \\ &= [g_{1d}(\omega) + g_{2d}(\omega)] + g_{3d}(\omega), \end{aligned} \quad (4.11)$$

where $g_{1d}(\omega)$ represents the uncertainty due to the prediction of the random effects and is of order $O(1)$ for large D, $g_{2d}(\omega)$ is due to the estimation of β and is of order $O(D^{-1})$, and the last term measures the uncertainty of the Spatial EBLUP arising from the estimation of the variance components and is of lower order (see Singh et al.(2005)). Exact analytical expressions for the first two terms are easily calculated because the Spatial BLUP $\tilde{\theta}_d(\omega)$ is a linear function of the data vector $\mathbf{y}$, and they are given by

$$g_{1d}(\omega) = \mathbf{b}_d'[\mathbf{G}(\omega) - \mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)\mathbf{G}(\omega)]\mathbf{b}_d, \tag{4.12}$$
$$g_{2d}(\omega) = \mathbf{b}_d'[\mathbf{I}_D - \mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)]\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X})^{-1}\mathbf{X}'[\mathbf{I}_D - \mathbf{V}^{-1}(\omega)\mathbf{G}(\omega)]\mathbf{b}_d. \tag{4.13}$$

However, for the last term $g_{3d}(\omega) = E\{[\tilde{\theta}_d(\hat{\omega}) - \tilde{\theta}_d(\omega)]^2\}$, an exact analytical expression does not exist due to the non-linearity of the EBLUP $\tilde{\theta}_d(\hat{\omega})$ in the data vector $\mathbf{y}$. Under the FH model (4.3) with independent random effects v_d (diagonal covariance matrix $\mathbf{V}$), Prasad and Rao(1990) obtained an approximation up to $o(D^{-1})$ terms of $g_{3d}(\omega)$ through Taylor linearization. Their formula can be taken as a naive approximation of the true $g_{3d}(\omega)$ under model (4.4)–(4.5). Straightforward application of this formula to model (4.4)–(4.5) yields

$$g_{3d}^{PR}(\omega) = \text{trace}\left\{\mathbf{L}_d(\omega)\mathbf{V}(\omega)\mathbf{L}_d'(\omega)\mathcal{I}^{-1}(\omega)\right\},$$

where

$$\mathbf{L}_d(\omega) = \begin{pmatrix} \mathbf{b}_d'\left[\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega) - \sigma_u^2\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega)\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega)\right] \\ \mathbf{b}_d'\left[\mathbf{A}(\omega)\mathbf{V}^{-1}(\omega) - \sigma_u^2\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega)\mathbf{A}(\omega)\mathbf{V}^{-1}(\omega)\right] \end{pmatrix}.$$

Then the full MSE can be approximated by

$$MSE^{PR}[\tilde{\theta}_d(\hat{\omega})] = g_{1d}(\omega) + g_{2d}(\omega) + g_{3d}^{PR}(\omega). \tag{4.14}$$

Singh et al.(2005) arrived to the same formula (4.14) for the true MSE under a Fay-Herriot model with random effects following a SAR process. However, this formula is not accounting for the extra uncertainty of the Spatial EBLUP $\tilde{\theta}_d(\hat{\omega})$ due to the estimation of the autocorrelation parameter ρ.

Concerning MSE estimation, under FH models with diagonal covariance matrix, Prasad and Rao (1990) obtained an approximately unbiased estimator of (4.14). Following the results of Harville and Jeske (1992), Zimmerman and Cressie (1992) extended the Prasad-Rao MSE estimator to models with more general covariance structure. The authors refer to geostatistical models, in which the correlation matrix is directly specified, and they assume that the covariance function is linear in the parameters. This situation is likely to occur under geostatistical models where the covariance function depends on the distance between locations. Under SAR models, the covariance is assumed to depend on a proximity matrix that specifies the proximity between the areas. Even so, SAR models lead to a covariance function that is similar to the Bessel variogram model by Griffith and Csillag (1993). Then, following the results of Zimmerman and Cressie (1992), when $\hat{\omega}$ is the REML estimator of ω, an approximately unbiased estimator of the MSE is

$$mse^{PR}[\tilde{\theta}_d(\hat{\omega})] = g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega}) + 2g_{3d}^{PR}(\hat{\omega}), \tag{4.15}$$

which is the same estimator derived by Prasad and Rao (1990). In formula (4.15), the term $g_{3d}^{PR}(\hat{\omega})$ appears twice due to a bias correction of $g_{1d}(\hat{\omega})$. If $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$ is obtained by ML, then an approximately unbiased estimator of the MSE is

$$mse_{ML}^{PR}[\tilde{\theta}_d(\hat{\omega})] = g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega}) + 2\tilde{g}_{3d}(\hat{\omega}) - \mathbf{b}_{ML}^T(\hat{\omega})\nabla g_{1d}(\hat{\omega}), \tag{4.16}$$

where $\nabla g_{1d}(\omega) = \partial g_{1d}(\omega)/\partial\omega$ is the gradient of $g_{1d}(\omega)$ and $\mathbf{b}_{ML}(\hat{\omega})$ is the bias of the ML estimator $\hat{\omega}$ up to order $o(D^{-1})$. This bias is equal to $\mathbf{b}_{ML}(\hat{\omega}) = \mathcal{I}^{-1}(\hat{\omega})\mathbf{h}(\hat{\omega})/2$ with $\mathbf{h}(\hat{\omega}) = (h_1(\hat{\omega}), h_2(\hat{\omega}))'$ and

$$h_k(\omega) = \text{trace}\left\{ \left[\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}\right]^{-1} \frac{\partial\left[\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}\right]}{\partial\omega_k} \right\}, \quad k = 1,2.$$

Ignoring the last term in (4.16) could lead to underestimation of the MSE (see e.g. Petrucci and Salvati (2006)). Finally, Singh et al. (2005) derived a different MSE estimator. When $\hat{\omega}$ is obtained by REML method, their estimator is given by

$$mse^{SSK}[\tilde{\theta}_d(\hat{\omega})] = g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega}) + 2g_{3d}^{PR}(\hat{\omega}) - g_{4d}(\hat{\omega}). \tag{4.17}$$

This estimator differs from (4.15) and (4.16) in the subtraction of the extra term $g_{4d}(\hat{\omega})$, where $g_{4d}(\omega)$ is given by

$$g_{4d}(\omega) = \frac{1}{2}\sum_{k=1}^{2}\sum_{\ell=1}^{2} \mathbf{b}_d'\Psi\mathbf{V}^{-1}(\omega)\frac{\partial^2\mathbf{V}(\omega)}{\partial\omega_k\partial\omega_\ell}\mathbf{V}^{-1}(\omega)\Psi\,\mathcal{I}_{k\ell}^{-1}(\omega)\mathbf{b}_d.$$

When $\hat{\omega}$ is obtained by ML, their estimator is obtained by subtracting $g_{4d}(\hat{\omega})$ in (4.16).

4.5 Parametric bootstrap estimation of the MSE

In the previous section, we decompose the MSE of the spatial EBLUP in three components, $g_{1d}(\omega)$, $g_{2d}(\omega)$ and $g_{3d}(\omega)$. The first two have exact closed formulas which does not happen for the third component. This reflects the additional uncertainty coming from the estimation of the variance components $\omega = (\sigma_u^2, \rho)$.

In this section, we propose to use the parametric bootstrap of González-Manteiga et al. (2007) extended to the FH model with spatial correlation (4.4)–(4.5), to derive an estimator for the full MSE, which is consistent if the estimators of the model parameters are consistent. In order to check the consistency of the full MSE, we can use, as in González-Manteiga et al. (2007), the asymptotic formula of the MSE obtained by Singh et al. (2005). The extended parametric bootstrap is composed of 8 steps as follows:

(step 1) Obtain the estimates $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$ and $\hat{\beta} = \tilde{\beta}(\hat{\omega})$ by fitting the model (4.8) to the initial data $\mathbf{y} = (y_1, \ldots, y_D)'$.

(step 2) Generate a vector $\mathbf{t}_1^*$ whose D elements are independent $N(0,1)$. Build bootstrap vectors $\mathbf{u}^* = \hat{\sigma}_u\mathbf{t}_1^*$ and $\mathbf{v}^* = (\mathbf{I}_D - \hat{\rho}\mathbf{W})^{-1}\mathbf{u}^*$, and calculate $\theta^* = \mathbf{X}\hat{\beta} + \mathbf{v}^*$, where $\hat{\beta}$ and $\hat{\omega}$ are viewed as the true values of the parameters.

(step 3) Generate a vector $\mathbf{t}_2^*$ with D independent $N(0,1)$ elements, which is independent of $\mathbf{t}_1^*$. Then, construct the vector of random errors as $\mathbf{e}^* = \Psi^{1/2}\,\mathbf{t}_2^*$.

(step 4) Obtain bootstrap data $\mathbf{y}^*$ directly applying the model, $\mathbf{y}^* = \theta^* + \mathbf{e}^* = \mathbf{X}\hat{\beta} + \mathbf{v}^* + \mathbf{e}^*$.

(step 5) Fit the model (4.8) to the bootstrap data $\mathbf{y}^*$ using $\hat{\beta}$ and $\hat{\omega}$ as the true values of β and ω. The estimates of the "true" $\hat{\beta}$ and $\hat{\omega}$ are obtained based on bootstrap data $\mathbf{y}^*$, by calculating the estimator of $\hat{\beta}$ at the "true" $\hat{\omega}$,

$$\tilde{\beta}^*(\hat{\omega}) = \left[\mathbf{X}'\mathbf{V}^{-1}(\hat{\omega})\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{V}^{-1}(\hat{\omega})\mathbf{y}^*;$$

then, obtain the estimator $\hat{\omega}^*$ based on $\mathbf{y}^*$. Finally, the estimator of $\hat{\beta}$ calculated at $\hat{\omega}^*$ is $\tilde{\beta}^*(\hat{\omega}^*)$.

(step 6) Calculate the bootstrap Spatial BLUP from bootstrap data $\mathbf{y}^*$ using $\hat{\omega}$ as the true value of ω,

$$\tilde{\theta}_d^*(\hat{\omega}) = \mathbf{x}_d\tilde{\beta}^*(\hat{\omega}) + \mathbf{b}_d'\mathbf{G}(\hat{\omega})\mathbf{V}(\hat{\omega})^{-1}[\mathbf{y}^* - \mathbf{X}\tilde{\beta}^*(\hat{\omega})].$$

Then, compute the bootstrap Spatial EBLUP using $\hat{\omega}^*$ in place of the "true" $\hat{\omega}$ as,

$$\tilde{\theta}_d^*(\hat{\omega}^*) = \mathbf{x}_d\tilde{\beta}^*(\hat{\omega}^*) + \mathbf{b}_d'\mathbf{G}(\hat{\omega}^*)\mathbf{V}^{-1}(\hat{\omega}^*)[\mathbf{y}^* - \mathbf{X}\tilde{\beta}^*(\hat{\omega}^*)].$$

(step 7) Repeat steps (2)–(6) B times. In the b-th bootstrap replication, $\theta_d^{*(b)}$ is the quantity of interest for d-th area, $\hat{\omega}^{*(b)}$ the bootstrap estimate of ω, $\tilde{\theta}_d^{*(b)}(\hat{\omega})$ the bootstrap Spatial BLUP and $\tilde{\theta}_d^{*(b)}(\hat{\omega}^{*(b)})$ is the bootstrap Spatial EBLUP for d-th area.

(step 8) A parametric bootstrap estimator of $g_{3d}(\omega)$ is

$$g_{3d}^{PB}(\hat{\omega}) = B^{-1}\sum_{b=1}^{B}\left[\tilde{\theta}_d^{*(b)}(\hat{\omega}^{*(b)}) - \tilde{\theta}_d^{*(b)}(\hat{\omega})\right]^2,$$

and a naive parametric bootstrap estimator of the full MSE is is given by

$$mse^{naPB}[\tilde{\theta}_d(\hat{\omega})] = B^{-1}\sum_{b=1}^{B}\left[\tilde{\theta}_d^{*(b)}(\hat{\omega}^{*(b)}) - \theta_d^{*(b)}\right]^2. \tag{4.18}$$

We can also obtain an alternative estimator of the MSE by adding the analytical estimates $g_{1d}(\hat{\omega})$ and $g_{2d}(\hat{\omega})$, the bootstrap estimate $g_{3d}^{PB}(\hat{\omega})$, and a bootstrap bias correction of $g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega})$ to obtain a MSE estimate similar to the one of Pfeffermann and Tiller (2006). The alternative final estimator is

$$mse^{bcPB}[\tilde{\theta}_d(\hat{\omega})] = 2\left[g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega})\right] - B^{-1}\sum_{b=1}^{B}\left[g_{1d}(\hat{\omega}^{*(b)}) + g_{2d}(\hat{\omega}^{*(b)})\right] + g_{3d}^{PB}(\hat{\omega}). \tag{4.19}$$

4.6 Nonparametric bootstrap

The aim of this section is to present a nonparametric bootstrap for the MSE estimation. The random effects $\{u_1^*,\dots,u_D^*\}$ and random errors $\{e_1^*,\dots,e_D^*\}$ are obtained by resampling respectively from both the empirical distribution of predicted random effects $\{\hat{u}_1,\dots,\hat{u}_D\}$ and of residuals $\{\hat{r}_1,\dots,\hat{r}_D\}$, where $r_d = y_d - \tilde{\theta}_d(\hat{\omega})$, $d = 1,\dots,D$, both previously standardized. The nonparametric bootstrap is more robust to the non-normality of any of the random components of the model since it does not assume any distribution for them.

Under model (4.4)–(4.5), the BLUPs of $\mathbf{u}$ and $\mathbf{v}$ are respectively

$$\tilde{\mathbf{v}}(\omega) = \mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)[\mathbf{y} - \mathbf{X}\tilde{\beta}(\omega)], \quad \tilde{\mathbf{u}}(\omega) = (\mathbf{I} - \rho\mathbf{W})\tilde{\mathbf{v}}(\omega),$$

with $\tilde{\mathbf{u}}(\omega)$ covariance matrix given by

$$\mathbf{V}_{\mathbf{u}}(\omega) = (\mathbf{I} - \rho\mathbf{W})\mathbf{G}(\omega)\mathbf{P}(\omega)\mathbf{G}(\omega)(\mathbf{I} - \rho\mathbf{W}').$$

Furthermore, the vector of residuals is

$$\tilde{\mathbf{r}}(\omega) = \mathbf{y} - \mathbf{X}\tilde{\beta}(\omega) - \tilde{\mathbf{v}}(\omega) = (y_1 - \tilde{\theta}_1(\omega), \ldots, y_D - \tilde{\theta}_D(\omega))',$$

with covariance matrix

$$\mathbf{V}_{\mathbf{r}}(\omega) = \Psi \mathbf{P}(\omega)\Psi.$$

$\mathbf{V}_{\mathbf{u}}(\omega)$ and $\mathbf{V}_{\mathbf{r}}(\omega)$ are not diagonal since the elements of the vectors $\tilde{\mathbf{u}}(\omega)$ and $\tilde{\mathbf{r}}(\omega)$ are correlated and, therefore, we have to proceed by applying a standardization to these vectors (resampling from the empirical distribution requires an *iid* setup). Our proposal relies on transforming both $\hat{\mathbf{u}} = \tilde{\mathbf{u}}(\hat{\omega})$ and $\hat{\mathbf{r}} = \tilde{\mathbf{r}}(\hat{\omega})$ to make them as close as possible to vectors with uncorrelated elements and unit variances. The method is the following for the $\hat{\mathbf{u}}$ vector (the same applies to the $\hat{\mathbf{r}}$ vector): First, we obtain the spectral decomposition of $\hat{\mathbf{V}}_{\mathbf{u}} = \mathbf{V}_{\mathbf{u}}(\hat{\omega})$ as

$$\hat{\mathbf{V}}_{\mathbf{u}} = \mathbf{Q}_{\mathbf{u}}\Delta_{\mathbf{u}}\mathbf{Q}_{\mathbf{u}}',$$

where $\Delta_{\mathbf{u}}$ is a diagonal matrix with the $m-p$ non-zero eigenvalues of $\hat{\mathbf{V}}_{\mathbf{u}}$ and $\mathbf{Q}_{\mathbf{u}}$ is the matrix with the corresponding eigenvectors in the columns. Keep in view that $\tilde{\mathbf{u}}(\omega)$ lies in a $m-p$ dimension space. Second, we square the matrix $\hat{\mathbf{V}}_{\mathbf{u}}^{-1/2} = \mathbf{Q}_{\mathbf{u}}\Delta_{\mathbf{u}}^{-1/2}\mathbf{Q}_{\mathbf{u}}'$ to obtain a generalized inverse of $\hat{\mathbf{V}}_{\mathbf{u}}$. The transformed $\hat{\mathbf{u}}$ is obtaining as

$$\hat{\mathbf{u}}^S = \hat{\mathbf{V}}_{\mathbf{u}}^{-1/2}\hat{\mathbf{u}}.$$

The covariance matrix of $\hat{\mathbf{u}}^S$ is $Var(\hat{\mathbf{u}}^S) = \mathbf{Q}_{\mathbf{u}}\mathbf{Q}_{\mathbf{u}}'$, which is close to an identity matrix. The explicit expression of $\hat{\mathbf{u}}^S$ is

$$\hat{\mathbf{u}}^S = \mathbf{Q}_{\mathbf{u}}\Delta_{\mathbf{u}}^{-1/2}\mathbf{Q}_{\mathbf{u}}'\hat{\mathbf{u}},$$

where $\mathbf{Q}_{\mathbf{u}}'\hat{\mathbf{u}}$ contains the coordinates of $\hat{\mathbf{u}}$ in its principal components. The coordinates are uncorrelated with the covariance matrix $\Delta_{\mathbf{u}}$. Then, we multiply the coordinates by $\Delta_{\mathbf{u}}^{-1/2}$ in order to force them to have unit variance. Finally, we multiply the standardized vector in the space of the principal components by $\mathbf{Q}_{\mathbf{u}}$. This procedure assures that the standardized vector returns to the original space. Thus, the transformed vector $\hat{\mathbf{u}}^S$ contains the coordinates of the vector $\Delta_{\mathbf{u}}^{-1/2}\mathbf{Q}_{\mathbf{u}}'\hat{\mathbf{u}}$, with standard elements, in the original space.

The nonparametric bootstrap procedure works by replacing steps (2) and (3) of the parametric bootstrap by the new steps (2') and (3') as:

(step 2') Calculate predictors of $\mathbf{v}$ and $\mathbf{u}$ using the estimates $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$ and $\hat{\beta} = \tilde{\beta}(\hat{\omega})$ obtained in step (1) in the following way:

$$\hat{\mathbf{v}} = \mathbf{G}(\hat{\omega})\mathbf{V}(\hat{\omega})^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}), \quad \hat{\mathbf{u}} = (\mathbf{I} - \hat{\rho}\mathbf{W})\hat{\mathbf{v}} = (\hat{u}_1, \ldots, \hat{u}_D)'.$$

Make $\hat{\mathbf{u}}^S = \hat{\mathbf{V}}_{\mathbf{u}}^{-1/2}\hat{\mathbf{u}} = (\hat{u}_1^S, \ldots, \hat{u}_D^S)'$, where $\hat{\mathbf{V}}_{\mathbf{u}}^{1/2}$ is the square root of the generalized inverse of $\hat{\mathbf{V}}_{\mathbf{u}}$ obtained by the spectral decomposition. Then, re-scale the elements $\hat{u}_d^S$ to obtain elements with sample mean exactly equal to zero and sample variance $\hat{\sigma}_u^2$. The transformation is

$$\hat{u}_d^{SS} = \frac{\hat{\sigma}_u(\hat{u}_d^S - D^{-1}\sum_{\ell=1}^D \hat{u}_\ell^S)}{\sqrt{D^{-1}\sum_{k=1}^D(\hat{u}_k^S - D^{-1}\sum_{\ell=1}^D \hat{u}_\ell^S)^2}}, \quad d = 1, \ldots, D.$$

Build the vector $\mathbf{u}^* = (u_1^*, \ldots, u_D^*)'$. Its elements are obtained by extracting a simple random sample with replacement of size D from the set $\{\hat{u}_1^{SS}, \ldots, \hat{u}_D^{SS}\}$. Proceed by obtaining $\mathbf{v}^* = (\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{u}^*$ and calculating the bootstrap quantity of interest $\theta^* = \mathbf{X}\hat{\beta} + \mathbf{v}^* = (\theta_1^*, \ldots, \theta_D^*)'$

(step 3') Follow by computing the vector of residuals $\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}\hat{\beta} - \hat{\mathbf{v}} = (\hat{r}_1, \ldots, \hat{r}_D)'$ and proceed with its standardization $\hat{\mathbf{r}}^S = \hat{\mathbf{V}}_{\mathbf{r}}^{-1/2}\hat{\mathbf{r}} = (\hat{r}_1^S, \ldots, \hat{r}_D^S)'$, where $\hat{\mathbf{V}}_{\mathbf{r}} = \Psi\mathbf{P}(\hat{\omega})\Psi$ is the estimated covariance matrix and $\hat{\mathbf{V}}_{\mathbf{r}}^{-1/2}$ is a root square of the generalized inverse derived from the spectral decomposition of $\hat{\mathbf{V}}_{\mathbf{r}}$. Once more, re-standardize these values in the following way

$$\hat{r}_d^{SS} = \frac{\hat{r}_d^S - D^{-1}\sum_{\ell=1}^{D}\hat{r}_\ell^S}{\sqrt{D^{-1}\sum_{k=1}^{D}(\hat{r}_k^S - D^{-1}\sum_{\ell=1}^{D}\hat{r}_\ell^S)^2}}, \quad d = 1, \ldots, D.$$

Finally, build $\mathbf{r}^* = (r_1^*, \ldots, r_D^*)'$ by extracting a simple random sample with replacement of size D from the set $\{\hat{r}_1^{SS}, \ldots, \hat{r}_D^{SS}\}$ and let $\mathbf{e}^* = (e_1^*, \ldots, e_D^*)'$, where $e_d^* = \psi_d^{1/2} r_d^*$, $d = 1, \ldots, D$.

This procedure leads to naive and bias-corrected nonparametric bootstrap estimators analogous to (4.18) and (6.13), which are denoted as $mse^{naNPB}[\tilde{\theta}_d(\hat{\omega})]$ and $mse^{bcNPB}[\tilde{\theta}_d(\hat{\omega})]$, respectively.

When the normality assumption is suspected to be violated either for the random effects or for the errors but not for both, it is possible to combine step (2') with (3), or step (2) with (3') of the two bootstrap procedures. This comes out in a semiparametric bootstrap that avoids the normality assumption on the desired component of the model.

4.7 Simulation study

In this section we describe some simulation experiments carried out with the following purposes: (a) to check whether taking into account the spatial correlation between small areas in the model improves the precision of small area estimators; (b) to study the small-sample behavior of the different MSE estimators introduced in this chapter, for different values of the spatial correlation parameter ρ and for different patterns of sampling variances ψ_d; (c) to analyze the robustness of the proposed bootstrap procedures to non-normality of the random effects and errors.

The experiments are based on a real population, the map of the $D = 287$ municipalities (small areas) of Tuscany. We considered a model with $p = 2$, that is, one explanatory variable and a constant, with an $D \times 2$ design matrix $\mathbf{X} = [\mathbf{1}_D\ \mathbf{x}]$, where $\mathbf{1}_D$ is a column vector of ones of size D and $\mathbf{x} = (x_1, \ldots, x_D)'$ contains the values of the explanatory variable. These values x_d were generated from a uniform distribution in the interval $(0, 1)$. The true model coefficients were $\beta = (1, 2)'$, the random effects variance $\sigma_u^2 = 1$ and the spatial correlation parameter $\rho \in \{0.25, 0.5, 0.75\}$. The matrix of sampling variances $\psi = \text{diag}(\psi_1, \ldots, \psi_D)$ was taken as $\psi_d = 0.7$ for $1 \le d \le 60$; $\psi_d = 0.6$ for $61 \le d \le 120$; $\psi_d = 0.5$ for $121 \le d \le 180$; $\psi_d = 0.4$ for $181 \le d \le 240$ and finally $\psi_d = 0.3$ for $241 \le d \le 287$ (see Datta et al. (2005)). The $D \times D$ row-standardized proximity matrix $\mathbf{W}$ was obtained from the neighborhood structure of the municipalities in Tuscany. This matrix was kept constant for all simulations. We considered three possible probability distributions for the random area effects and errors, namely Normal, Gumbel and Student t distribution with 6 degrees of freedom, all standardized to have zero mean and unit variance. The last two distributions represent two different sources of discrepancy to normality, since the Gumbel distribution is skewed and the Student t has heavy tails.

Taking into account the simulation results of Molina et al. (2008) on the comparison of fitting methods for the Spatial Fay-Herriot model, we have decided to use only REML method in these simulations.

Concerning target (a), $L = 1000$ Monte Carlo data sets were generated as described before, taking Normal distribution for the random effects and errors. Then two models were fitted to each data set: the spatial model (4.4)-(4.5), and the non-spatial model obtained by assuming that in model (4.4), the vector of random effects $\mathbf{v} = (v_1,\ldots,v_D)'$ has independent and identically distributed elements v_d, with zero mean and variance σ_u^2. Figures 4.1 and 4.2 plot the empirical values of the mean squared errors of the Spatial EBLUP obtained from the former model, and the NonSpatial EBLUP resulting from the latter model, for the $D = 287$ small areas, for $\rho = 0.75$ and $\rho = 0.25$, respectively. The piecewise decreasing shape that we observe in the level of these two figures is due to the decreasing patterns of sampling variances ψ_d. Figure 4.1 shows that ignoring the spatial correlation structure of small areas leads to an increase in the MSE. However, this increase is smaller for areas with smaller sampling variances and in the case of weak spatial correlation, see Figure 4.2 for $\rho = 0.25$. This last figure also suggests that modelling the spatial correlation seems to be convenient even when this correlation is weak, since there is no loss in efficiency.

Target (b) deals with comparing the analytical estimators of the MSE given in (4.15) and (4.17) with the bootstrap estimators. For this, $L = 250$ Monte Carlo data sets were generated, and for each data set, we calculated the different MSE estimators introduced in this paper, namely, the two analytical estimators $mse^{PR}[\tilde{\theta}_d(\hat{\omega})]$ and $mse^{SSK}[\tilde{\theta}_d(\hat{\omega})]$, the two estimators obtained by parametric bootstrap $mse^{naPB}[\tilde{\theta}_d(\hat{\omega})]$ and $mse^{bcPB}[\tilde{\theta}_d(\hat{\omega})]$ (drawing from the Normal distribution), and the two estimators based on nonparametric bootstrap $mse^{naNPB}[\tilde{\theta}_d(\hat{\omega})]$ and $mse^{bcNPB}[\tilde{\theta}_d(\hat{\omega})]$. The bootstrap procedures were applied with $B = 250$ replicates. The empirical values of the MSE, which are the reference values for comparison, were computed previously with 10000 Monte Carlo replicates to ensure better accuracy. As output of simulations, we obtained for each small area d, the mean mse_d^a and the mean squared error E_d^a over Monte Carlo samples of each estimator $mse^a[\tilde{\theta}_d(\hat{\omega})]$, for $a \in A$ with $A = \{PR, SSK, naPB, bcPB, naNPB, bcNPB\}$. Then, to summarize the results over small areas, we computed the average over the $D = 274$ small areas of the relative bias and of the relative root mean squared error of each MSE estimator, as

$$\mathrm{ARB}^a = \frac{1}{D}\sum_{d=1}^{D}\left(\frac{mse_d^a}{MSE_d} - 1\right), \quad \mathrm{ARE}^a = \frac{1}{D}\sum_{d=1}^{D}\frac{\sqrt{E_d^a}}{MSE_d}, \qquad a \in A,$$

where MSE_d stands for the empirical value of the true $MSE[\tilde{\theta}_d(\hat{\omega})]$. Tables 4.1 and 4.2 report the resulting percent values of ARE^a and ARB^a respectively. In terms of relative root mean squared error, the bcNPB estimator does not behave bad in comparison with the other estimators in any case. As expected, under Normal distribution, the analytical PR estimator is less biased for $\rho = 0.25$ and $\rho = 0.75$, although not for $\rho = 0.5$. In this last case, the parametric bootstrap gets a better ARB. For Gumbel distribution, the bcNPB is less biased than the rest of estimators for $\rho = 0.75$ and in all cases in has a moderate ARE. In the case of the Student t_6, the bcNPB shows less ARB for $\rho = 0.25$, and similar to the best estimator for the other two values of ρ, and has the smaller values of ARE for $\rho = 0.25$ and $\rho = 0.5$. Finally, results indicate that the bias correction is necessary.

4.8 Conclusions

This chapter considers a Fay-Herriot model with correlated random area effects according to a simultaneously autoregressive process. It revises the different analytical estimators of the MSE of the Spatial EBLUP and proposes two new estimators based on both a parametric and a nonparametric bootstrap.

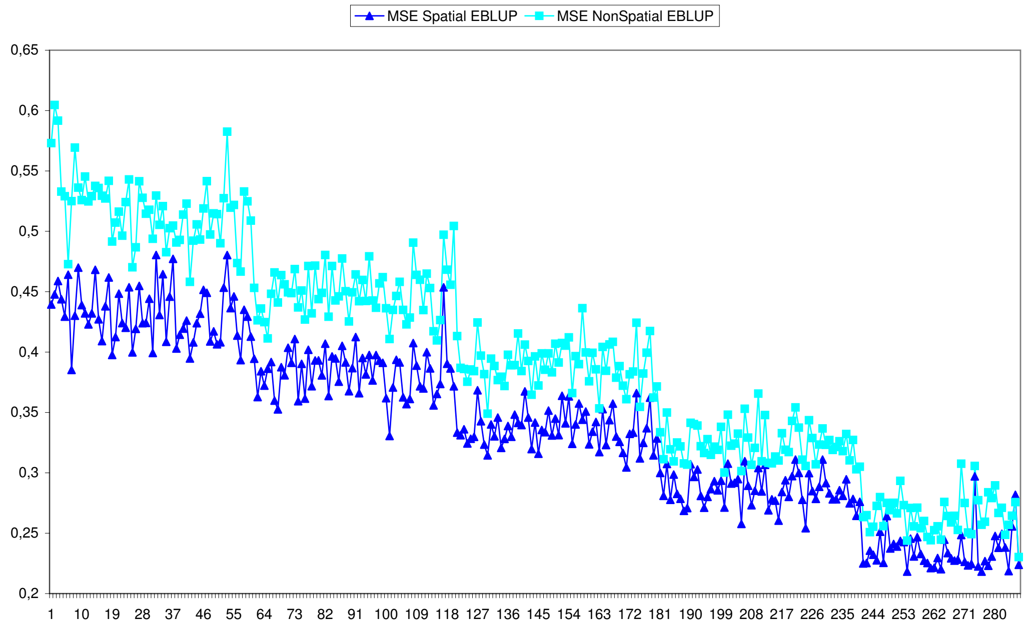


Figure 4.1: Empirical MSE of the Spatial EBLUP and the NonSpatial EBLUP for the $D = 287$ small areas, for $\rho = 0.75$.

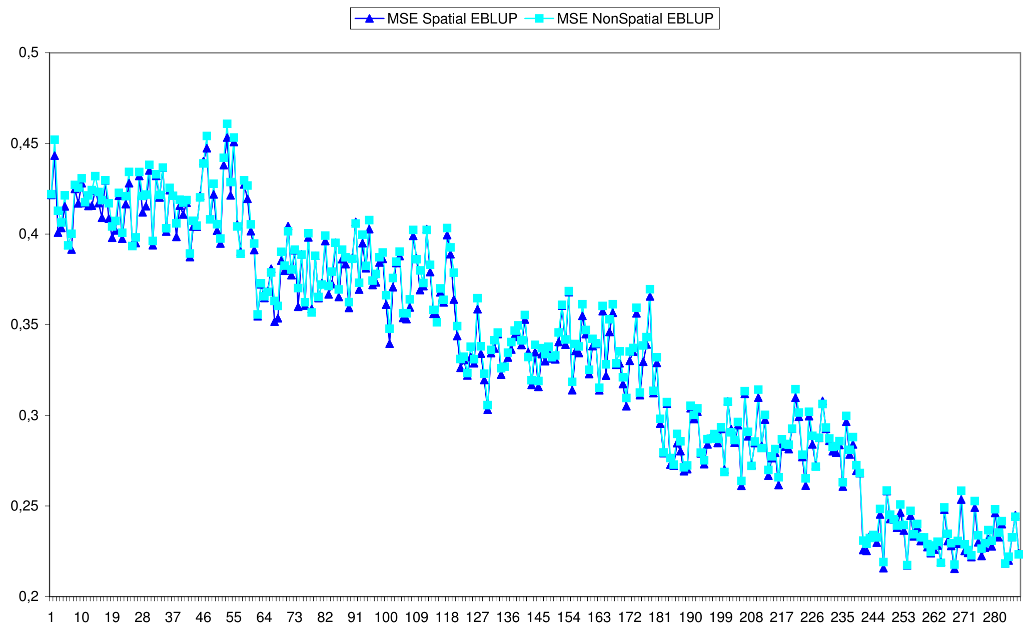


Figure 4.2: Empirical MSE of the Spatial EBLUP and the NonSpatial EBLUP for the $D = 287$ small areas, for $\rho = 0.25$.

Table 4.1: ARE ($\times$100) of the different MSE estimators, when data are simulated from standard Normal, Gumbel and Student t_6 distributions.

Normal

ρ	PR	SSK	naPB	bcPB	naNPB	bcNPB
0.25	4.15	4.17	9.79	4.02	9.91	4.08
0.5	4.23	4.23	9.90	4.22	9.80	4.12
0.75	4.20	4.20	9.80	4.20	9.98	4.57

Gumbel

ρ	PR	SSK	naPB	bcPB	naNPB	bcNPB
0.25	5.15	5.18	10.44	5.41	11.50	5.28
0.5	5.27	5.28	10.53	5.56	11.93	6.05
0.75	5.17	5.16	10.68	5.78	11.74	5.75

Student t_6

ρ	PR	SSK	naPB	bcPB	naNPB	bcNPB
0.25	6.15	6.23	10.64	5.68	11.97	5.55
0.5	6.18	6.21	10.68	5.85	11.87	5.85
0.75	6.07	6.08	10.58	5.70	12.37	7.23

Our simulation experiments supported the results of Molina et al. (2008), in the sense that the estimator derived from the nonparametric bootstrap performed well in terms of average relative error and bias, as compared with the other estimators, under Gumbel and Student t_6 distributions, and it performed acceptably well also under Normal distribution. Thus, this method is expected to be more reliable when the distribution is not exactly normal.

In the simulations of Section 4.7 there were not municipalities without sample data. For and area d without sample data but for which the values of the covariates at the area-level are available from any other data source, a possible estimator is $\tilde{\theta}_d(\hat{\omega}) = \mathbf{x}_d\tilde{\beta}(\hat{\omega})$. Estimation of the MSE for these areas either by analytical estimators based on Taylor expansion or using bootstrap should not be a problem.

In the SAR process, the proximity matrix must be specified in advance. The structure of this matrix can be determined by specifying a neighborhood rule or a distance function between areas. These distances can be either related to physical distances, or to other socioeconomic variables. However, the best specification of this matrix for a particular problem is not clear and this issue deserves deep investigation. Several specifications for the proximity matrix between Spanish provinces are studied and compared in the next chapter.

Table 4.2: ARB (×100) of the different MSE and g_{3d} estimators when data are simulated from standard Normal, Gumbel and Student t_6 distributions.

Normal

ρ	PR	SSK	naPB	bcPB	naNPB	bcNPB
0.25	0.27	-0.04	-0.88	-0.29	-0.15	0.44
0.5	0.26	0.07	-0.28	0.20	-0.63	-0.19
0.75	-0.16	-0.24	-1.08	-0.79	-1.26	-0.94

Gumbel

ρ	PR	SSK	naPB	bcPB	naNPB	bcNPB
0.25	0.07	-0.24	-1.41	-0.82	-1.23	-0.48
0.5	0.20	-0.01	-0.49	-0.02	-1.65	-1.06
0.75	0.26	0.17	-0.57	-0.31	-0.40	0.03

Student t_6

ρ	PR	SSK	naPB	bcPB	naNPB	bcNPB
0.25	-0.63	-0.95	-1.37	-0.80	-0.61	0.20
0.5	-0.65	-0.85	-1.39	-0.89	-0.27	0.30
0.75	-1.10	-1.18	-1.93	-1.67	-1.58	-1.13

Chapter 5

Proximities based on semi-metrics for socioeconomic functional data

5.1 Introduction

The specification of the weight matrix **W** of (4.5) introduced in Section 4.2 is one of the challenges in analyzing spatial data. The literature on spatial econometrics and statistics specifies mainly two ways of modeling this matrix. The first consists of distances among units, it is flexible since the spatial effects are different from different distances and it is determined in advance by Kakamu (2005). This matrix is often call typicality matrix, and is defined in the following way: if two small areas are neighbours the corresponding entry in **W** is 1, and 0 otherwise. Therefore, matrix **W** contains the geographic dependence between small areas, an information that can be relevant in the prediction of physical or environmental variables, such as temperature or pollution, but not in describing latent variables, such as poverty. An alternative approach is to estimate the weights together with the model's parameters. In this field the paper by LeSage and Pace (2007) became a reference. They proposed the matrix exponential spatial specification (MESS) procedure that replaces the geometric pattern of decay in the spatial autoregressive model (SAR) by an exponential decay. Among the MESS advantages we can emphasize the simplification of the log-likelihood and the consequent simplification of the Bayesian estimation of the model. The procedure produces estimates and inferences similar to those from the conventional spatial autoregressive models. Nevertheless, it can be applied only to the SAR model. Kakamu (2005) proposed a distance functional weight matrix model that is suitable to be applied to the SAR model as well to the spatial error model (SEM), however, it is more time consuming.

Our approach belongs to the first branch of the literature, and uses the information on socioeconomic variables to estimate the weight matrix, that we denote hereafter proximity matrix. It is not restrictive to the use of socioeconomic variables, one can use any set of information that considers relevant for the estimation of the variable of interest. In our particular case, we use the unemployment rate and the illiteracy rate because our main interest is to estimate the poverty incidences in Spanish provinces. Our purpose is to construct a proximity matrix that considers as neighbors those provinces whose socioeconomic information is similar. The variables chosen are related to poverty and are not included in the regressor set. We used only two variables since data by provinces are not very abundant and are difficult to obtain.

In particular, we develop two alternative methods for constructing matrix $\mathbf{W}$ instead of considering the classical typicality matrix. In both proposals the proximities between small areas are computed attending to some socioeconomic information. The first one, described in Section 5.2, consists in constructing the proximity matrix using classical multivariate analysis, whereas the second approach, described in Section 5.3, uses functional data techniques. In both cases $\mathbf{W}$ is obtained from a matrix $\mathbf{D}^{(2)} = (\delta_{ij}^{(2)})$, containing the squared distances between small areas. In Section 5.4 we draw a smooth-parametric Bootstrap scheme in order to study the gain obtained in the prediction of the poverty level, in the sense of mean squared prediction error, when considering the classical, multivariate and functional approaches in the computation of matrix $\mathbf{W}$. The empirical results show that the functional method performs better than the multivariate and classical ones, since it gives more accurate predictions of the poverty level.

5.2 The multivariate approach

Let $\mathbf{X}$ be a $D \times p$ matrix containing the information of p socioeconomic (continuous) variables observed on D small areas, at a fixed instant of time. Given two small areas i and j, the entries in matrix $\mathbf{D}^{(2)}$ are obtained from Mahalanobis distance as:

$$\delta_{ij}^{(2)} = (\mathbf{x}_i - \mathbf{x}_j)'\mathbf{S}^{-1}(\mathbf{x}_i - \mathbf{x}_j),$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ denotes the i-th row of $\mathbf{X}$ (analogous for $\mathbf{x}_j$) and $\mathbf{S}$ is the covariance matrix of $\mathbf{X}$,

$$\mathbf{S} = \frac{1}{D}\mathbf{X}'\left(\mathbf{I} - \frac{1}{D}\mathbf{1}\mathbf{1}'\right)\mathbf{X},$$

where $\mathbf{I}$ is the identity matrix of size D and $\mathbf{1}$ is the vector of ones. Since $\mathbf{W}$ is a proximity (or similarity) matrix it can be obtained as

$$\mathbf{W} = \mathbf{1}\mathbf{1}' - \frac{1}{2}\mathbf{D}^{(2)}.$$

Finally, one can consider $\mathbf{W}$ in row standardized form, if necessary.

5.3 The functional approach

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be $D \times J$ matrices containing two socioeconomic time series observed on D small areas through a fixed period of time $T = \{t_1, t_2, \ldots, t_J\}$. The historical socioeconomic information of the D small areas contained in the rows of each matrix $\mathbf{X}_\ell$, $\ell = 1, 2$, can be seen as a set of D curves, $\mathbf{x}_{\ell 1}, \ldots, \mathbf{x}_{\ell D}$, that is called a functional dataset. Ferraty and Vieu (2006) proposed several methodologies for obtaining semi-metrics from a functional dataset. In this chapter we explore *Functional Principal Components Analysis (FPCA)*, a technique that is well adapted for rough curves, and for each $\mathbf{X}_\ell$, $\ell = 1, 2$ we obtain the corresponding functional semi-metric. Then, following Cuadras and Fortiana (1998), we construct matrix $\mathbf{D}^{(2)}$ combining the two functional semi-metrics without including redundant information.

5.3.1 Functional PCA

In the context of multivariate analysis, the classical Principal Components Analysis is considered as a useful tool for displaying data in a reduced dimensional space. More recently, the PCA methods were extended to functional data and used for many different statistical purposes. In particular, Ferraty and Vieu (2006) propose FPCA as a tool for computing proximities between curves in a reduced dimensional space. In the following we describe this technique for a general functional dataset.

Let $\mathcal{X}$ be a random element of a functional space (typically a real function from $T = [a,b] \subseteq \mathbb{R}$ to $\mathbb{R}$) and let $\mathcal{X}_1,\dots,\mathcal{X}_D$ be D independent and identically distributed observations from $\mathcal{X}$. As long as $E(\int_T \mathcal{X}^2(t)\,dt) < \infty$, the FPCA of $\mathcal{X}$ allows us to obtain the following expansion of Dauxois et al. (1982)

$$\mathcal{X} = \sum_{k=1}^{\infty} \left(\int_T \mathcal{X}(t)\, v_k(t)\, dt \right) v_k, \tag{5.1}$$

where $\{v_k\}_{k\geq 1}$ is the sequence of orthonormal eigenfunctions of the covariance operator

$$\Gamma_{\mathcal{X}}(s,t) = E(\mathcal{X}(s)\,\mathcal{X}(t)),$$

associated with the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots$. Now, let

$$\widetilde{\mathcal{X}}^{(q)} = \sum_{k=1}^{q} \left(\int_T \mathcal{X}(t)\, v_k(t)\, dt \right) v_k$$

be the truncated version of (5.1). The main interest of such a decomposition is that this truncated version is minimizing $E(\int_T (\mathcal{X}(t) - P_q\mathcal{X}(t))^2 dt)$ over all projections P_q of $\mathcal{X}$ into q-dimensional spaces. Thus, we can define a parametrized class of semi-norms from the classical L^2-norm in the following way:

$$\|\mathcal{X}\|_q^{PCA} = \sqrt{\int_T (\widetilde{\mathcal{X}}^{(q)}(t))^2\, dt} = \sqrt{\sum_{k=1}^{q} \left(\int_T \mathcal{X}(t)\, v_k(t)\, dt \right)^2},$$

which leads to the following parametrized family of semi-metrics:

$$d_q^{PCA}(\mathcal{X}_i,\mathcal{X}) = \sqrt{\sum_{k=1}^{q} \left(\int_T (\mathcal{X}_i(t) - \mathcal{X}(t))\, v_k(t)\, dt \right)^2}.$$

Here, q is not really a smoothing parameter but rather a tuning parameter indicating the resolution level at which the problem is considered. Note that in practice, $\Gamma_{\mathcal{X}}$ is unknown, and also the v_k's, but the covariance operator can be well approximated by its empirical version

$$\Gamma_{\mathcal{X}}^D(s,t) = \frac{1}{D} \sum_{i=1}^{D} \mathcal{X}_i(s)\mathcal{X}_i(t),$$

and the eigenfunctions of $\Gamma_{\mathcal{X}}^D$ are consistent estimators of those of $\Gamma_{\mathcal{X}}$ (see Cardot et al. (1999)). Indeed we never observe directly $\{\mathcal{X}_i = \{\mathcal{X}_i(t), t \in T\}\}_{i=1,\dots,D}$ but only a discretized version $\{\mathbf{x}_i = (\mathcal{X}_i(t_1),\dots,\mathcal{X}_i(t_J))\}_{i=1,\dots,D}$ (notice that this is implicitly assuming that the data are balanced, which means that all units are measured at the same points). So, from a practical point of view, according to Castro et al. (1986) we can approximate the integral in the following way

$$\int_T (\mathcal{X}_i(t) - \mathcal{X}(t))\, v_k(t)\, dt \approx \sum_{j=1}^{J} w_j\, (\mathcal{X}_i(t_j) - \mathcal{X}(t_j))\, v_k(t_j),$$

where $w_1,\ldots,w_J$ are quadrature weights which define the approximate integration. To fix ideas, note that the standard choice could be $w_j = t_j - t_{j-1}$. If we have two discretized curves $\mathbf{x}_i$ and $\mathbf{x}_{i'}$, the quantity $d_q^{PCA}(\chi_i,\chi_{i'})$ will be approximated by its empirical version:

$$d_q^{PCA}(\mathbf{x}_i,\mathbf{x}_{i'}) = \sqrt{\sum_{k=1}^{q}\left(\sum_{j=1}^{J} w_j(\chi_i(t_j)-\chi_{i'}(t_j))v_{jk}\right)}, \tag{5.2}$$

where $\mathbf{v}_k = (v_{1k},\ldots,v_{Jk})'$, $k=1,\ldots,q$, are the $\Delta_\mathbf{w}$-orthonormal eigenvectors of the covariance matrix $(\Delta_\mathbf{w} = \mathrm{diag}(w_1,\ldots,w_J))$

$$\Gamma^n\,\Delta_\mathbf{w} = \frac{1}{D}\sum_{i=1}^{D}\mathbf{x}_i'\,\mathbf{x}_i\,\Delta_\mathbf{w},$$

associated with the eigenvalues $\lambda_{1,D}\geq\lambda_{2,D}\geq\ldots\geq\lambda_{q,D}$. Note that $d_q^{PCA}(\mathbf{x}_i,\mathbf{x}_{i'})$ is close to $d_q^{PCA}(\chi_i,\chi_{i'})$ as soon as the grid $(t_1,\ldots,t_J)$ is sufficiently fine.

5.3.2 Related metric scaling applied to functional semi-metrics

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be $D\times J$ matrices containing two socioeconomic time series observed on D small areas through a fixed period of time $T=\{t_1,t_2,\ldots,t_J\}$. Using expression (5.2) and for a given $q\geq 1$ we can obtain matrices $\mathbf{D}_{1,q}^{(2)}$ and $\mathbf{D}_{2,q}^{(2)}$, which contain the distances between the D small areas, according to these two socioeconomic indicators. Since both socioeconomic indicators are referred to the same set of D small areas, the two functional semi-metrics $\mathbf{D}_{1,q}^{(2)}$ and $\mathbf{D}_{2,q}^{(2)}$ may contain redundant information. Cuadras and Fortiana (1998) proposed a technique, called related metric scaling, that is an extension of metric scaling whose aim is to join several distance matrices referred to the same group of individuals, taking into consideration the possibility of redundant information. See Cuadras and Fortiana (1998) and also Cuadras and Fortiana (1995) for the details.

In the following we apply their methodology in the case of functional semi-metrics. We start by imposing that the two marginal distances $\mathbf{D}_{1,q}^{(2)}$ and $\mathbf{D}_{2,q}^{(2)}$ have the same geometric variability, that is

$$\frac{1}{D^2}\sum_{i,j=1}^{D}\delta_{1,q}^2(i,j) = \frac{1}{D^2}\sum_{i,j=1}^{D}\delta_{2,q}^2(i,j).$$

Note that this condition can always be assumed to hold, since multiplying one of the marginal distances by an appropriate constant amounts to a change of measurement unit. Then, for each marginal distance $\mathbf{D}_{i,q}^{(2)}$, $i=1,2$, we consider the associated inner product matrix

$$\mathbf{G}_i = -\frac{1}{2}\mathbf{H}\,\mathbf{D}_{i,q}^{(2)}\,\mathbf{H},$$

where $\mathbf{H} = \mathbf{I}-(1/D)\,\mathbf{1}\mathbf{1}'$ is the centering matrix, and we define the inner product matrix associated to the joint distance $\mathbf{D}^{(2)}$ as

$$\mathbf{G} = \mathbf{G}_1+\mathbf{G}_2-\frac{1}{2}\left(\mathbf{G}_1^{1/2}\mathbf{G}_2^{1/2}+\mathbf{G}_2^{1/2}\mathbf{G}_1^{1/2}\right),$$

where $\mathbf{G}_i^{1/2} = \mathbf{U}_i \Lambda_i^{1/2} \mathbf{U}_i'$, $\mathbf{U}_i$ is the $D \times k$ matrix containing the orthonormal eigenvectors of the symmetric matrix $\mathbf{G}_i$ corresponding to the first eigenvalues, ordered as $\lambda_1 \geq \ldots \geq \lambda_k > 0$, $k \leq D-1$, and $\Lambda_i = \text{diag}(\lambda_1, \ldots, \lambda_k)$. Finally, the joint distance $\mathbf{D}^{(2)}$ is obtained form the inner product matrix as

$$\mathbf{D}^{(2)} = \mathbf{g}\,\mathbf{1}' + \mathbf{1}\,\mathbf{g}' - 2\,\mathbf{G},$$

where $\mathbf{g} = \text{diag}(\mathbf{G})$. Proceeding as in Section 5.2, the proximity matrix $\mathbf{W}$ is obtained as

$$\mathbf{W} = \mathbf{1}\,\mathbf{1}' - \frac{1}{2}\mathbf{D}^{(2)}.$$

5.4 Simulation study

In this Section we study the gain obtained in the prediction of direct estimators of the FGT poverty measures (for $\alpha = 0$), in the sense of mean squared error of small area predictors, when considering three different approaches in the computation of matrix $\mathbf{W}$ in (4.5). The first one is the classical approach, where $\mathbf{W}$ is a typicality matrix, whereas the second and third approaches consist in implementing the techniques described in Sections 5.2 and 5.3, respectively.

We start by describing the data to be used in the model described in Section 4.2. They consist of official data from the Spanish Survey of Income and Living Conditions corresponding to year 2006 for $D = 51$ Spanish provinces (the small areas). The response variable is the direct estimator of the FGT poverty measure (for $\alpha = 0$), that is the proportion of poor in the area. The auxiliary covariates are the intercept and the following proportions (in the area) of Spanish people, people of ages from 16 to 24, from 25 to 49, from 50 to 64, equal or greater than 65, people with no studies up to primary studies, Graduate people, employees, unemployed people, inactive people.

We have selected from the *Instituto Nacional de Estadística* website (`http://www.ine.es`), the more relevant socioeconomic variables related with poverty, being the unemployment rate and share of illiterate population over 16 years old. These variables have been measured in the $D = 51$ provinces from 1991 to 2005 ($J = 15$ years). Therefore, in practice we have two matrices, $\mathbf{X}_1$ and $\mathbf{X}_2$ of size 51×15.

In order to compute matrix $\mathbf{W}$ with the multivariate approach of Section 5.2, we only consider the information contained in J-th columns of $\mathbf{X}_1$ and $\mathbf{X}_2$, which leads to a matrix of size 51×2. We call $\mathbf{W}_M$ the proximity matrix computed with the methodology described in Section 5.2.

To compute matrix $\mathbf{W}$ using the functional approach of Section 5.3, we have obtained two semi-metrics $\mathbf{D}_{1,q}^{(2)}$ and $\mathbf{D}_{2,q}^{(2)}$, one for each data set (see Ferraty and Vieu (2006)), for $q = 4$ functional principal components, since $q = 4$ is enough to collect the most part of the observed variability. Finally, in order to obtain a square matrix of joint distances from the previous two, we have used the related metric scaling technique, introduced by Cuadras and Fortiana (1998), which provides a joint metric from different metrics on the same individuals, taking into account the possible redundant information that can be added simply by adding distance matrices. We call $\mathbf{W}_F$ the proximity matrix computed in this way.

We call $\mathbf{W}_T$ the typicality matrix between the Spanish provinces, whose entries are 1 if the corresponding two provinces are neighbours or 0 otherwise.

We undertake a simulation study to compare the performance of the three matrices $\mathbf{W}_T$, $\mathbf{W}_M$ and $\mathbf{W}_F$. Therefore, we apply a resampling approach and obtain 500 samples from the original database, by using a parametric Bootstrap technique. Under this approach, it is added a small amount of normally distributed random noise to each resampled observation. This is equivalent to sample from a kernel density estimate of the data.

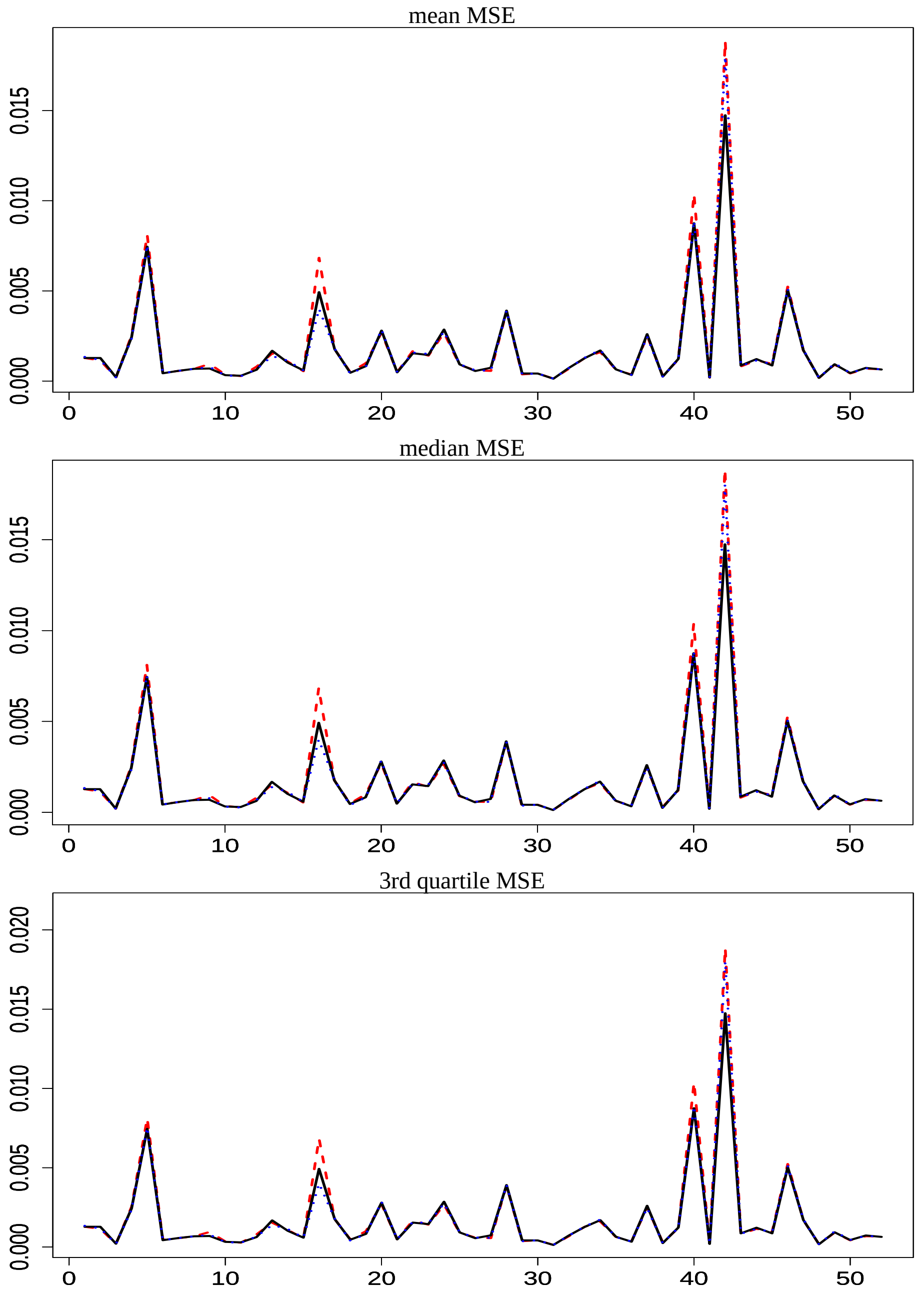
We compute the estimators of the model in the original sample and in each of the resampled data considering each of the matrices $\mathbf{W}_T$, $\mathbf{W}_M$ and $\mathbf{W}_F$. As a measure of errors we compute the corresponding sum of squares among these estimators. Table 5.1 contains the global mean, median and third quartile of MSE obtained using matrices $\mathbf{W}_T$, $\mathbf{W}_M$ and $\mathbf{W}_F$.

Table 5.1: Global mean, median and third quartile of the MSE obtained using three methodologies (classical, multivariate and functional) in computing matrix $\mathbf{W}$.

	mean MSE	median MSE	3rd quartile MSE
$\mathbf{W}_T$	0.096991	0.082082	0.127506
$\mathbf{W}_M$	0.090955	0.076425	0.119072
$\mathbf{W}_F$	0.088751	0.073952	0.117385

Figure 5.1 contains the plots of the mean, median and third quartile of the MSE for each Spanish province. Both from Table 5.1 and Figure 5.1 we can see that the best performance is obtained with $\mathbf{W}_F$, that is using functional data analysis approach.

Figure 5.1: Comparisons of mean MSE, median MSE and third quartile MSE, for each Spanish province, computed on 500 samples from the original database by smooth-parametric Bootstrap, using matrices W_F (black solid line), W_T (red dashed line) and W_M (blue dotted line).



Chapter 6

Semiparametric Fay-Herriot model using penalized splines

Traditional Fay-Herriot small area estimation models are based on linear mixed models, characterized by random area effects which allow for between area heterogeneity apart from that explained by the auxiliary variables (see Rao (2003)). These models however are based on the hypothesis of a linear relationship between the variable of interest and the covariates, hypothesis that can represent a serious restriction in many real data applications. Furthermore, traditional linear mixed models do not handle spatial proximity effects between the areas, an important feature in environmental studies where detailed geo-referenced information for the units of analysis is usually available. Indeed, in recent years extensions to random effects models have been proposed to allow for spatially correlated random area effects taking into account the information provided by neighboring areas (see Petrucci and Salvati (2006) and Pratesi and Salvati (2009)), but these models still rely on the linearity assumption.

Here we present a semiparametric version of the basic Fay-Herriot model that is based on P-splines and can also handle situations where the functional form of the relationship between the variable of interest and the covariates cannot be specified a priori (Giusti et al., in preparation). This is often the case when the data are supposed to be affected by spatial proximity effects. In these cases P-spline bivariate smoothing can easily introduce spatial effects in the area level model. Opsomer et al. (2008) proposed a similar small area model based on P-splines but under the assumption that all the data are available at the unit level, and this can be a restriction in some situations.

6.1 Estimation of small area means

Let θ be the $d \times 1$ vector of the parameter of inferential interest (small area total y_d, small area mean $\bar{y}_d$ with $d = 1, ..., D$) and assume that the $d \times 1$ vector of the direct estimator $\hat{\theta}$ is available and design unbiased. Denote the corresponding $d \times p$ matrix of the area level auxiliary variables by $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$. Fay and Herriot (1979) introduced a model that can be expressed as:

$$\hat{\theta} = \mathbf{X}\alpha + \mathbf{Z}\mathbf{u} + \varepsilon. \tag{6.1}$$

Here $\mathbf{u}$ is $m \times 1$ vector of independent and identically distributed random variables with mean $\mathbf{0}$ and $m \times m$ variance matrix $\Sigma_u = \sigma_u^2 \mathbf{I}_m$, $\mathbf{Z}$ is a $m \times m$ matrix of known positive constants, ε is the $m \times 1$ vector of independent sampling errors with mean $\mathbf{0}$ and known diagonal variance matrix $\mathbf{R} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ and α is the $q \times 1$ vector of regression parameters. The Fay-Herriot model is a general linear mixed model with diagonal covariance structure $\Sigma(\sigma_u^2) = \mathbf{Z}\Sigma_u\mathbf{Z}^T + \mathbf{R}$.

The Fay-Herriot model produces reliable small area estimates by combining the design model and the regression model and then borrowing strength from other domains. It assumes that the direct survey estimators are linear function of the covariates. When this assumption fails down, the Fay-Herriot model can lead to biased estimators of the small area parameters. A semiparametric specification of the Fay-Herriot model, which allows non linearities in the relationship between $\hat{\theta}$ and the auxiliary variables $\mathbf{X}$, can be obtained by penalized-splines. This approach may have significant advantages compared to the linear Fay-Herriot model.

A semiparametric model with one covariate x_1 can be written as $\tilde{m}(\mathbf{x}_1)$, where the function $\tilde{m}(\cdot)$ is unknown, but assumed to be sufficiently well approximated by the function

$$m(\mathbf{x}_1; \eta, \gamma) = \eta_0 + \eta_1\mathbf{x}_1 + \ldots + \eta_p\mathbf{x}_1^p + \sum_{k=1}^{K} \gamma_k(\mathbf{x}_1 - \kappa_k)_+^p \tag{6.2}$$

where $\eta = (\eta_0, \eta_1, \ldots, \eta_p)^T$ is the $(p+1)$ vector of the coefficients of the polynomial function, $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_K)^T$ is the coefficient vector of the truncated polynomial spline basis (P-spline) and p is the degree of the spline $(t)_+^p = t^p$ if $t > 0$ and 0 otherwise. The latter portion of the model allows for handling departures from a p-polynomial t in the structure of the relationship. In this portion κ_k for $k = 1, \ldots, K$ is a set of fixed knots and if K is sufficiently large, the class of functions in (6.2) is very large and can approximate most smooth functions. Details on bases and knots choice can be found in Chapters 3 and 5 of Ruppert et al. (2003). Since a P-spline model can be viewed as a random-effects model (see Ruppert et al. (2003) and Opsomer et al. (2008)), it can be combined with the Fay-Herriot model for obtaining a semiparametric small area estimation framework based on linear mixed model regression. Given the η and γ vectors, define

$$\mathbf{X}_1 = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1m} & \cdots & x_{1m}^p \end{bmatrix},$$

and

$$\mathbf{S} = \begin{bmatrix} (x_{11} - \kappa_1)_+^p & \cdots & (x_{11} - \kappa_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_{1m} - \kappa_1)_+^p & \cdots & (x_{1m} - \kappa_K)_+^p \end{bmatrix}.$$

Using only one covariate, x_1, the semiparametric Fay-Herriot can be written as

$$\hat{\beta} = \mathbf{X}\beta + \mathbf{S}\gamma + \mathbf{Z}\mathbf{u} + \varepsilon, \tag{6.3}$$

where $\mathbf{X} = \mathbf{X}_1$, β is a $(p+1)$ vector of regression coefficients, the γ component can be treated as a $K \times 1$ vector of independent and identically distributed random variables with mean $\mathbf{0}$ and $K \times K$ variance matrix $\Sigma_\gamma = \sigma_\gamma^2\mathbf{I}_K$. The variance-covariance matrix of the model (6.3) is $\Sigma(\psi) = \mathbf{S}\Sigma_\gamma\mathbf{S}^T + \mathbf{Z}\Sigma_u\mathbf{Z}^T + \mathbf{R}$ where $\psi = (\sigma_\gamma^2, \sigma_u^2)^T$.

Model-based estimation of the small area parameters can be obtained by using the best linear unbiased prediction (see Henderson (1975):

$$\tilde{\theta}^B(\psi) = \mathbf{X}\tilde{\beta}(\psi) + \Lambda(\psi)[\hat{\theta} - \mathbf{X}\tilde{\beta}(\psi)] \tag{6.4}$$

with $\Lambda(\psi) = (\mathbf{S}\Sigma_\gamma \mathbf{S}^T + \mathbf{Z}\Sigma_u \mathbf{Z}^T)\Sigma^{-1}(\psi)$ and $\tilde{\beta}(\psi) = (\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}(\psi)\hat{\theta}$.

Extension to bivariate smoothing can be handled by assuming $\tilde{m}(\mathbf{x}_1, \mathbf{x}_2) = m(\mathbf{x}_1, \mathbf{x}_2; \eta, \gamma)$. See details in Opsomer et al. (2008). This is of central interest in a number of application areas as environment, public health and poverty mapping. It has particular relevance when referenced responses need to be converted to maps.

6.2 Estimation of the MSE

The Mean Squared Error estimator (MSE) of $\tilde{\theta}^B(\psi)$, depending on the variance components $\psi = (\sigma_\gamma^2, \sigma_u^2)^T$, can be expressed as in Rao (2003):

$$MSE[\tilde{\theta}^B(\psi)] = g_1(\psi) + g_2(\psi) \tag{6.5}$$

where the first term

$$g_1(\psi) = \Lambda(\psi)\mathbf{R} = \mathbf{R} - \mathbf{R}\Sigma^{-1}(\psi)\mathbf{R} \tag{6.6}$$

is due to the estimation of random effects and it is of order $O(1)$, while the second term

$$g_2(\psi) = \mathbf{R}\Sigma^{-1}(\psi)\mathbf{X}(\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{R} \tag{6.7}$$

is due to the estimation of β and it is of order $O(m^{-1})$ for large m.

The estimator $\tilde{\theta}^B(\psi)$ depends on the unknown variance components σ_γ^2 and σ_u^2. Replacing the parameters with estimators $\hat{\sigma}_\gamma^2$, $\hat{\sigma}_u^2$, a two stage estimator $\tilde{\theta}^E(\hat{\psi})$ is

$$\tilde{\theta}^E(\hat{\psi}) = \mathbf{X}\hat{\beta}(\hat{\psi}) + \hat{\Lambda}(\hat{\psi})[\hat{\theta} - \mathbf{X}\hat{\beta}(\hat{\psi})] \tag{6.8}$$

where $\hat{\beta}(\hat{\psi}) = (\mathbf{X}^T\hat{\Sigma}^{-1}(\hat{\psi})\mathbf{X})^{-1}\mathbf{X}^T\hat{\Sigma}^{-1}(\hat{\psi})\hat{\theta}$. Assuming normality of the random effects, σ_γ^2 and σ_u^2 can be estimated both by Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) procedures (see Prasad and Rao (1990)).

The ML and REML estimators possess the following properties (see Datta et al. (2005)): (i) they are $m^{1/2}$-consistent; (ii) they are even functions of $\hat{\theta}$, so that $\hat{\psi}(-\hat{\theta}) = \hat{\psi}(\hat{\theta})$; (iii) they are translation invariant functions, so that $\hat{\psi}(\hat{\theta} + \mathbf{G}c) = \hat{\psi}(\hat{\theta})$, for any $m \times (g+1)$ matrix, $c \in \mathbb{R}^{g+1}$ and for all $\hat{\theta}$.

For any $\hat{\psi}$ satisfying (ii) and (iii), the MSE of $\tilde{\theta}^E(\hat{\psi})$ can be decomposed as

$$MSE[\tilde{\theta}^E(\hat{\psi})] = g_1(\psi) + g_2(\psi) + E\left\{[\tilde{\theta}^E(\hat{\psi}) - \tilde{\theta}^B(\psi)]^2\right\} = g_1(\psi) + g_2(\psi) + g_3(\psi). \tag{6.9}$$

Under the model (6.1) with diagonal covariance matrix $\Sigma(\sigma_u^2)$, Prasad and Rao (1990) obtained an approximation up to $o(m^{-1})$ terms of $g_3(\psi)$ through Taylor linearization. In case of the semiparametric Fay & Herriot model the structure of the covariance matrix is not diagonal due to the introduction of the spline random component, then the results of Prasad and Rao (1990) can not be applied directly. The results of Opsomer et al. (2008) can be used for deriving a second order approximation to the $g_3(\psi)$ term. It can be given by

$$g_3(\psi) = \mathbf{L}^T(\psi)\left[I^{-1}(\psi) \otimes \Sigma(\psi)\right]\mathbf{L}(\psi) + o(\delta_m/m) \tag{6.10}$$

where

$$\mathbf{L}(\psi) = [\mathbf{L}_{\sigma_\gamma^2}(\psi), \mathbf{L}_{\sigma_u^2}(\psi)]^T, \mathbf{L}_i(\psi) = \frac{\partial \Lambda(\psi)}{\partial \psi_i},\ i = 1,2.$$

Here $\otimes$ represents Kronecker product, $\mathcal{I}^{-1}(\psi)$ is the inverse of the information matrix with $\mathcal{I}_{ij}^{-1}(\psi) = 0.5tr[\mathbf{P}(\psi)\mathbf{B}_i\mathbf{P}(\psi)\mathbf{B}_j]$, $i,j = 1,2$, $\mathbf{P}(\psi) = \Sigma^{-1}(\psi) - \Sigma^{-1}(\psi)\mathbf{X}(\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}(\psi)$, $\mathbf{B}_1 = \mathbf{S}\mathbf{S}^T$ and $\mathbf{B}_2 = \mathbf{Z}\mathbf{Z}^T$ and $\delta_m = o(\sqrt{m})$.

In practical applications, the EBLUP $\tilde{\theta}^E(\hat{\psi})$ should be accompanied with an estimate of the MSE. Again, under Fay & Herriot models with diagonal covariance matrix, Prasad and Rao (1990) obtained an approximately unbiased estimator of the MSE (6.9). Following the results of Prasad and Rao (1990) and Das et al. (2004), Opsomer et al. (2008) extended the Prasad-Rao MSE estimator to models with more general covariance structure. An approximately unbiased estimator of the MSE is

$$mse[\tilde{\theta}^E(\hat{\psi})] = g_1(\hat{\psi}) + g_2(\hat{\psi}) + 2g_3(\hat{\psi}). \tag{6.11}$$

which is the same estimator derived by Prasad and Rao (1990). In formula (6.11), the term $g_3(\hat{\psi})$ appears twice due to a bias correction of $g_1(\hat{\psi})$.

This section describes an alternative procedure for estimating the MSE of the EBLUP $\tilde{\theta}^E(\hat{\psi})$ based on bootstrapping according to the bootstrap procedure proposed by González-Manteiga et al. (2007), Opsomer et al. (2008) and Molina et al. (2009). In this procedure, the bootstrap random effects $(\gamma_1^*,\ldots,\gamma_K^*)^T$, $(u_1^*,\ldots,u_m^*)^T$ and the random errors $(\varepsilon_1^*,\ldots,\varepsilon_m^*)^T$ are obtained by resampling respectively from the empirical distribution of the predicted random elements $\hat{\gamma} = (\hat{\gamma}_1,\ldots,\hat{\gamma}_K)^T$, $\hat{\mathbf{u}} = (\hat{u}_1,\ldots,\hat{u}_m)^T$, and the residuals $\hat{\mathbf{r}} = \hat{\theta} - \mathbf{X}\hat{\beta} - \mathbf{S}\hat{\gamma} - \mathbf{Z}\hat{\mathbf{u}} = (\hat{r}_1,\ldots,\hat{r}_m)^T$, previously standardized. This method avoids the need of distributional assumptions; therefore, it is expected to be more robust to non-normality of any of the random components of the model. The procedure works as follows:

1. Fit model (6.3) to the initial direct estimates $\hat{\theta}$, obtaining estimates $(\hat{\sigma}_\gamma^2, \hat{\sigma}_u^2)$ and $\hat{\beta}$.

2. With estimates obtained in step 1, calculate predictors of $\hat{\gamma} = (\hat{\gamma}_1,\ldots,\hat{\gamma}_K)^T$ and $\hat{\mathbf{u}} = (\hat{u}_1,\ldots,\hat{u}_m)^T$. Then take $\hat{\gamma}^S = \hat{\Sigma}_\gamma^{-1/2}\hat{\gamma}$ and $\hat{\mathbf{u}}^S = \hat{\Sigma}_u^{-1/2}\hat{\mathbf{u}}$ where $\hat{\Sigma}_\gamma^{-1/2}$ and $\hat{\Sigma}_u^{-1/2}$ are the root square of the generalized inverse of $\hat{\Sigma}_\gamma^{-1/2} = \mathbf{S}\hat{\Sigma}_\gamma\mathbf{S}^T\mathbf{P}(\hat{\psi})\mathbf{S}^T\hat{\Sigma}_\gamma\mathbf{S}$ and $\hat{\Sigma}_u^{-1/2} = \mathbf{Z}\hat{\Sigma}_u\mathbf{Z}^T\mathbf{P}(\hat{\psi})\mathbf{Z}^T\hat{\Sigma}_u\mathbf{Z}$ respectively, obtained by the spectral decomposition. It is convenient re-scale the elements $\hat{\gamma}_k^S$ and $\hat{\mathbf{u}}_i^S$ so that they have sample means exactly equal to zero and sample variances $\hat{\sigma}_\gamma^2, \hat{\sigma}_u^2$. This is achieved by the transformation

$$\hat{\gamma}_k^{SS} = \frac{\hat{\sigma}_\gamma\left\{\hat{\gamma}_k^S - K^{-1}\sum_{j=1}^K \hat{\gamma}_j^S\right\}}{\sqrt{K^{-1}\sum_{d=1}^K\left\{\hat{\gamma}_d^S - K^{-1}\sum_{j=1}^K\hat{\gamma}_j^S\right\}^2}},\ k = 1,\ldots,K$$

$$\hat{u}_i^{SS} = \frac{\hat{\sigma}_u\left\{\hat{u}_i^S - m^{-1}\sum_{j=1}^m \hat{u}_j^S\right\}}{\sqrt{m^{-1}\sum_{d=1}^m\left\{\hat{u}_d^S - m^{-1}\sum_{j=1}^m\hat{u}_j^S\right\}^2}},\ i = 1,\ldots,m.$$

Construct the vectors $\gamma^* = (\gamma_1^*,\ldots,\gamma_K^*)^T$ and $\mathbf{u}^* = (u_1^*,\ldots,u_m^*)^T$, whose elements are obtained by extracting a simple random sample with replacement of size K and m from the sets $\hat{\gamma}^{SS} = (\hat{\gamma}_1^{SS},\ldots,\hat{\gamma}_K^{SS})^T$ and $\hat{\mathbf{u}}^{SS} = (\hat{u}_1^{SS},\ldots,\hat{u}_m^{SS})^T$, respectively. Then calculate the bootstrap quantity of interest $\theta^* = \mathbf{X}\hat{\beta} + \mathbf{S}\gamma^* + \mathbf{Z}\mathbf{u}^* = (\theta_1^*,\ldots,\theta_m^*)^T$.

3. Compute the vector of residuals $\hat{\mathbf{r}} = \hat{\theta} - \mathbf{X}\hat{\beta} - \mathbf{S}\hat{\gamma} - \mathbf{Z}\hat{\mathbf{u}} = (\hat{r}_1, \ldots, \hat{r}_m)^T$. Standardize the residuals by $\hat{\mathbf{r}}^S = (\mathbf{R}\mathbf{P}(\hat{\psi})\mathbf{R})^{-1/2}\hat{\mathbf{r}}$. Re-standardized these values

$$\hat{r}_i^{SS} = \frac{\left\{\hat{r}_i^S - m^{-1}\sum_{j=1}^m \hat{r}_j^S\right\}}{\sqrt{m^{-1}\sum_{d=1}^m \left\{\hat{r}_d^S - m^{-1}\sum_{j=1}^m \hat{r}_j^S\right\}^2}}, \; i = 1, \ldots, m.$$

   Construct the vector $\mathbf{r}^* = (r_1^*, \ldots, r_m^*)^T$, whose elements are obtained by extracting a simple random sample with replacement of size m from the set $\hat{\mathbf{r}}^{SS} = (\hat{r}_1^{SS}, \ldots, \hat{r}_m^{SS})^T$. Then take $\varepsilon^* = (\varepsilon_1^*, \ldots, \varepsilon_m^*)^T$ where $\varepsilon_i^* = \sigma_i r_i^*$.

4. Construct bootstrap data from the model,

$$\hat{\theta}^* = \theta^* + \varepsilon^* = \mathbf{X}\hat{\beta} + \mathbf{S}\gamma^* + \mathbf{Z}\mathbf{u}^* + \varepsilon^* = (\hat{\theta}_1^*, \ldots, \hat{\theta}_m^*)^T.$$

5. Regarding $\hat{\beta}$, $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_u^2$ as the true values of β, σ_γ^2 and σ_u^2, fit the model (6.3) to the bootstrap data $\hat{\theta}^*$. The obtained estimates β^*, σ_γ^{2*} and σ_u^{2*} will be called bootstrap estimators.

6. Calculate the bootstrap small area estimator using β^*, σ_γ^{2*} and σ_u^{2*} in place of the 'true' $\hat{\beta}$, $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_u^2$,

$$\tilde{\theta}^{E*}(\hat{\psi}^*) = \mathbf{X}\hat{\beta}^*(\hat{\psi}^*) + \hat{\Lambda}^*(\hat{\psi}^*)[\hat{\theta}^* - \mathbf{X}\hat{\beta}^*(\hat{\psi}^*)]$$

7. Repeat steps 2-6 B times. In the b-th bootstrap replication, let $\theta_i^{*(b)}$ be the quantity of interest in area i, $\tilde{\theta}_i^{E*}(\hat{\psi}^{*(b)})$ be the bootstrap estimator for area i.

A naïve bootstrap estimator for the MSE for area i is

$$mse_i^{naNPB}[\tilde{\theta}_i^E(\hat{\psi})] = B^{-1}\sum_{b=1}^{B}\left\{\tilde{\theta}_i^{E*(b)}(\hat{\psi}^{*(b)}) - \theta_i^{*(b)}\right\}^2. \tag{6.12}$$

Another MSE estimate can be obtained by adding the bootstrap estimate $g_{3i}^{NPB}(\hat{\psi})$ and the analytical estimates $g_{1i}(\hat{\psi})$ and $g_{2i}(\hat{\psi})$, and then including a bootstrap bias correction of $g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi})$ (see Pfeffermann & Tiller(2006)), as

$$mse_i^{bcNPB}[\tilde{\theta}_i^E(\hat{\psi})] = 2\left[g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi})\right] - B^{-1}\sum_{b=1}^{B}\left[g_{1i}(\hat{\psi}^{*(b)}) + g_{2i}(\hat{\psi}^{*(b)})\right] + g_{3i}^{NPB}(\hat{\psi}). \tag{6.13}$$

where $g_{3i}^{NPB}(\hat{\psi}) = B^{-1}\sum_{b=1}^{B}\left\{\tilde{\theta}_i^{E*(b)}(\hat{\psi}^{*(b)}) - \tilde{\theta}^{BLUP*(b)}(\hat{\psi})\right\}^2$ with $\tilde{\theta}^{BLUP*(b)} = \mathbf{X}\hat{\beta}^* + \mathbf{S}\gamma^* + \mathbf{Z}\mathbf{u}^*$.

6.3 Simulations for semiparametric Fay-Herriot model

In this section we develop a simulation study to compare the performance of the $\tilde{\theta}^E(\hat{\psi})$ estimator of the small area mean under the proposed semiparametric specification (denoted by NPEBLUP hereafter) to that under the traditional Fay-Herriot specification (denoted by EBLUP).

We consider five synthetic populations generated using the following models for creating the true underlying relationship between the covariate x and the expected value of the response variable y $E(y|x) = m(x)$:

Linear. $m(x) = 10 + 2(x)$;

Jump. $m(x) = 1 + 2(x - 1.5)I(x \leqslant 1.5) + 2I(x > 1.5)$.

Exponential. $m(x) = 2 + exp(3x)/400$.

Bump. $m(x) = 10 + 2(x - 1.5) + 5exp(-200(x - 1.5)^2)$.

Cycle. $m(x) = 10 + 10\sin(2\pi x)$;

Population values of y in small area $i = 1, \ldots, 200$ are generated under the random intercepts model

$$y_i = m(x) + u_i + \varepsilon_i$$

with x drawn from a Uniform distribution $[0, 3]$, the area effects u_i drawn from $N(0, 1)$ and the error effects ε_i independently generated from $N(0, 1)$.

The linear case represents a situation in which the EBLUP is based on a good representation of the true model, while the NPEBLUP may be too complex and overparametrized. The jump model is a discontinuous function for which EBLUP and NPEBLUP are based on a misspecified model; the Exponential, Bump and Cycle models define increasingly more complicated structures of the relationship between y and x.

For each of the five generated populations a total of $T = 250$ simulations were carried out. For each sample the EBLUP and the NPEBLUP estimators have been used to estimate the small area means $\bar{y_i}$, $i = 1, \ldots, 200$.

Then, for each estimator and for each small area we computed the Monte Carlo estimate of the Bias

$$\mathrm{B}_{MC} = \frac{1}{T}\sum_{t=1}^{T}(\hat{\bar{y}}_{it} - \bar{y_i}) \tag{6.14}$$

and with it the percentage relative bias

$$\mathrm{RB\%} = \frac{\mathrm{B}_{MC}}{\bar{y_i}}100; \tag{6.15}$$

the Root Mean Squared Error

$$\mathrm{RMSE}_{MC} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\bar{y}}_{it} - \bar{y_i})^2}, \tag{6.16}$$

and the corresponding percentage Relative Root Mean Squared Error

$$\mathrm{RRMSE\%} = \frac{\mathrm{RMSE}_{MC}}{\bar{y_i}}100. \tag{6.17}$$

To evaluate the RB% and the RRMSE% across the 200 small areas we consider these summary statistics: the minimum value, the first quartile, the mean and the median value, the third quartile and the maximum value.

Tables 6.1 and 6.2 report respectively the summary statistics for the RB% and the RRMSE% values obtained for the estimation of the small area means under the Linear, Jump, Exponential, Bump and Cycle signals.

Table 6.1: Percentage Relative Bias (RB%) of the estimators of the small area means.

RB% Point Estimation						
Estimator	Min	1st quartile	Mean	Median	3rd quartile	Max
Linear Signal						
EBLUP	-0.74	-0.12	-0.01	0.00	0.13	0.57
NPEBLUP	-0.44	-0.13	0.00	-0.01	0.11	0.47
Jump Signal						
EBLUP	-204.21	-12.84	0.64	-4.70	10.25	399.83
NPEBLUP	-108.79	-0.63	4.36	0.52	3.45	79.01
Exponential Signal						
EBLUP	-15.92	-4.54	1.29	1.78	8.24	13.42
NPEBLUP	-2.74	-0.82	-0.06	-0.14	0.67	2.80
Bump Signal						
EBLUP	-12.08	0.59	0.19	0.82	1.01	1.79
NPEBLUP	-10.46	-0.11	0.12	0.13	0.60	3.49
Cycle Signal						
EBLUP	-9.23	-2.24	33.14	-0.49	6.13	769.18
NPEBLUP	-46.69	-0.16	-0.62	-0.01	0.22	11.68

The results are promising. First note that the performance of the two estimators is essentially equivalent under the Linear signal, both in terms of bias and variability. Then, from Table 6.1 we see that the mean and median bias of the NPEBLUP estimator are always lower with respect to the EBLUP estimator, with the only exception of the mean value under the Jump signal. Moreover, in many cases there is a high gain also in terms of minimum and maximum values of the RB%: that is, the bias of the NPEBLUP estimator in the estimating the 200 small area means varies in a range of smaller size than the EBLUP. In terms of variability (Table 6.2) the results show a similar behavior: the NPEBLUP is always a good competitor the the EBLUP.

6.4 Estimation of the Mean Squared Error

In this Section we present a simulation experiment carried out to contrast the three alternative estimators of the Mean Squared Error of the NPEBLUP estimator $\tilde{\theta}^E(\hat{\psi})$ described in Section 6.2. Namely, the estimators we consider are the analytical estimator (6.11), the naïve nonparametric bootstrap estimator mse^{naNPB} (6.12) and the combined analytical and bootstrap estimator mse^{bcNPB} (6.13).

The simulation study is carried out using real data coming from the Italian Agricultural Census of year 2000 for the Tuscany region, as in Molina et al. (2009), under two different settings. The small areas of interest are the 287 municipalities of the region, with N_i, $i = 1, \ldots, m$, given by the census and the n_i randomly generated from a Binomial distribution with parameters N_i and $p = 0.05$. These sampling data were used to compute, for each municipality i, the direct estimator of the mean agrarian surface area used for production of grape in hectares (θ_i) and its sampling variance (ψ_i). Information on the agrarian

Table 6.2: Percentage Relative Root Mean Squared Error (RRMSE%) of the estimators of the small area means.

RRMSE% Point Estimation						
Estimator	Min	1st quartile	Mean	Median	3rd quartile	Max
Linear Signal						
EBLUP	4.37	5.03	5.66	5.54	6.28	7.25
NPEBLUP	4.45	5.00	5.67	5.60	6.24	7.36
Jump Signal						
EBLUP	23.29	26.81	108.66	42.74	106.62	2114.17
NPEBLUP	22.61	24.34	113.22	42.40	109.19	2487.47
Exponential Signal						
EBLUP	8.41	27.16	34.90	41.02	43.89	47.84
NPEBLUP	4.58	21.05	27.99	33.32	35.70	38.55
Bump Signal						
EBLUP	5.51	6.61	7.87	7.71	8.91	13.10
NPEBLUP	5.63	6.51	7.80	7.89	8.74	11.62
Cycle Signal						
EBLUP	4.85	5.74	86.00	8.13	23.03	1884.93
NPEBLUP	3.45	4.10	55.80	6.22	18.35	1317.73

surface area used for production in hectares (x_{1i}) and on the average number of working days in the reference year (x_{2i}) for each municipality i is also available from the census data.

Thus, in the simulation study the goal is the estimation of the mean agrarian surface area used for production of grape in hectares (y_i) for all the municipalities of the region, using as explicative variables x_{1i} and x_{2i}, which have a linear relation with y_i, and an intercept term. The centroids of the small areas are also available as spatial reference points (latitude and longitude) and are used in the $\mathbf{S}$ matrix when fitting the semiparametric model under both settings. Since the true sampling variance ψ_i resulted equal to 0 for nine areas, in the simulation experiment we consider $m = 278$. Note that the true sampling variances ψ_i have a highly right-skewed distribution with a range of 102745; this skewness is caused by few municipalities with atypically large sampling variances.

More in detail, in the first simulation setting the Monte Carlo samples are generated at each step as follows: first, the random errors e_i are generated from a normal distribution with mean 0 and variance ψ_i; second, the random effects u_i are generated from a normal distribution with mean 0 and variance σ_u^2 taken equal to the estimated value obtained fitting a linear model with random area effects to the census data, that is $\sigma_u^2 = 56.23$ for all the iterations; then, using the values of the covariates $\mathbf{x}_i = (1, x_{1i}, x_{2i})$ obtained from the census together with the true vector of coefficients $\beta = (-3.72, -0.0095, 0.51)$, the vector $\mathbf{y}$ of responses is generated under model (6.1). In a second alternative setting the steps of Monte Carlo experiment are the same as in the first setting but the vector $\mathbf{y}$ of responses is generated under the model (6.3), with γ random errors generated under a normal distribution with mean 0 and variance $\sigma_\gamma^2 = 15$.

Under both settings we considered $L = 500$ Monte Carlo samples and we computed the three MSE

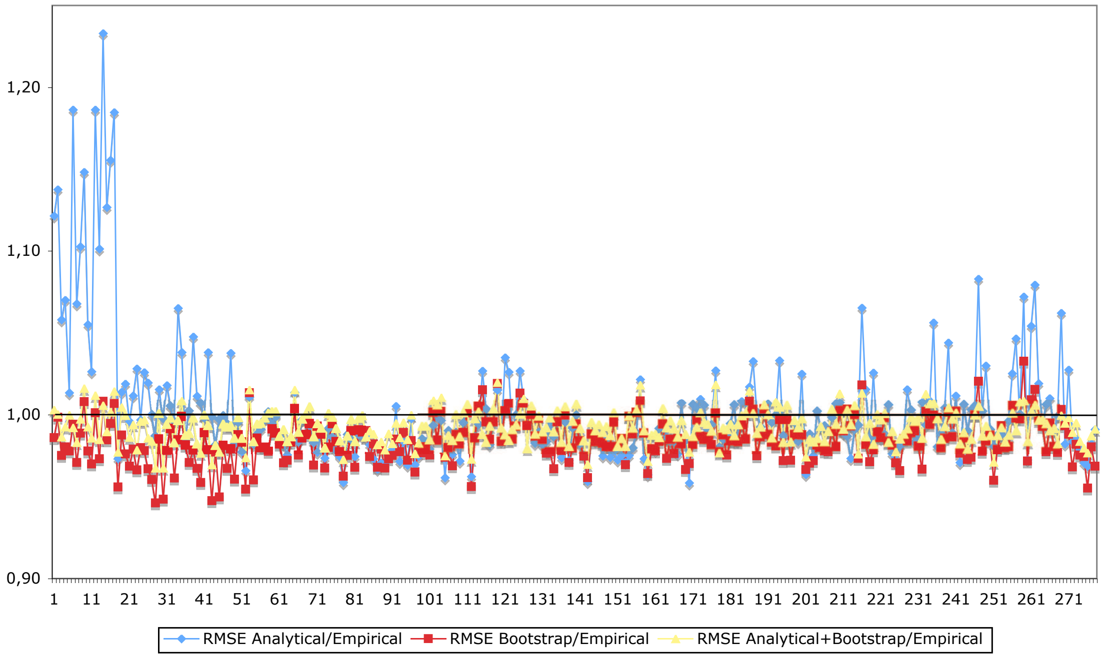


Figure 6.1: Ratios of analytical Root Mean Squared Error (RMSE), naïve nonparametric bootstrap RMSE and combined analytical and bootstrap RMSE over empirical values for the $m = 278$ small areas, model with $\sigma_\gamma^2 = 0$.

estimators of interest, setting the replicates of the two bootstrap procedures to $B = 250$; the final estimates were computed taking the mean over the replicates. The empirical values of the MSEs, that is the reference values, were computed previously under both settings with 1000 Monte Carlo replicates to ensure better accuracy. Figures 6.1 and 6.2 represent for each of the $m = 278$ small areas the ratios of the three estimated Root Mean Squared Error (analytical, naïve nonparametric bootstrap and combined analytical and bootstrap) over the empirical values (represented by the straight lines), under the first and the second setting respectively. Note that to allow a better comparison of the results, the scale used in the two Figures has been zoomed out to the interval 0.9-1.25.

The main result standing from the simulation results is that the two proposed bootstrap estimators of the MSE outperform the analytical one, under both settings. As regards the comparison between the estimator mse^{naNPB} and the estimator mse^{bcNPB}, the first seems to better follow the empirical values (see Figure 6.1). This behavior is the same even considering the second setting, were the model used to

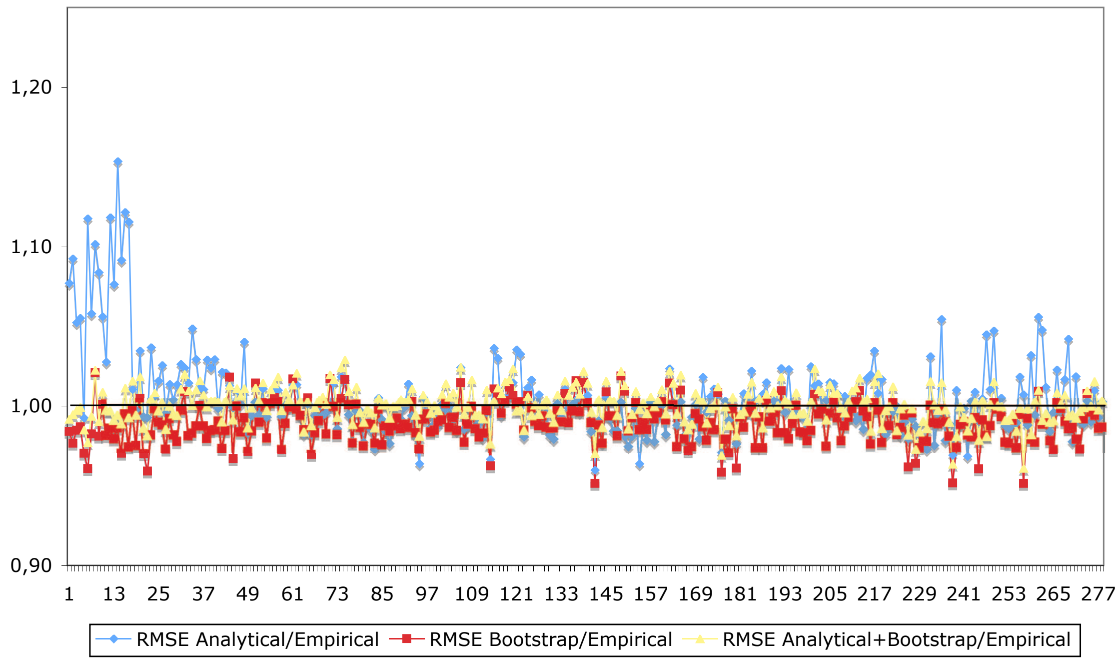


Figure 6.2: Ratios of analytical Root Mean Squared Error (RMSE), naïve nonparametric bootstrap RMSE and combined analytical and bootstrap RMSE over empirical values for the $m = 278$ small areas, model with $\sigma_\gamma^2 = 15$.

generate the y_i values has a spline component: in this case we can observe a slightly higher variability of the estimates, while the estimators are more correct with respect to the empirical values, as expected. Thus, the estimation of the g_3 term of the MSE seems to play an important role in this estimation context.

Chapter 7

Area-level time models

7.1 Area-level model with correlated time effects

7.1.1 Introduction

In the field of small area estimation, data are often available for many small areas simultaneously, although possibly for only a few time points. In such cases, it is desired to borrow information both cross-sectionally and over time. Rao and Yu (1994) gave a simple way of borrowing information cross-sectionally and over time by introducing a model containing both contemporary random effects and time varying effects. They proposed the extension of the basic Fay Herriot model

$$y_{dt} = \mathbf{x}_{dt}\beta + v_d + u_{dt} + e_{dt}, \quad d = 1,\dots,D, \quad t = 1,\dots,T, \tag{7.1}$$

where y_{dt} is a direct estimator of the indicator of interest and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. They assume that $v_1,\dots,v_D$ are i.i.d. normal, $(u_{d1},\dots,u_{dT})$'s follow i.i.d. AR(1) processes (i.e. they follow autoregressive processes of order 1), $e_{11},\dots,e_{DT}$ are i.i.d. normal, and the v_d's, the $(u_{d1},\dots,u_{dT})$'s and the e_{dt}'s are independent.

In this section we introduce a model that it is related to the model (7.1) in the sense that only u_{dt} is considered to take into account the area-by-time variability through specific random effects. The model is

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1,\dots,D, \quad t = 1,\dots,m_d, \tag{7.2}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We further assume that the random vectors $(u_{d1},\dots,u_{dm_d})$, $d = 1,\dots,D$, follow i.i.d. AR(1) processes with variance and auto-correlation parameters σ_u^2 and ρ respectively, the errors e_{dtj}'s are independent $N(0,\sigma_{dt}^2)$ with known σ_{dt}^2's, and the u_{dt}'s are independent of the e_{dt}'s.

In matrix notation the model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{7.3}$$

where $\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$, $\mathbf{y}_d = \underset{1\le t\le m_d}{\text{col}}(y_{dt})$, $\mathbf{u} = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_d)$, $\mathbf{u}_d = \underset{1\le t\le m_d}{\text{col}}(u_{dt})$, $\mathbf{e} = \underset{1\le d\le D}{\text{col}}(\mathbf{e}_d)$, $\mathbf{e}_d = \underset{1\le t\le m_d}{\text{col}}(e_{dt})$, $\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d = \underset{1\le t\le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \underset{1\le i\le p}{\text{col}'}(x_{dti})$, $\beta = \underset{1\le i\le p}{\text{col}}(\beta_i)$, $\mathbf{Z} = \mathbf{I}_{M\times M}$ and $M = \sum_{d=1}^D m_d$. In this notation, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \sigma_u^2 \Omega(\rho), \quad \Omega(\rho) = \underset{1\le d\le D}{\text{diag}} (\Omega_d(\rho)), \quad \mathbf{V}_e = \underset{1\le d\le D}{\text{diag}} (\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1\le t\le m_d}{\text{diag}} (\sigma_{dt}^2),$$

where the σ_{dt}^2 are known and

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \dots & \rho & 1 \end{pmatrix}_{m_d \times m_d}.$$

If the variance components are known, then the BLUE of β and the BLUP of $\mathbf{u}$ are

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u \mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}),$$

where

$$var(\mathbf{y}) = \mathbf{V} = \sigma_u^2 \underset{1\le d\le D}{\text{diag}} (\Omega_d(\rho)) + \mathbf{V}_e = \underset{1\le d\le D}{\text{diag}} (\sigma_u^2 \Omega_d(\rho) + \mathbf{V}_{ed}) = \underset{1\le d\le D}{\text{diag}} (\mathbf{V}_d).$$

To calculate $\widehat{\beta}$ and $\widehat{\mathbf{u}}$ we apply the formulas

$$\widehat{\beta} = \left(\sum_{d=1}^{D} \mathbf{X}_d' \mathbf{V}_d^{-1} \mathbf{X}_d \right)^{-1} \left(\sum_{d=1}^{D} \mathbf{X}_d' \mathbf{V}_d^{-1} \mathbf{y}_d \right), \quad \widehat{\mathbf{u}} = \sigma_u^2 \underset{1\le d\le D}{\text{col}} \left(\Omega_d(\rho) \mathbf{V}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \widehat{\beta}) \right).$$

7.1.2 REML estimators of model parameters

The REML log-likelihood is

$$l_{REML}(\sigma_u^2, \rho) = -\frac{M-p}{2} \log 2\pi + \frac{1}{2} \log |\mathbf{X}'\mathbf{X}| - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2} \mathbf{y}'\mathbf{P}\mathbf{y},$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \quad \mathbf{PVP} = \mathbf{P}, \quad \mathbf{PX} = \mathbf{0}.$$

Let us define $\theta = (\theta_1, \theta_2) = (\sigma_u^2, \rho)$, $\mathbf{V}_1 = \frac{\partial \mathbf{V}}{\partial \sigma_u^2} = \underset{1\le d\le D}{\text{diag}} (\Omega_d(\rho))$ and $\mathbf{V}_2 = \frac{\partial \mathbf{V}}{\partial \rho} = \sigma_u^2 \underset{1\le d\le D}{\text{diag}} (\dot{\Omega}_d(\rho))$. Then

$$\mathbf{P}_a = \frac{\partial \mathbf{P}}{\partial \theta_a} = -\mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_a} \mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P}, \quad a = 1, 2.$$

By taking partial derivatives of l_{REML} with respect to θ_a, we get

$$S_a = \frac{\partial l_{REML}}{\partial \theta_a} = -\frac{1}{2} \text{tr}(\mathbf{P}\mathbf{V}_a) + \frac{1}{2} \mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y}, \quad a = 1, 2.$$

If we take again partial derivatives with respect to θ_a and θ_b, we take expectations and we change the sign, we obtain the elements of the REML Fisher information matrix. These elements are

$$F_{ab} = \frac{1}{2}\,\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a,b = 1,2.$$

We use the Fisher-scoring algorithm to calculate the REML estimates of θ. The updating formula is

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k).$$

As seeds we use $\rho = 0$ and $\sigma_u^{2(0)} = \widehat{\sigma}_{uH}^2$, where $\widehat{\sigma}_{uH}^2$ is the Henderson 3 estimator of σ_u^2 under the model restricted to $\rho = 0$. The REML estimator of β is calculated by applying the formula

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

The asymptotic distributions of the REML estimators of θ and β are

$$\hat{\theta} \sim N_2(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_i are

$$\hat{\theta}_a \pm z_{\alpha/2}\,\nu_{aa}^{1/2},\ a = 1,2, \quad \hat{\beta}_i \pm z_{\alpha/2}\,q_{ii}^{1/2},\ i = 1,\ldots,p,$$

where $\hat{\theta} = \theta^\kappa$, $\mathbf{F}^{-1}(\theta^\kappa) = (\nu_{ab})_{a,b=1,2}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^\kappa)\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\ldots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$). Observed $\hat{\beta}_i = \beta_0$, the p-value for testing the hypothesis $H_0: \beta_i = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_i > |\beta_0|) = 2P(N(0,1) > \beta_0/\sqrt{q_{ii}}).$$

In what follows we present some matrix calculation that are useful to implement the Fisher-scoring algorithm. The target here is to avoid calculations of $M \times M$ matrices.

$$\begin{aligned}
\mathbf{Q} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}, \\
\mathbf{P} &= \underset{1\le d\le D}{\mathrm{diag}}(\mathbf{V}_d^{-1}) - \underset{1\le d\le D}{\mathrm{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\mathrm{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}), \\
\mathbf{P}\mathbf{V}_a &= \underset{1\le d\le D}{\mathrm{diag}}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \underset{1\le d\le D}{\mathrm{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\mathrm{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}), \\
\mathrm{tr}(\mathbf{P}\mathbf{V}_a) &= \sum_{d=1}^{D}\mathrm{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \sum_{d=1}^{D}\mathrm{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q}), \\
\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b) &= \sum_{d=1}^{D}\mathrm{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}) - 2\sum_{d=1}^{D}\mathrm{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q}) \\
&+ \mathrm{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\right\}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y} &= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)' \\
&- \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d\right) \\
&+ \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'.
\end{aligned}$$

Finally, the derivative of matrix $\Omega_d(\rho)$ with respect to ρ is

$$\dot{\Omega}_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 0 & 1 & \dots & \dots & (m_d-1)\rho^{m_d-2} \\ 1 & 0 & \ddots & & (m_d-2)\rho^{m_d-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ (m_d-2)\rho^{m_d-3} & & \ddots & 0 & 1 \\ (m_d-1)\rho^{m_d-2} & \dots & \dots & 1 & 0 \end{pmatrix} + \frac{2\rho\Omega_d(\rho)}{(1-\rho^2)^2}.$$

7.1.3 The mean squared error of the EBLUP

We are interested in predicting the value of $\mu_{dt} = \mathbf{x}_{dt}\beta + u_{dt}$ by using the EBLUP $\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt} = \mathbf{a}'\mathbf{y}$, where $\mathbf{a} = \underset{1\le\ell\le D}{\text{col}}(\underset{1\le k\le m_\ell}{\text{col}}(\delta_{d\ell}\delta_{tk}))$ is a vector having one 1 in the position $t + \sum_{\ell=1}^{d-1} m_\ell$ and 0's in the remaining cells. To estimate $\overline{Y}_{dt}$ we use $\widehat{\overline{Y}}_{dt}^{eblup} = \widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where $\theta = (\sigma_u^2, \rho)$,

$$\begin{aligned}
g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\
g_2(\theta) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}], \\
g_3(\theta) &\approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}
\end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}).$$

Calculation of $g_1(\theta)$

In the formula of $g_1(\theta) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, we have that $\mathbf{Z} = \mathbf{I}_{M\times M}$, and

$$\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \sigma_u^2 \underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho)) - \sigma_u^4 \underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho)) \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}) \underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho)).$$

Let us write $\Omega_d = \Omega_d(\rho)$ and $\mathbf{a}_d = \underset{1\le k\le m_d}{\text{col}}(\delta_{tk})$. Then, $g_1(\theta)$ can be expressed in the form

$$g_1(\theta) = \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d.$$

Calculation of $g_2(\theta)$

We have that $g_2(\theta)=[\mathbf{a}'\mathbf{X}-\mathbf{a}'\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a}-\mathbf{X}'\mathbf{V}_e^{-1}\mathbf{ZTZ}'\mathbf{a}]$, where

$$\begin{aligned}\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} &= \left[\sigma_u^2 \underset{1\le d\le D}{\text{diag}}(\Omega_d)-\sigma_u^4 \underset{1\le d\le D}{\text{diag}}(\Omega_d)\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1})\underset{1\le d\le D}{\text{diag}}(\Omega_d)\right]\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1})\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)\\ &= \sigma_u^2 \underset{1\le d\le D}{\text{col}}(\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d)-\sigma_u^4 \underset{1\le d\le D}{\text{col}}(\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d).\end{aligned}$$

Therefore

$$\begin{aligned}g_2(\theta) &= \left[\mathbf{a}_d'\mathbf{X}_d-\sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d+\sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q}\\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d-\sigma_u^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d+\sigma_u^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\right].\end{aligned}$$

Calculation of $g_3(\theta)$

We have that

$$g_3(\theta)\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where

$$\mathbf{b}'=\mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}=\sigma_u^2\mathbf{a}'\underset{1\le\ell\le D}{\text{diag}}(\Omega_\ell)\underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell^{-1})=\sigma_u^2\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell\Omega_\ell\mathbf{V}_\ell^{-1}).$$

It holds that

$$\begin{aligned}\frac{\partial\mathbf{b}'}{\partial\sigma_u^2} &= \underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1})-\sigma_u^2\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell u}\mathbf{V}_\ell^{-1}),\quad \mathbf{V}_{\ell u}=\frac{\partial\mathbf{V}_\ell}{\partial\sigma_u^2}=\Omega_\ell,\\ \frac{\partial\mathbf{b}'}{\partial\rho} &= \sigma_u^2\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\dot{\Omega}_\ell\mathbf{V}_\ell^{-1})-\sigma_u^2\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell\rho}\mathbf{V}_\ell^{-1}),\quad \mathbf{V}_{\ell\rho}=\frac{\partial\mathbf{V}_\ell}{\partial\rho}=\sigma_u^2\dot{\Omega}_\ell.\end{aligned}$$

We define

$$\begin{aligned}q_{11}=\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\right)' &= \mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d-2\sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\\ &+ \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d,\\ q_{12}=\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\rho}\right)' &= \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d-\sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\\ &- \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d+\sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d,\\ q_{22}=\frac{\partial\mathbf{b}'}{\partial\rho}\underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\rho}\right)' &= \sigma_u^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d-2\sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d\\ &+ \sigma_u^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d.\end{aligned}$$

Finally

$$g_3(\theta)=\text{tr}\left\{\begin{pmatrix} q_{11} & q_{12}\\ q_{21} & q_{22}\end{pmatrix}\begin{pmatrix} F_{11} & F_{12}\\ F_{21} & F_{22}\end{pmatrix}^{-1}\right\},$$

where F_{ab} is the element of the REML Fisher information matrix.

7.1.4 Simulations

Simulation 1

For $d = 1,\dots,D$, $t = 1,\dots,m_d$, the explanatory and target variables are

$$\begin{aligned} x_{dt} &= (b_{dt}-a_{dt})U_{dt}+a_{dt},\ U_{dt}=\frac{t}{m_d+1},\ a_{dt}=1,\ b_{dt}=1+\frac{1}{D}(m_d(d-1)+t),\\ y_{dt} &= \beta_1+\beta_2 x_{dt}+u_{dt}+e_{dt},\ \beta_1=0,\ \beta_2=1, \end{aligned}$$

where $e_{dt} \sim N(0,\sigma_{dt}^2)$ and

$$\sigma_{dt}^2 = \frac{(\alpha_1-\alpha_0)(m_d(d-1)+t-1)}{M-1}+\alpha_0, \quad \alpha_0=0.8,\ \alpha_1=1.2.$$

For $d = 1,\dots,D$, the random effects u_{dt} are calculated as follows:

$$u_{d1} = (1-\rho^2)^{-1/2}\varepsilon_{d1}, \qquad u_{dt}=\rho u_{dt-1}+\varepsilon_{dt},\ t=2,\dots,m_d,$$

where $\varepsilon_{dt} \sim N(0,\sigma_A^2)$ if $d \le D_A$, $\varepsilon_{dt} \sim N(0,\sigma_B^2)$ if $d > D_A$, and $\rho = 0.5$. The first simulation experiment has the following steps:

1. Repeat $K = 10^4$ times ($k = 1,\dots,K$)
   1.1. Generate a sample of size $m = \sum_{d=1}^{D} m_d$ and calculate $\mu_{dt}^{(k)} = \beta_1^{(k)}+\beta_2^{(k)}x_{dt}+u_{dt}^{(k)}$.
   1.2. Calculate $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)},\hat{\beta}_2^{(k)},\hat{\sigma}_u^{2(k)},\widehat{\rho}^{(k)}\}$ and $\hat{\mu}_{dt}^{(k)}$ by using the REML estimation method.
2. For each $\hat{\tau} \in \{\beta_1,\beta_2,\sigma_u^2,\rho\}$ and for $\hat{\mu}_{dt}$, $d = 1,\dots,D$, $t = 1,\dots,m_d$, calculate

$$BIAS(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)}-\tau),\ BIAS_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)}-\mu_{dt}^{(k)}),\ BIAS = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d}BIAS_{dt},$$

$$MSE(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)}-\tau)^2,\ MSE_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)}-\mu_{dt}^{(k)})^2,\ MSE = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d}MSE_{dt}.$$

The simulations are carried out for the 6 combinations of sample sizes appearing in Table 7.1.4.1.

D	50	100	200	300	400	500
m_d	5	5	5	5	5	5
m	250	500	1000	1500	2000	2500

Table 7.1.4.1: Sample sizes.

Table 7.1.4.2 presents the results of the simulation experiment.

D	50	100	200	300	400	500
$BIAS(\hat{\beta}_1)$	0.0020	0.0018	-0.0012	-0.0011	-0.0004	-0.0010
$MSE(\hat{\beta}_1)$	0.0784	0.0410	0.0208	0.0134	0.0100	0.0080
$BIAS(\hat{\beta}_2)$	0.0130	0.0067	0.0034	0.0022	0.0017	0.0013
$MSE(\hat{\beta}_2)$	0.0009	-0.0003	0.0004	0.0005	0.0003	0.0004
$BIAS(\hat{\sigma}_u^2)$	-0,0164	-0,0052	-0,0020	-0.0040	-0.0030	-0.0029
$MSE(\hat{\sigma}_u^2)$	0.0414	0.0213	0.0107	0.0070	0.0053	0.0044
$BIAS(\hat{\rho})$	-0.0018	-0.0009	-0.0002	0.0005	0.0009	0.0013
$MSE(\hat{\rho})$	0.0115	0.0056	0.0027	0.0018	0.0014	0.0011
$BIAS$	0.0005	-0.0003	0.0001	0.0002	0.0000	0.0004
MSE	0.5196	0.5149	0.5121	0.5117	0.5114	0.5113

Table 7.1.4.2. Results of simulation experiment 1.

Table 7.1.4.2 shows that bias is always close to zero and that MSE decreases when the number of domains increases, so that the REML estimates are consistent.

Simulation 2

In the second simulation experiment we investigate the behavior of the estimator mse_{dt} of the MSE of the EBLUP of μ_{dt}. For this task we compare the mse_{dt} with the empirical MSE of $\hat{\mu}_{dt}$ obtained from experiment 1.

1. For $D = 50, 100, 200, 300, 400, 500$,take the values of MSE_d obtained in experiment 1 and repeat $I = 10^4$ times ($k = 1, \dots, K$)

   1.1. Generate the sample $(y_{dt}^{(k)}, \mathbf{x}_{dt})$, $d = 1, \dots, D$, $t = 1, \dots, m_d$.

   1.2. Calculate $\hat{\beta}_1^{(k)}$, $\hat{\beta}_2^{(k)}$, $\hat{\sigma}_u^{2(k)}$ and $mse_{dt}^{(k)} = mse_{dt}(\hat{\sigma}_u^{2(k)})$.

2. Calculate the performance measure of estimator mse_{dt}

$$B_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt}), \quad E_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt})^2, \quad d = 1, \dots, D,$$

$$B = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} B_{dt}, \quad E = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} E_{dt}.$$

Table 7.1.4.3 presents the obtained results.

D	50	100	200	300	400	500
B	-1.4366	-0.5348	-0.0949	-1.1423	-1.0755	-1.2366
E	3.0978	2.1816	1.6443	1.4613	1.3800	1.3508

Table 7.1.4.3. Results of simulation experiment 2.

Tables 7.1.4.3 shows that BIAS and MSE tend to zero as D increases.

7.2 Area-level model with independent time effects

7.2.1 Introduction

This section presents a simplification of model (7.9) that is useful for those cases where survey data is only available for a reduced number of time instants. The new model is defined in the same way as model (7.9), but assuming that $\rho = 0$. Parameter estimates of model (7.4) can also be used as seeds for an iterative fitting method in model (7.9). We assume that

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1,\dots,D, \quad t = 1,\dots,m_d, \tag{7.4}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We assume that the vectors u_{dt}'s are $N(0,\sigma_u^2)$, the errors e_{dt}'s are independent $N(0,\sigma_{dt}^2)$, and the u_{dt}'s are independent of the e_{dt}'s.

Model (7.4) can be alternatively written in the form

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{7.5}$$

where $\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$, $\mathbf{y}_d = \underset{1\le t\le m_d}{\text{col}}(y_{dt})$, $\mathbf{u} = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_d)$, $\mathbf{u}_d = \underset{1\le t\le m_d}{\text{col}}(u_{dt})$, $\mathbf{e} = \underset{1\le d\le D}{\text{col}}(\mathbf{e}_d)$, $\mathbf{e}_d = \underset{1\le t\le m_d}{\text{col}}(e_{dt})$, $\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d = \underset{1\le t\le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \underset{1\le i\le p}{\text{col}'}(x_{dti})$, $\beta = \underset{1\le i\le p}{\text{col}}(\beta_i)$, $\mathbf{Z} = \mathbf{I}_M$, $M = \sum_{d=1}^D m_d$. We assume that $\mathbf{u} \sim N(\mathbf{0},\mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \sigma_u^2\mathbf{I}_M, \quad \mathbf{I}_M = \underset{1\le d\le D}{\text{diag}}(\mathbf{I}_{m_d}), \quad \mathbf{V}_e = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1\le t\le m_d}{\text{col}}(\sigma_{dt}^2),$$

and known variances σ_{dt}^2.

The BLUE of β and the BLUP of $\mathbf{u}$ are

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}),$$

where

$$var(\mathbf{y}) = \mathbf{V} = \sigma_u^2 \underset{1\le d\le D}{\text{diag}}(\mathbf{I}_{m_d}) + \mathbf{V}_e = \underset{1\le d\le D}{\text{diag}}(\sigma_u^2\mathbf{I}_{m_d} + \mathbf{V}_{ed}) = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d).$$

To calculate $\widehat{\beta}$ and $\widehat{\mathbf{u}}$ we apply the formulas

$$\widehat{\beta} = \left(\sum_{d=1}^D \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}\left(\sum_{d=1}^D \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{y}_d\right), \quad \widehat{\mathbf{u}} = \sigma_u^2 \underset{1\le d\le D}{\text{col}}\left(\mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta})\right).$$

7.2.2 The Henderson 3 method

For the linear mixed model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

with $\mathbf{u} \sim N_D(0, \sigma_u^2 \mathbf{I}_D)$ and $\mathbf{e} \sim N_n(0, \sigma_e^2 \mathbf{W}^{-1})$ independent, the Henderson 3 method gives unbiased estimators of σ_e^2 and σ_u^2 by considering the expectations

$$\begin{aligned} E[SSE(\beta, \mathbf{u})] &= \sigma_e^2[n - \text{rg}(\mathbf{X}, \mathbf{Z})], \\ E[SSE(\mathbf{u}|\beta)] &= \text{tr}\left\{\mathbf{Z}'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}']\mathbf{W}\mathbf{Z}\right\}\sigma_u^2 + \sigma_e^2[\text{rg}(\mathbf{X}, \mathbf{Z}) - \text{rg}(\mathbf{X})], \end{aligned}$$

where $SSE(\mathbf{u}|\beta) = SSE(\beta) - SSE(\beta, \mathbf{u})$ y $SSE(\beta)$, $SSE(\beta, \mathbf{u})$ are the sum of squares of residuals of the fixed effect models $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ and $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$ respectively. It hold that

$$\begin{aligned} E[SSE(\beta)] &= E[SSE(\mathbf{u}|\beta)] + E[SSE(\beta, \mathbf{u})] \\ &= \text{tr}\left\{\mathbf{Z}'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}']\mathbf{W}\mathbf{Z}\right\}\sigma_u^2 + \sigma_e^2[n - \text{rg}(\mathbf{X})]. \end{aligned}$$

The Henderson 3 estimators of σ_u^2 is

$$\widehat{\sigma}_{uH}^2 = \frac{SSE(\beta) - \sigma_e^2[n - \text{rg}(\mathbf{X})]}{\text{tr}\{\mathbf{Z}'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}']\mathbf{W}\mathbf{Z}\}},$$

where $SSE(\beta) = \mathbf{y}'\mathbf{P}_2\mathbf{y}$ and

$$\mathbf{P}_2 = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]'\mathbf{W}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}] = \mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}.$$

For the model (7.3) with $\rho = 0$ we have $\sigma_e^2 = 1$, $\mathbf{W} = \mathbf{V}_e^{-1}$, $\mathbf{Z} = \mathbf{I}_M$, $n = M = \sum_{d=1}^D m_d$ and $\text{rg}(\mathbf{X}) = p$. Therefore,

$$\widehat{\sigma}_{uH}^2 = \frac{\mathbf{y}'\mathbf{P}_2\mathbf{y} - (M - p)}{\text{tr}\{\mathbf{P}_2\}},$$

where

$$\begin{aligned} \mathbf{Q}_2 &= (\mathbf{X}'\mathbf{V}_e^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^D (\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)^{-1}, \\ \mathbf{P}_2 &= \mathbf{V}_e^{-1} - \mathbf{V}_e^{-1}\mathbf{X}\mathbf{Q}_2\mathbf{X}'\mathbf{V}_e^{-1} = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1}) - \underset{1\le d\le D}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d)\mathbf{Q}_2 \underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1}), \\ \text{tr}\{\mathbf{P}_2\} &= \sum_{d=1}^D\sum_{t=1}^{m_d}\sigma_{dt}^{-2} - \sum_{d=1}^D \text{tr}\{\mathbf{X}_d'\mathbf{V}_{ed}^{-2}\mathbf{X}_d\mathbf{Q}_2\}, \end{aligned}$$

$$\begin{aligned} \mathbf{y}'\mathbf{P}_2\mathbf{y} &= \underset{1\le d\le D}{\text{col}'}(\mathbf{y}_d)\left[\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1}) - \underset{1\le d\le D}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d)\mathbf{Q}_2 \underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1})\right]\underset{1\le d\le D}{\text{col}}(\mathbf{y}_d) \\ &= \sum_{d=1}^D\sum_{t=1}^{m_d}\sigma_{dt}^{-2}y_{dt}^2 - \left(\sum_{d=1}^D \mathbf{y}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)\mathbf{Q}_2\left(\sum_{d=1}^D \mathbf{y}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)'. \end{aligned}$$

7.2.3 The REML method

The REML log-likelihood is

$$l_{REML}(\sigma_u^2) = -\frac{M-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}'\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where $\mathbf{P}=\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, $\mathbf{PVP}=\mathbf{P}$ and $\mathbf{PX}=\mathbf{0}$. Let us define $\mathbf{V}_u=\frac{\partial \mathbf{V}}{\partial \sigma_u^2}=\mathbf{I}_M$, $\mathbf{P}_u=\frac{\partial \mathbf{P}}{\partial \sigma_u^2}=-\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_u^2}\mathbf{P}=-\mathbf{P}\mathbf{V}_u\mathbf{P}=-\mathbf{P}^2$. The derivative of l_{REML} with respect to $\theta=\sigma_u^2$ is

$$S=S(\theta)=\frac{\partial l_{REML}}{\partial \theta}=-\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_u)+\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_u\mathbf{P}\mathbf{y}=-\frac{1}{2}\text{tr}(\mathbf{P})+\frac{1}{2}\mathbf{y}'\mathbf{P}^2\mathbf{y}.$$

The minus expectation of the second order derivative of l_{REML} with respect to $\theta=\sigma_u^2$ is

$$F=F(\theta)=\frac{1}{2}\,\text{tr}(\mathbf{P}\mathbf{V}_u\mathbf{P}\mathbf{V}_u)=\frac{1}{2}\,\text{tr}(\mathbf{P}^2). \tag{7.6}$$

The updating formula of the Fisher-scoring algorithm is

$$\theta^{k+1}=\theta^k+F^{-1}(\theta^k)S(\theta^k).$$

The Henderson 3 estimator $\widehat{\sigma}_{uH}^2$ can be used as seed of the Fisher-scoring algorithm. The REML estimator of β is

$$\widehat{\beta}_{REML}=(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of the REML estimators of σ_u^2 and β are

$$\hat{\sigma}_u^2\sim N_2(\theta,F^{-1}(\sigma_u^2)),\quad \hat{\beta}\sim N_p(\beta,(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for σ_u^2 and β_i are

$$\hat{\sigma}_u^2\pm z_{\alpha/2}\,\nu^{1/2},\qquad \hat{\beta}_i\pm z_{\alpha/2}\,q_{ii}^{1/2},\ i=1,\ldots,p,$$

where $\hat{\sigma}_u^2=\sigma_u^{2,(\kappa)}$, $\nu=F^{-1}(\sigma_u^{2,(\kappa)})$, $(\mathbf{X}'\mathbf{V}^{-1}(\sigma_u^{2,(\kappa)})\mathbf{X})^{-1}=(q_{ij})_{i,j=1,\ldots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$. Observed $\hat{\beta}_i=\beta_0$, the p-value for testing the hypothesis $H_0:\beta_i=0$ is

$$p=2P_{H_0}(\hat{\beta}_i>|\beta_0|)=2P(N(0,1)>\beta_0/\sqrt{q_{ii}}).$$

In what follows we present some matrix calculation that are useful to implement the Fisher-scoring algorithm. The target here is to avoid calculations of $M\times M$ matrices.

$$\begin{aligned}
\mathbf{Q} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}=\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1},\\
\mathbf{P} &= \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1})-\underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}),\\
\text{tr}(\mathbf{P}) &= \sum_{d=1}^{D}\text{tr}(\mathbf{V}_d^{-1})-\sum_{d=1}^{D}\text{tr}(\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\mathbf{Q}),
\end{aligned}$$

$$\begin{aligned}
\text{tr}(\mathbf{P}^2) &= \sum_{d=1}^{D}\text{tr}(\mathbf{V}_d^{-2})-2\sum_{d=1}^{D}\text{tr}(\mathbf{X}_d'\mathbf{V}_d^{-3}\mathbf{X}_d\mathbf{Q})\\
&+ \text{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\right\}.
\end{aligned}$$

$$\begin{aligned}\mathbf{y}'\mathbf{P}^2\mathbf{y} &= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-2}\mathbf{y}_d - 2\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{y}_d\right)\\ &+ \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'.\end{aligned}$$

7.2.4 Mean squared error of the EBLUP

We are interested in predicting $\mu_{dt} = \mathbf{x}_{dt}\beta + u_{dt}$ with the EBLUP $\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{dt}$. No taking into account the error e_{dt}, this is equivalent to predict $y_{dt} = \mathbf{a}'\mathbf{y}$, where $\mathbf{a} = \underset{1\le\ell\le D}{\text{col}}(\underset{1\le k\le m_\ell}{\text{col}}(\delta_{d\ell}\delta_{tk}))$ is a vector having one “1” in the cell $t+\sum_{\ell=1}^{d-1} m_\ell$ and ”0”’s in the remaining cells. The total $\overline{Y}_{dt}$ is estimated with $\widehat{\overline{Y}}_{dt}^{eblup} = \widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where $\theta = \sigma_u^2$ and

$$\begin{aligned} g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a},\\ g_2(\theta) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}] \qquad \text{y}\\ g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}\end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}).$$

Calculation of $g_1(\sigma_u^2)$

We have that $g_1(\sigma_u^2) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, where $\mathbf{Z} = \mathbf{I}_{M\times M}$ and

$$\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \sigma_u^2\mathbf{I}_M - \sigma_u^4 \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}).$$

We define $\mathbf{a}_d = \underset{1\le k\le m_d}{\text{col}}(\delta_{tk})$. Then, we have

$$g_1(\sigma_u^2) = \sigma_u^2\mathbf{a}_d'\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d = \frac{\sigma_u^2\sigma_{dt}^2}{\sigma_u^2+\sigma_{dt}^2}.$$

Calculation of $g_2(\sigma_u^2)$

We have that $g_2(\sigma_u^2) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}]$, where

$$\begin{aligned}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X} &= \left[\sigma_u^2\mathbf{I}_M - \sigma_u^4\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1})\right]\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1})\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)\\ &= \sigma_u^2\underset{1\le d\le D}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_u^4\underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d).\end{aligned}$$

Therefore,

$$\begin{aligned} g_2(\sigma_u^2) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_u^2\mathbf{a}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_u^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_u^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{a}_d + \sigma_u^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{V}_d^{-1}\mathbf{a}_d\right] \end{aligned}$$

Calculation of $g_3(\sigma_u^2)$

We have that

$$g_3(\sigma_u^2) \approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where

$$\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \sigma_u^2\mathbf{a}' \underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell^{-1}) = \sigma_u^2 \underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}).$$

It holds that

$$\frac{\partial\mathbf{b}'}{\partial\sigma_u^2} = \underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) - \sigma_u^2 \underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell u}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell u} = \frac{\partial\mathbf{V}_\ell}{\partial\sigma_u^2} = \mathbf{I}_{m_\ell}.$$

We define

$$\begin{aligned} q &= \frac{\partial\mathbf{b}'}{\partial\sigma_u^2} \underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\right)' = \mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d - 2\sigma_u^2\mathbf{a}_d'\mathbf{V}_d^{-2}\mathbf{a}_d + \sigma_u^4\mathbf{a}_d'\mathbf{V}_d^{-3}\mathbf{a}_d \\ &= \frac{1}{\sigma_u^2+\sigma_{dt}^2} - \frac{2\sigma_u^2}{(\sigma_u^2+\sigma_{dt}^2)^2} + \frac{\sigma_u^4}{(\sigma_u^2+\sigma_{dt}^2)^3}, \end{aligned}$$

Finally, we get

$$g_3(\sigma_u^2) = qF^{-1}(\sigma_u^2),$$

where F is the REML Fisher amount of information calculated in the updating equation of the Fisher-scoring algorithm (cf. (7.6)).

7.2.5 Simulations

Simulation 1

For $d = 1,\dots,D$, $t = 1,\dots,m_d$, The explanatory and target variables are

$$\begin{aligned} x_{dt} &= (b_{dt}-a_{dt})U_{dt} + a_{dt},\ U_{dt} = \frac{t}{m_d+1},\ a_{dt} = 1,\ b_{dt} = 1 + \frac{1}{D}\left(m_d(d-1)+t\right), \\ y_{dt} &= \beta_1 + \beta_2 x_{dt} + u_{dt} + e_{dt},\ \beta_1 = 0,\ \beta_2 = 1, \end{aligned}$$

where $u_{dt} \sim N(0,\sigma_u^2)$, $e_{dt} \sim N(0,\sigma_{dt}^2)$, $\sigma_u^2 = 1$ ands

$$\sigma_{dt}^2 = \frac{(\alpha_1-\alpha_0)\left(m_d(d-1)+t-1\right)}{M-1} + \alpha_0, \quad \alpha_0 = 0.8, \alpha_1 = 1.2.$$

The first simulation experiment has the following steps:

1. Repeat $K = 10^4$ times $(k = 1,\dots,K)$

1.1. Generate a sample of size M and calculate $\mu_{dt}^{(k)} = \beta_1^{(k)} + \beta_2^{(k)} x_{dt} + u_{dt}^{(k)}$.

1.2. Calculate $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\sigma}_u^{2(k)}\}$ and $\hat{\mu}_{dt}^{(k)}$ by using the REML method.

2. For each $\hat{\tau} \in \{\beta_1, \beta_2, \sigma_u^2\}$ and for $\hat{\mu}_{dt}$, $d = 1, \dots, D$, $t = 1, \dots, m_d$, calculate

$$BIAS(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)} - \tau), \quad MSE(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)} - \tau)^2.$$

$$BIAS_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)} - \mu_{dt}^{(k)}), \quad MSE_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)} - \mu_{dt}^{(k)})^2,$$

$$BIAS = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d} BIAS_{dt}, \quad MSE = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d} MSE_{dt}.$$

The simulation experiment is carried out for the 6 combinations of sample sizes appearing in Table 7.2.5.1.

D	50	100	200	300	400	500
m_d	5	5	5	5	5	5
M	250	500	1000	1500	2000	2500

Table 7.2.5.1: Sample sizes.

The Table 7.2.5.2 presents the results of the simulation experiment.

D	50	100	200	300	400	500
$BIAS(\hat{\beta}_1)$	0.0010	0.0020	-0.0008	-0.0008	-0.0005	-0.0007
$MSE(\hat{\beta}_1)$	0.0472	0.0245	0.0122	0.0080	0.0059	0.0047
$BIAS(\hat{\beta}_2)$	0.0007	-0.0006	0.0003	0.0004	0.0003	0.0004
$MSE(\hat{\beta}_2)$	0.0083	0.0043	0.0022	0.0014	0.0011	0.0008
$BIAS(\hat{\sigma}_u^2)$	-0.0038	0.0010	0.0017	-0.0008	-0.0001	-0.0001
$MSE(\hat{\sigma}_u^2)$	0.0319	0.0159	0.0081	0.0052	0.0040	0.0032
$BIAS$	0.0020	0.0010	-0.0002	-0.0001	0.0002	-0.0003
MSE	0.5064	0.5025	0.5000	0.4997	0.4994	0.4992

Table 7.2.5.2. Results of simulation experiment 1.

The Table 7.2.5.2 shows that the bias is always close to zero and that the MSE decreases as the number of domains increases, so that the REML estimates are consistent.

Simulation 2

The second simulation experiment investigates the behavior of the estimator mse_{dt} of the MSE of the EBLUP of μ_{dt}. We compare mse_{dt} with the empirical MSE of $\hat{\mu}_{dt}$ obtained from Experiment 1.

1. For $D = 50, 100, 200, 300, 400, 500$, take the values of MSE_{dt} obtained in simulation 1 and repeat $I = 10^4$ times $(k = 1, \dots, K)$

1.1. Generate the sample $(y_{dt}^{(k)}, \mathbf{x}_{dt})$, $d=1,\ldots,D$, $t=1,\ldots,m_d$.

1.2. Calculate $\hat{\sigma}_u^{2(k)}$ and $mse_{dt}^{(k)} = mse_{dt}(\hat{\sigma}_u^{2(k)})$.

2. Calculate the performance measures of estimator mse_{dt}

$$B_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt}), \quad E_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt})^2, \quad d=1,\ldots,D,$$

$$B = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} B_{dt}, \quad E = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} E_{dt}.$$

The Table 7.2.5.3 presents the obtained results.

D	50	100	200	300	400	500
B	-0.8957	0.1581	0.7045	-0.1818	-0.0684	-0.1334
E	2.8852	1.8964	1.3884	1.1960	1.1179	1.0805

Table 7.2.5.3. Results of simulation experiment 2.

The Table 7.2.5.3 shows that the BIAS and the MSE tends to zero as D increases.

7.2.6 The impact of the correlation parameter

Two simulation experiments for analyzing the behavior of the EBLUP and its mean squared error estimator are presented in this section. The scope of the simulations is to investigate when it is worthwhile and what is gained when using the more complicated model (7.9) with correlation parameter ρ instead of the simplified model (7.4) restricted to $\rho = 0$. For $d=1,\ldots,D$, $t=1,\ldots,m_d$, the explanatory and target variables are

$$\begin{aligned} x_{dt} &= (b_{dt} - a_{dt})U_{dt} + a_{dt},\ U_{dt} = \frac{t}{m_d+1},\ a_{dt} = 1,\ b_{dt} = 1 + \frac{1}{D}(m_d(d-1)+t), \\ y_{dt} &= \beta_1 + \beta_2 x_{dt} + u_{dt} + e_{dt},\ \beta_1 = 0,\ \beta_2 = 1, \end{aligned}$$

where $e_{dt} \sim N(0, \sigma_{dt}^2)$, $\sigma_{dt}^2 = \alpha_0 + \frac{(\alpha_1-\alpha_0)(m_d(d-1)+t-1)}{M-1}$, $\alpha_0 = 0.8$ and $\alpha_1 = 1.2$. For $d=1,\ldots,D$, the random vectors $(u_{d1},\ldots,u_{dm_d})$ are generated as follows:

$$u_{d1} = (1-\rho^2)^{-1/2}\varepsilon_{d1}, \quad u_{dt} = \rho u_{dt-1} + \varepsilon_{dt},\ t = 2,\ldots,m_d,$$

where $\varepsilon_{dt} \sim N(0, \sigma_u^2)$, $d=1,\ldots,D$, $t=1,\ldots,m_d$, and $\sigma_u^2 = 1$.

The first simulation experiment is dedicated to investigated the gain of efficiency achieved by the EBLUP based on model (7.9) as a function of the correlation parameter ρ. The experiment has the following steps:

1. For $\rho = 0, 1/4, 1/2, 3/4$, repeat $K = 10^4$ times $(k=1,\ldots,K)$

1.1. Generate a sample of size $m = \sum_{d=1}^{D} m_d$. Calculate $\mu_{dt}^{(k)} = \beta_1 + \beta_2 x_{dt} + u_{dt}^{(k)}$.

1.2. Calculate $\hat{\beta}_1^{(k,0)}, \hat{\beta}_2^{(k,0)}, \hat{\sigma}_u^{2(k,0)}$ and EBLUP0 $\hat{\mu}_{dt}^{(k,0)}$ by using REML method under (7.4) restricted to $\rho = 0$.

1.3. Calculate $\hat{\beta}_1^{(k,1)}, \hat{\beta}_2^{(k,1)}, \hat{\sigma}_u^{2(k,1)}, \widehat{\rho}^{(k,1)}$ and EBLUP1 $\hat{\mu}_{dt}^{(k,1)}$ by using REML method under model (7.9).

2 For $d = 1, \ldots, D$, $t = 1, \ldots, m_d$, calculate

$$BIAS_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\mu}_{dt}^{(k,a)} - \mu_{dt}^{(k)}\right), \quad MSE_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k,a)} - \mu_{dt}^{(k)})^2, \quad a = 0, 1,$$

$$BIAS^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} BIAS_{dt}^{(a)}, \quad MSE^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} MSE_{dt}^{(a)}, \quad a = 0, 1.$$

Mean squared errors $MSE^{(0)}$ and $MSE^{(1)}$ are presented in the Table 7.2.6.1 (left). Biases $BIAS^{(0)}$ and $BIAS^{(1)}$ are presented in the Table 7.2.6.1 (right). In the Figure 7.2.6.1 the MSE_{dm_d}-values are plotted for $D = 100$, $m_d = 5$ and $\rho = 0$ (top-left), $\rho = 0.25$ (top-right), $\rho = 0.5$ (bottom-left) and $\rho = 0.75$ (bottom-right). In the Figure 7.2.6.2 the $BIAS_{dm_d}$-values are plotted for $D = 100$, $m_d = 5$ with the same configuration as in the Figure 7.2.6.1.

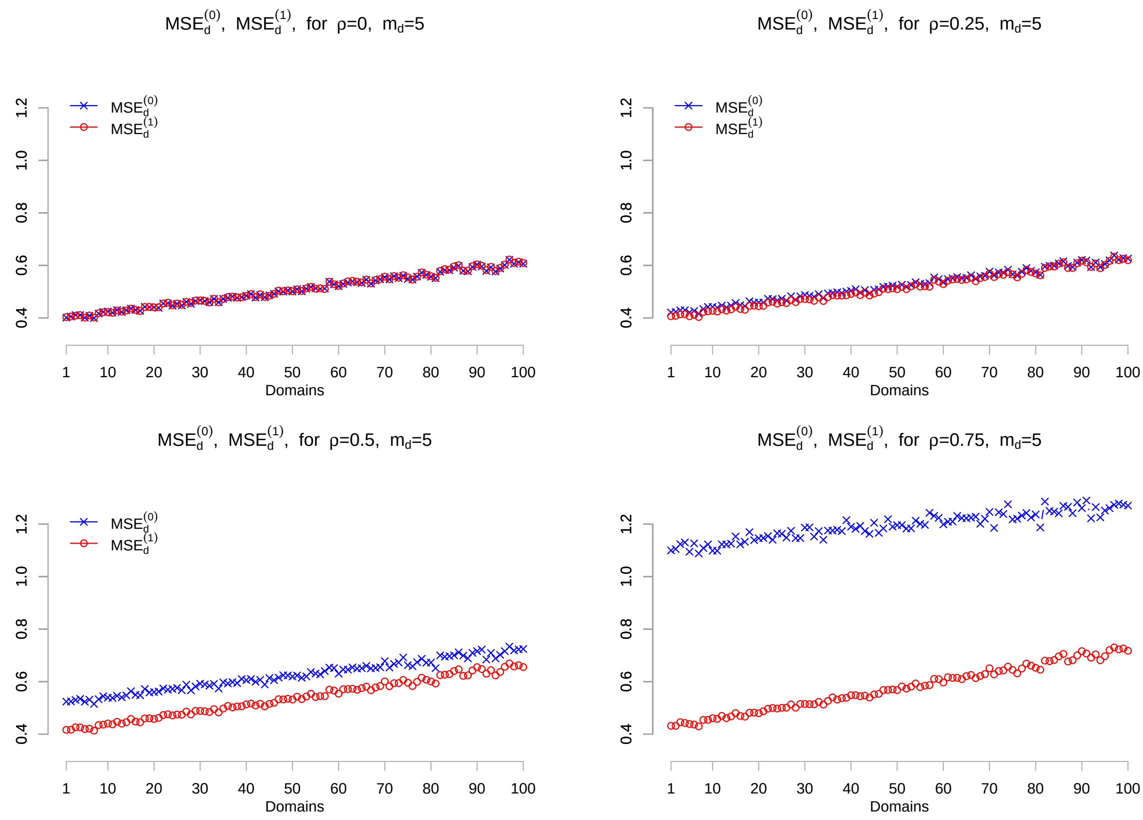


Figure 7.2.6.1. MSE_{dm_d}'s of EBLUP0 and EBLUP1 for $D = 100$, $m_d = 5$.

When the true model is model (7.4) restricted to $\rho = 0$, the best results in MSE are obtained if we work all the time under the assumption that $\rho = 0$. However if we use the EBLUP derived under the incorrect model (7.9) the increment of MSE is almost negligible. This can be appreciated in the two first rows of the Table 7.2.6.1 (left) and on the Figure 7.2.6.1. If we look at the bias, no increment is observed for incorrectly using model (7.9).

		m_d				m_d			
ρ	a	2	5	10	20	2	5	10	20
0	0	0.5086	0.5026	0.5003	0.4996	0.00078	-0.00011	0.00053	-0.00001
0	1	0.5138	0.5046	0.5014	0.5001	0.00078	-0.00011	0.00053	-0.00001
0.25	0	0.5263	0.5204	0.5185	0.5176	0.00079	-0.00011	0.00053	-0.00002
0.25	1	0.5214	0.5074	0.5026	0.5007	0.00078	-0.00011	0.00052	-0.00001
0.5	0	0.6263	0.6189	0.6183	0.6193	-0.00020	-0.00133	0.00196	0.00103
0.5	1	0.5457	0.5133	0.5052	0.5015	-0.00021	-0.00132	0.00193	0.00104
0.75	0	1.2021	1.1903	1.1930	1.1971	-0.00030	-0.00130	0.00197	0.00106
0.75	1	0.5953	0.5230	0.5029	0.4935	-0.00032	-0.00129	0.00192	0.00106

Table 7.2.6.1. *MSE*'s (left) and *BIAS*'s (right) of EBLUP0 and EBLUP1 for $D = 100$

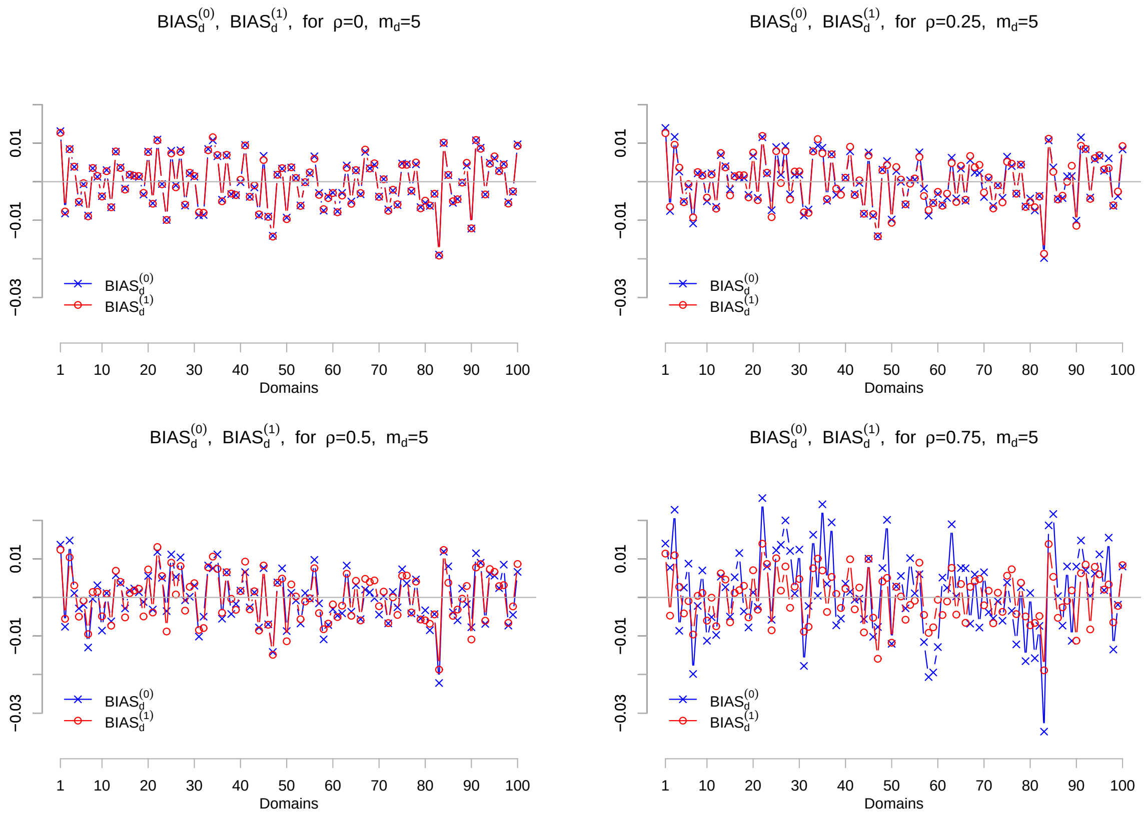


Figure 7.2.6.2. $BIAS_{dm_d}$'s of EBLUP0 and EBLUP1 for $D = 100$, $m_d = 5$.

When the true model is model (7.9) and the correlation parameter is small ($\rho = 0.25$), there is almost no difference in MSE or BIAS by using the true model or the incorrect model (7.4). If the correlation parameter is of medium size ($\rho = 0.5$) there is a clear increment of MSE and BIAS by using the incorrect model. Finally if the correlation parameter is high ($\rho = 0.75$) the use of the incorrect model produce sever increments of MSE and BIAS.

The second simulation experiment takes the MSEs obtained in the first experiment and includes the following additional steps:

1.4 Calculate $mse(\hat{\mu}_{dt}^{(k,0)})$ and $mse(\hat{\mu}_{dt}^{(k,1)})$.

3 For $d = 1,\dots,D$, $t = 1,\dots,m_d$, calculate

$$B_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}\left(mse(\hat{\mu}_{dt}^{(k,a)} - MSE_{dt}^{(a)}\right), E_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}\left(mse(\hat{\mu}_{dt}^{(k,a)} - MSE_{dt}^{(a)}\right)^2, a = 0,1,$$

$$B^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} B_{dt}^{(a)}, \quad E^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} E_{dt}^{(a)}, \quad a = 0,1.$$

Mean squared errors $E^{(0)}$ and $E^{(1)}$ are presented in the Table 7.2.6.2 (left). Biases $B^{(0)}$ and $B^{(1)}$ are presented in the Table 7.2.6.2 (right). For $D = 100$ and $m_d = 5$, in the Figure 7.2.6.3 the B_{dm_d}-values are plotted on the top for $\rho = 0$ and $\rho = 0.75$ and the E_{dm_d}-values are plotted in the bottom for the same values of ρ. We observe that in the case $\rho = 0$ there is no difference between working under the true model (7.4) or under the incorrect model (7.9). On the other hand, if $\rho = 0.75$ then we get higher bias and mean squared error in the estimation of the MSE of the EBLUP by working under model (7.4). Again we conclude that if true model is model (7.9), then there is a loss of efficiency by using model (7.4). The cases $\rho = 0.25$ and $\rho = 0.5$ has been also analyzed, but not presented here as they represent a smooth transition between the two extreme considered cases.

		m_d				m_d			
ρ	a	2	5	10	20	2	5	10	20
0	0	0.00347	0.00194	0.00140	0.00112	-0.00118	-0.00015	0.00014	-0.00038
0	1	0.00350	0.00194	0.00140	0.00112	-0.00086	-0.00018	0.00013	-0.00038
0.25	0	0.00350	0.00202	0.00150	0.00122	-0.00118	-0.00006	-0.00007	-0.00023
0.25	1	0.00352	0.00203	0.00146	0.00118	-0.00116	-0.00047	-0.00007	-0.00059
0.5	0	0.00365	0.00242	0.00195	0.00168	-0.00139	-0.00028	-0.00052	-0.00030
0.5	1	0.00398	0.00222	0.00161	0.00132	-0.00198	-0.00109	-0.00073	-0.00113
0.75	0	0.00465	0.00395	0.00361	0.00336	-0.00307	-0.00209	-0.00232	-0.00190
0.75	1	0.00513	0.00243	0.00173	0.00141	-0.00405	-0.00225	-0.00165	-0.00162

Table 7.2.6.2. E's (left) and B's (right) of EBLUP0 and EBLUP1 for $D = 100$

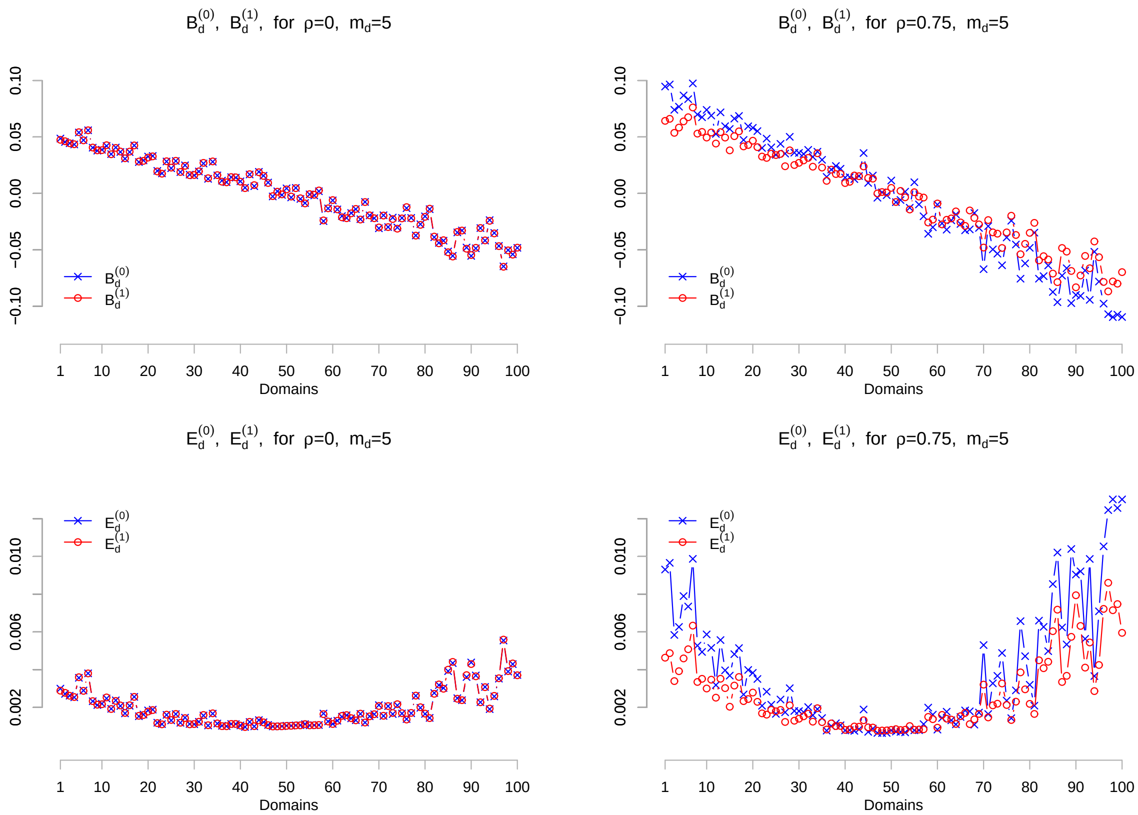


Figure 7.2.6.3. B_{dm_d}'s (top) and E_{dm_d}'s (bottom) of EBLUP0 and EBLUP1 for $D = 100$, $m_d = 5$.

7.3 Partitioned Fay-Herriot model 1

7.3.1 The model

Let us consider the model (*model*1)

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1,\dots,D = D_A + D_B, \quad t = 1,\dots,m_d, \tag{7.7}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We assume that the random effects u_{dt}'s are i.i.d. $N(0,\sigma_A^2)$ if $d \le D_A$ and i.i.d. $N(0,\sigma_B^2)$ if $d > D_A$. We further assume that the errors e_{dt}'s are independent $N(0,\sigma_{dt}^2)$ with known σ_{dt}^2's. Finally we assume that the u_{dt}'s and the e_{dt}'s are mutually independent. In matrix notation the model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where vectors $\mathbf{y}$, $\mathbf{u}$ and $\mathbf{e}$ can be decomposed in the form $\mathbf{v} = (\mathbf{v}_A', \mathbf{v}_B')'$, with $\mathbf{v}_A = \underset{d \le D_A}{\text{col}}(\mathbf{v}_d)$, $\mathbf{v}_B = \underset{d > D_A}{\text{col}}(\mathbf{v}_d)$ and $\mathbf{v}_d = \underset{1 \le t \le m_d}{\text{col}}(v_{dt})$, matrix $\mathbf{X}$ can be similarly decomposed in the form $\mathbf{X} = (\mathbf{X}_A', \mathbf{X}_B')'$, with $\mathbf{X}_A = \underset{d \le D_A}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_B = \underset{d > D_A}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d = \underset{1 \le t \le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \underset{1 \le j \le p}{\text{col}'}(x_{dtj})$, $\beta = \beta_{p \times 1}$, $\mathbf{Z} = \mathbf{I}_M$, $M = M_A + M_B$, $M_A = \sum_{d \le D_A} m_d$, $M_B = \sum_{d > D_A} m_d$ an $\mathbf{I}_M$ denotes the identity $M \times M$ matrix. In this notation, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u)$

and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \text{var}(\mathbf{u}) = \text{diag}(\sigma_A^2 \mathbf{I}_{M_A}, \sigma_B^2 \mathbf{I}_{M_B}), \quad \mathbf{V}_e = \text{var}(\mathbf{e}) = \underset{1 \le d \le D}{\text{diag}} (\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1 \le t \le m_d}{\text{diag}} (\sigma_{dt}^2).$$

The covariance matrix of vector $\mathbf{y}$ is $\mathbf{V} = \text{var}(\mathbf{y}) = \text{diag}(\mathbf{V}_A, \mathbf{V}_B)$, where $\mathbf{V}_A = \underset{d \le D_A}{\text{diag}} (\mathbf{V}_d)$, $\mathbf{V}_B = \underset{d > D_A}{\text{diag}} (\mathbf{V}_d)$, $\mathbf{V}_d = \sigma_A^2 \mathbf{I}_{m_d} + \mathbf{V}_{ed}$ if $d \le D_A$ and $\mathbf{V}_d = \sigma_B^2 \mathbf{I}_{m_d} + \mathbf{V}_{ed}$ if $d > D_A$.

If $\sigma_A^2 > 0$ and $\sigma_B^2 > 0$ are known, the best linear unbiased estimator (BLUE) of β is

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

and the best linear unbiased predictor (BLUP) of $\mathbf{u}$ is

$$\widehat{\mathbf{u}} = \mathbf{V}_u \mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}) = \text{diag}\left(\sigma_A^2 \mathbf{I}_{M_A}, \sigma_B^2 \mathbf{I}_{M_B}\right) \underset{1 \le d \le D}{\text{col}} (\mathbf{V}_d^{-1})(\mathbf{y} - \mathbf{X}\widehat{\beta}),$$

so that

$$\widehat{\mathbf{u}}_d = \begin{cases} \sigma_A^2 \mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta}), & d = 1, \dots, D_A, \\ \sigma_B^2 \mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta}), & d = D_A + 1, \dots, D, \end{cases}$$

or equivalently

$$\hat{u}_{dt} = \left[\frac{\sigma_A^2}{\sigma_A^2 + \sigma_{dt}^2} I_{\{d \le D_A\}}(d) + \frac{\sigma_B^2}{\sigma_B^2 + \sigma_{dt}^2} I_{\{d > D_A\}}(d) \right] (y_{dt} - \mathbf{x}_{dt}\hat{\beta}),\ d = 1, \dots, D, t = 1, \dots, m_d.$$

The loglikelihood of the restricted (residual) maximum likelihood method is

$$\begin{aligned} l_{reml} &= l_{reml}(\sigma_A^2, \sigma_B^2) = -\frac{M-p}{2} \log 2\pi + \frac{1}{2} \log |\mathbf{X}'\mathbf{X}| - \frac{1}{2} \log |\mathbf{V}_A| - \frac{1}{2} \log |\mathbf{V}_B| \\ &- \frac{1}{2} \log |\mathbf{X}_A'\mathbf{V}_A^{-1}\mathbf{X}_A + \mathbf{X}_B'\mathbf{V}_B^{-1}\mathbf{X}_B| - \frac{1}{2} \mathbf{y}'\mathbf{P}\mathbf{y}, \end{aligned}$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \quad \mathbf{P}\mathbf{V}\mathbf{P} = \mathbf{P}, \quad \mathbf{P}\mathbf{X} = \mathbf{0}.$$

Let $\theta = (\theta_1, \theta_2) = (\sigma_A^2, \sigma_B^2)$, then

$$\mathbf{V}_1 = \frac{\partial \mathbf{V}}{\partial \sigma_A^2} = \text{diag}\Big(\mathbf{I}_{M_A}, \underset{d > D_A}{\text{diag}} (\mathbf{0}_{m_d \times m_d})\Big), \quad \mathbf{V}_2 = \frac{\partial \mathbf{V}}{\partial \sigma_B^2} = \text{diag}\Big(\underset{d \le D_A}{\text{diag}} (\mathbf{0}_{m_d \times m_d}), \mathbf{I}_{M_B}\Big).$$

Then

$$\mathbf{P}_a = \frac{\partial \mathbf{P}}{\partial \theta_a} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_a}\mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P}, \quad a = 1, 2.$$

By taking partial derivatives of l_{reml} with respect to θ_a, we get the scores

$$S_a = \frac{\partial l_{reml}}{\partial \theta_a} = -\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y}, \quad a = 1, 2.$$

By taking again partial derivatives with respect to θ_a and θ_b, taking expectations and changing the sign, we get the Fisher information matrix components

$$F_{ab} = \frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a,b=1,2.$$

To calculate the REML estimate we apply the Fisher-scoring algorithm with the updating fórmula

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k),$$

where $\mathbf{S}$ and $\mathbf{F}$ are the column vector of scores and the Fisher information matrix respectively. As seeds we use $\sigma_A^{2(0)} = \sigma_B^{2(0)} = \widehat{\sigma}_{uH}^2$, where $\widehat{\sigma}_{uH}^2$ is the Henderson 3 estimator under model with $\sigma_A^2 = \sigma_B^2$. The REML estimator of β is

$$\widehat{\beta}_{reml} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of the REML estimators of θ and β are

$$\hat{\theta} \sim N_2(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}\nu_{aa}^{1/2},\ a=1,2, \quad \hat{\beta}_j \pm z_{\alpha/2}q_{jj}^{1/2},\ j=1,\ldots,p,$$

where $\hat{\theta} = \theta^\kappa$, $\mathbf{F}^{-1}(\theta^\kappa) = (\nu_{ab})_{a,b=1,2}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^\kappa)\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\ldots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$). Observed $\hat{\beta}_j = \beta_0$, the p-value for testing the hypothesis $H_0: \beta_j = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_j > |\beta_0|) = 2P(N(0,1) > \beta_0/\sqrt{q_{jj}}).$$

In what follows we present some matrix calculation that are useful to implement the Fisher-scoring algorithm. The target here is to avoid calculations of $M \times M$ matrices. For ease of exposition we define the sets of indexes $\mathcal{D}_1 = \{1,\ldots,D_A\}$ and $\mathcal{D}_2 = \{D_A+1,\ldots,D\}$.

$$\begin{aligned}
\mathbf{Q} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}, \\
\mathbf{P} &= \underset{1\le d\le D}{\mathrm{diag}}(\mathbf{V}_d^{-1}) - \underset{1\le d\le D}{\mathrm{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\mathrm{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}), \\
\mathbf{P}\mathbf{V}_1 &= \mathrm{diag}\big(\underset{d\le D_A}{\mathrm{diag}}(\mathbf{V}_d^{-1}), \mathbf{0}_{M_B\times M_B}\big) - \underset{1\le d\le D}{\mathrm{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\mathrm{col}'\big(\underset{d\le D_A}{\mathrm{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}), \mathbf{0}_{p\times M_B}\big), \\
\mathbf{P}\mathbf{V}_2 &= \mathrm{diag}\big(\mathbf{0}_{M_A\times M_A}, \underset{d> D_A}{\mathrm{diag}}(\mathbf{V}_d^{-1})\big) - \underset{1\le d\le D}{\mathrm{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\mathrm{col}'\big(\mathbf{0}_{p\times M_A}, \underset{d> D_A}{\mathrm{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1})\big), \\
\mathrm{tr}(\mathbf{P}\mathbf{V}_a) &= \sum_{d\in\mathcal{D}_a}\mathrm{tr}(\mathbf{V}_d^{-1}) - \sum_{d\in\mathcal{D}_a}\mathrm{tr}(\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\mathbf{Q}), \quad a=1,2, \\
\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_a) &= \sum_{d\in\mathcal{D}_a}\mathrm{tr}(\mathbf{V}_d^{-2}) - 2\sum_{d\in\mathcal{D}_a}\mathrm{tr}(\mathbf{X}_d'\mathbf{V}_d^{-3}\mathbf{X}_d\mathbf{Q}) \\
&+ \mathrm{tr}\left\{\left(\sum_{d\in\mathcal{D}_a}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d\in\mathcal{D}_a}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\right\}, \quad a=1,2, \\
\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b) &= \mathrm{tr}\left\{\left(\sum_{d\in\mathcal{D}_a}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d\in\mathcal{D}_a}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\right\}, \quad a,b=1,2, \text{ with } a\neq b,
\end{aligned}$$

$$\begin{aligned}
\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y} &= \sum_{d\in\mathcal{D}_a}\mathbf{y}_d'\mathbf{V}_d^{-2}\mathbf{y}_d - \left(\sum_{d\in\mathcal{D}_a}\mathbf{y}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)' \\
&- \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d\in\mathcal{D}_a}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{y}_d\right) \\
&+ \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d\in\mathcal{D}_a}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)', \quad a=1,2, \\
\mathbf{y}'\mathbf{P}\mathbf{y} &= \underset{1\le d\le D}{\text{col}'}(\mathbf{y}_d')\left(\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}) - \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1})\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d')\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1})\right)\underset{1\le d\le D}{\text{col}}(\mathbf{y}_d) \\
&= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{y}_d - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{y}_d\right).
\end{aligned}$$

7.3.2 The mean squared error of the EBLUP

We are interested in predicting the value of $\mu_{dt} = \mathbf{x}_{dt}\beta + u_{dt}$ by using the EBLUP $\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt} = \mathbf{a}'\mathbf{y}$, where $\mathbf{a} = \underset{1\le\ell\le D}{\text{col}}\big(\underset{1\le k\le m_\ell}{\text{col}}(\delta_{d\ell}\delta_{tk})\big)$ is a vector having one 1 in the position $t + \sum_{\ell=1}^{d-1} m_\ell$ and 0's in the remaining cells. To estimate $\overline{Y}_{dt}$ we use $\widehat{\overline{Y}}_{dt}^{eblup} = \widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where $\theta = (\sigma_A^2, \sigma_B^2)$,

$$\begin{aligned}
g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\
g_2(\theta) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}], \\
g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}
\end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}).$$

Calculation of $g_1(\theta)$

In the formula of $g_1(\theta) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, we have that $\mathbf{Z} = \mathbf{I}_M$, and $\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \text{diag}(\mathbf{T}_A, \mathbf{T}_B)$, where

$$\mathbf{T}_A = \sigma_A^2\mathbf{I}_{M_A} - \sigma_A^4\underset{d\le D_A}{\text{diag}}(\mathbf{V}_d^{-1}), \quad \mathbf{T}_B = \sigma_B^2\mathbf{I}_{M_B} - \sigma_B^4\underset{d>D_A}{\text{diag}}(\mathbf{V}_d^{-1}).$$

Let us write $\mathbf{a}_d = \underset{1\le k\le m_d}{\text{col}}(\delta_{tk})$. Then, $g_1(\theta)$ can be expressed in the form

$$g_1(\theta) = \begin{cases} \sigma_A^2 - \sigma_A^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d = \frac{\sigma_A^2\sigma_{dt}^2}{\sigma_A^2+\sigma_{dt}^2} & \text{if } d \le D_A, \\ \sigma_B^2 - \sigma_B^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d = \frac{\sigma_B^2\sigma_{dt}^2}{\sigma_B^2+\sigma_{dt}^2} & \text{if } d > D_A. \end{cases}$$

Calculation of $g_2(\theta)$

We have that $g_2(\theta) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{ZTZ}'\mathbf{a}]$, where $\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} =$

$$\begin{aligned}\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} &= \begin{pmatrix} \left[\sigma_A^2\mathbf{I}_{M_A} - \sigma_A^4 \underset{d\le D_A}{\text{diag}}(\mathbf{V}_d^{-1})\right] \underset{d\le D_A}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{d\le D_A}{\text{col}}(\mathbf{X}_d) \\ \left[\sigma_B^2\mathbf{I}_{M_B} - \sigma_B^4 \underset{d> D_A}{\text{diag}}(\mathbf{V}_d^{-1})\right] \underset{d> D_A}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{d> D_A}{\text{col}}(\mathbf{X}_d) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_A^2 \underset{d\le D_A}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_A^4 \underset{d\le D_A}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \\ \sigma_B^2 \underset{d> D_A}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_B^4 \underset{d> D_A}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \end{pmatrix}.\end{aligned}$$

Therefore

$$\begin{aligned} g_2(\theta) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_A^2\mathbf{a}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_A^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_A^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{a}_d + \sigma_A^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{V}_d^{-1}\mathbf{a}_d\right] \quad \text{if } d \le D_A, \\ &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_B^2\mathbf{a}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_B^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_B^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{a}_d + \sigma_B^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{V}_d^{-1}\mathbf{a}_d\right] \quad \text{if } d > D_A. \end{aligned}$$

Calculation of $g_3(\theta)$

We have that

$$g_3(\theta) \approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where $\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \mathbf{a}'\text{diag}\left(\sigma_A^2\mathbf{I}_{M_A}, \sigma_B^2\mathbf{I}_{M_B}\right) \underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell^{-1}) = (\mathbf{b}_A', \mathbf{b}_B')$,

$$\mathbf{b}_A' = \sigma_A^2 \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) \quad \text{and} \quad \mathbf{b}_B' = \sigma_B^2 \underset{\ell> D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}).$$

It holds that $\frac{\partial\mathbf{b}'}{\partial\sigma_A^2} = (\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}, \mathbf{0})$, $\frac{\partial\mathbf{b}'}{\partial\sigma_B^2} = (\mathbf{0}, \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2})$, where

$$\begin{aligned} \frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2} &= \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) - \sigma_A^2 \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-2}), \\ \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2} &= \underset{\ell> D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) - \sigma_B^2 \underset{\ell> D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-2}). \end{aligned}$$

We define

$$\begin{aligned} q_{11} &= \frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2} \underset{\ell\le D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}\right)' = \left[\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d - 2\sigma_A^2\mathbf{a}_d'\mathbf{V}_d^{-2}\mathbf{a}_d + \sigma_A^4\mathbf{a}_d'\mathbf{V}_d^{-3}\mathbf{a}_d\right] I_{\{d\le D_A\}}(d) \\ &= \frac{\sigma_{dt}^4}{(\sigma_A^2+\sigma_{dt}^2)^3} I_{\{d\le D_A\}}(d) \\ q_{22} &= \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2} \underset{\ell> D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2}\right)' = \left[\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d - 2\sigma_B^2\mathbf{a}_d'\mathbf{V}_d^{-2}\mathbf{a}_d + \sigma_B^4\mathbf{a}_d'\mathbf{V}_d^{-3}\mathbf{a}_d\right] I_{\{d> D_A\}}(d) \\ &= \frac{\sigma_{dt}^4}{(\sigma_B^2+\sigma_{dt}^2)^3} I_{\{d> D_A\}}(d). \end{aligned}$$

Finally

$$g_3(\theta) = \begin{cases} q_{11}F_{11}^{-1}, & \text{if } d \le D_A \\ q_{22}F_{22}^{-1}, & \text{if } d > D_A, \end{cases}$$

where F_{ab} is the element of the REML Fisher information matrix.

7.3.3 Testing for $H_0 : \sigma_A^2 = \sigma_B^2$

Let $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ be the unrestricted REML estimators of σ_A^2 and σ_B^2 respectively. Let $\tilde{\sigma}_u^2$ be the REML estimator of the common value $\sigma_A^2 = \sigma_B^2$ under H_0. The REML likelihood ratio statistic (LRS) for testing $H_0 : \sigma_A^2 = \sigma_B^2$ is

$$\begin{aligned} \lambda &= -2[l_{REML}(\tilde{\sigma}_u^2) - l_{REML}(\hat{\sigma}_A^2, \hat{\sigma}_B^2)] = \log\frac{|\tilde{V}|}{|\hat{V}|} + \log\frac{|X'\tilde{V}^{-1}X|}{|X'\hat{V}^{-1}X|} + y'\tilde{P}y - y'\hat{P}y \\ &= \log|\tilde{V}| - \log|\hat{V}_A| - \log|\hat{V}_B| + \log|X'\tilde{V}^{-1}X| \\ &- \log|X_A'\hat{V}_A^{-1}X_A + X_B'\hat{V}_B^{-1}X_B| + y'\tilde{P}y - y'\hat{P}y. \end{aligned}$$

Asymptotic distribution of λ under H_0 is χ_1^2, so null hypothesis is rejected at the level α if $\lambda > \chi_{1,\alpha}^2$.

7.3.4 Simulations

Simulation 1

For $d = 1,\dots,D$, $t = 1,\dots,m_d$, The explanatory and target variables are

$$\begin{aligned} x_{dt} &= (b_{dt} - a_{dt})U_{dt} + a_{dt},\ U_{dt} = \frac{t}{m_d+1},\ a_{dt} = 1,\ b_{dt} = 1 + \frac{1}{D}(m_d(d-1)+t), \\ y_{dt} &= \beta_1 + \beta_2 x_{dt} + u_{dt} + e_{dt},\ \beta_1 = 0,\ \beta_2 = 1, \end{aligned}$$

where $u_{dt} \sim N(0,\sigma_A^2)$ if $d \le D_A$, $u_{dt} \sim N(0,\sigma_B^2)$ if $d > D_A$, $e_{dt} \sim N(0,\sigma_{dt}^2)$, $D_A = D/2$, $\sigma_A^2 = 1$, $\sigma_B^2 = 0.8, 1, 1.2$ and

$$\sigma_{dt}^2 = \frac{(\alpha_1 - \alpha_0)(m_d(d-1)+t-1)}{M-1} + \alpha_0, \quad \alpha_0 = 0.8,\ \alpha_1 = 1.2.$$

The first simulation experiment has the following steps:

1. Repeat $K = 10^3$ times ($k = 1,\dots,K$)
   1.1. Generate a sample of size M and calculate $\mu_{dt}^{(k)} = \beta_1^{(k)} + \beta_2^{(k)} x_{dt} + u_{dt}^{(k)}$.
   1.2. Calculate $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\sigma}_A^{2(k)}, \hat{\sigma}_B^{2(k)}\}$ and $\hat{\mu}_{dt}^{(k)}$ by using the REML method.
2. For each $\hat{\tau} \in \{\beta_1, \beta_2, \sigma_A^2, \sigma_B^2\}$ and for $\hat{\mu}_{dt}$, $d = 1,\dots,D$, $t = 1,\dots,m_d$, calculate

$$BIAS(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)} - \tau), \quad MSE(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)} - \tau)^2.$$

$$BIAS_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)} - \mu_{dt}^{(k)}), \quad MSE_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)} - \mu_{dt}^{(k)})^2,$$

$$BIAS = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d} BIAS_{dt}, \quad MSE = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d} MSE_{dt}.$$

The simulation experiments are carried out for the 6 combinations of sample sizes appearing in Table 7.3.4.1.

D	50	100	200	300	400	500
m_d	5	5	5	5	5	5
M	250	500	1000	1500	2000	2500

Table 7.3.4.1: Sample sizes.

The Table 7.3.4.2 presents the results of the simulation experiment for the case $\sigma_A^2 = 1$ and $\sigma_B^2 = 0.8$.

D	50	100	200	300	400	500
$BIAS(\hat{\beta}_1)$	0.0010	0.0020	-0.0007	-0.0008	-0.0004	-0.0006
$MSE(\hat{\beta}_1)$	0.0457	0.02359	0.0118	0.0077	0.0057	0.0046
$BIAS(\hat{\beta}_2)$	0.0007	-0.0005	0.0003	0.0004	0.0003	0.0003
$MSE(\hat{\beta}_2)$	0.0079	0.0040	0.0020	0.0013	0.0010	0.0008
$BIAS(\hat{\sigma}_{uA}^2)$	-0.0928	-0.0896	-0.0890	-0.0899	-0.0900	-0.0898
$MSE(\hat{\sigma}_{uA}^2)$	0.0608	0.0344	0.0213	0.0166	0.0145	0.0134
$BIAS(\hat{\sigma}_{uB}^2)$	0.1052	0.1112	0.1118	0.1079	0.1094	0.1092
$MSE(\hat{\sigma}_{uB}^2)$	0.0765	0.0455	0.0287	0.0222	0.0201	0.0184
$BIAS$	0.0021	0.0010	-0.0002	-0.0009	0.0002	-0.0003
MSE	0.5174	0.4974	0.4867	0.4833	0.4820	0.4811

Table 7.3.4.2. Results of simulation 1 under $\sigma_A^2 = 1$, $\sigma_B^2 = 0.8$.

The Table 7.3.4.2 shows that the bias is always close to zero and that the MSE decreases as the number of domains increases, so that the REML estimates are consistent.

The Table 7.3.4.3 presents the results of the simulation experiment for the case $\sigma_A^2 = 1$ and $\sigma_B^2 = 1$.

D	50	100	200	300	400	500
$BIAS(\hat{\beta}_1)$	0.0014	0.0017	-0.0005	-0.0005	-0.0002	-0.0004
$MSE(\hat{\beta}_1)$	0.0305	0.0156	0.0078	0.0051	0.0038	0.0030
$BIAS(\hat{\beta}_2)$	0.0009	-0.0006	0.0002	0.0003	0.0003	0.0001
$MSE(\hat{\beta}_2)$	0.0051	0.0026	0.0013	0.0008	0.0007	0.0005
$BIAS(\hat{\sigma}_{uA}^2)$	-0.2478	-0.2468	-0.2462	-0.2465	-0.2464	-0.2466
$MSE(\hat{\sigma}_{uA}^2)$	0.0846	0.0725	0.0665	0.0646	0.0636	0.0631
$BIAS(\hat{\sigma}_{uB}^2)$	-0.2493	-0.2459	-0.2453	-0.2479	-0.2469	-0.2470
$MSE(\hat{\sigma}_{uB}^2)$	0.0896	0.0745	0.0669	0.0659	0.0644	0.0637
$BIAS$	0.0011	0.0005	-0.0001	-0.0004	0.0009	-0.0001
MSE	0.2536	0.2456	0.2416	0.2410	0.2404	0.2401

Table 7.3.4.3. Results of simulation 1 under $\sigma_A^2 = 1$, $\sigma_B^2 = 1$.

The Table 7.3.4.3 shows that the bias is always close to zero and that the MSE decreases as the number of domains increases, so that the REML estimates are consistent.

The Table 7.3.4.4 presents the results of the simulation experiment for the case $\sigma_A^2 = 1$ and $\sigma_B^2 = 1.2$.

D	50	100	200	300	400	500
$BIAS(\hat{\beta}_1)$	0.0014	0.0017	-0.0005	-0.0005	-0.0002	-0.0004
$MSE(\hat{\beta}_1)$	0.0323	0.0166	0.0082	0.0054	0.0040	0.0032
$BIAS(\hat{\beta}_2)$	0.0007	-0.0007	0.0002	0.0003	0.0001	0.0003
$MSE(\hat{\beta}_2)$	0.0056	0.0029	0.0014	0.0009	0.0007	0.0006
$BIAS(\hat{\sigma}_{uA}^2)$	-0.2478	-0.2468	-0.2462	-0.2465	-0.2464	-0.2466
$MSE(\hat{\sigma}_{uA}^2)$	0.0846	0.0725	0.0665	0.0646	0.0636	0.0631
$BIAS(\hat{\sigma}_{uB}^2)$	-0.2496	-0.2457	-0.2451	-0.2480	-0.2469	-0.2470
$MSE(\hat{\sigma}_{uB}^2)$	0.0989	0.0791	0.0690	0.0674	0.0656	0.0646
$BIAS$	0.0011	0.0005	-0.0001	-0.0005	0.0008	-0.0001
MSE	0.2613	0.2454	0.2373	0.2350	0.2340	0.2333

Table 7.3.4.4. Results of simulation1 under $\sigma_A^2 = 1$, $\sigma_B^2 = 1.2$.

The Table 7.3.4.4 shows that the bias is always close to zero and that the MSE decreases as the number of domains increases, so that the REML estimates are consistent.

Simulation 2

The second simulation experiment investigates the behavior of the estimator mse_{dt} of the MSE of the EBLUP of μ_{dt}. We compare mse_{dt} with the empirical MSE of $\hat{\mu}_{dt}$ obtained from Experiment 1.

1. For $D = 50, 100, 200, 300, 400, 500$, take the values of MSE_{dt} obtained in simulation 1 and repeat $K = 10^3$ times ($k = 1, \ldots, K$)

   1.1. Generate the sample $(y_{dt}^{(k)}, \mathbf{x}_{dt})$, $d = 1, \ldots, D$, $t = 1, \ldots, m_d$.

   1.2. Calculate $mse_{dt}^{(k)} = mse_{dt}(\hat{\sigma}_u^{2(k)})$.

2. Calculate the performance measures of estimator mse_{dt}

$$B_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt}), \quad E_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt})^2, \quad d = 1, \ldots, D,$$

$$B = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} B_{dt}, \quad E = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} E_{dt}.$$

The Table 7.3.4.5 presents the obtained results for the case $\sigma_A^2 = 1$ and $\sigma_B^2 = 0.8$.

D	50	100	200	300	400	500
B	-0.0336	-0.0173	-0.0090	-0.0070	-0.0060	-0.0054
E	8.1547	3.4678	1.6412	1.1217	0.9084	0.7736

Table 7.3.4.5. Results of simulation 2 under $\sigma_A^2 = 1$, $\sigma_B^2 = 0.8$.

The Table 7.3.4.5 shows that the BIAS and the MSE tends to zero as D increases.

The Table 7.3.4.6 presents the obtained results for the case $\sigma_A^2 = 1$ and $\sigma_B^2 = 1$.

D	50	100	200	300	400	500
B	0.0588	0.0606	0.0613	0.0607	0.0608	0.0607
E	4.6616	4.1992	4.0266	3.8739	3.8572	3.8305

Table 7.3.4.6. Results of simulation 2 under $\sigma_A^2 = 1$, $\sigma_B^2 = 1$.

The Table 7.3.4.6 shows that the BIAS and the MSE tends to zero as D increases.

The Table 7.3.4.7 presents the obtained results for the case $\sigma_A^2 = 1$ and $\sigma_B^2 = 1.2$.

D	50	100	200	300	400	500
B	0.0694	0.0776	0.0818	0.0826	0.0830	0.0833
E	5.7342	6.5256	7.1781	7.3190	7.4025	7.4494

Table 7.3.4.7. Results of simulation 2 under $\sigma_A^2 = 1$, $\sigma_B^2 = 1.2$.

The Table 7.3.4.7 shows that the BIAS and the MSE don't tends to zero as D increases.

Simulation 3

This simulation experiment is planned to check the behavior of the REML log-likelihood test statistics. For $\sigma_A^2 = 1$ and $\sigma_B^2 = 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2$. The steps of the simulation experiment are:

1. Repeat $K = 10^3$ times ($k = 1, \dots, K$)

   1.1. Generate a sample of size M in the same way as in Simulation 1 and calculate the unrestricted and H_0-restricted REML estimates $\hat{\sigma}^2_{A(k)}$, $\hat{\sigma}^2_{B(k)}$ and $\tilde{\sigma}^2_{u(k)}$.

   1.2. Calculate $\lambda_{(k)} = \lambda(\hat{\sigma}^2_{A(k)}, \hat{\sigma}^2_{B(k)}; \tilde{\sigma}^2_{u(k)})$ and $\alpha_{(k)} = I\left\{\lambda_{(k)} > \chi^2_{1,0.05}\right\}$.

2. Output: $\alpha = \frac{1}{K}\sum_{k=1}^{K} \alpha_{(k)}$.

The simulation experiment is carried out for the 6 combinations of sample sizes appearing in Table 7.3.4.4.

D	50	100	200	300	400	500
m_d	5	5	5	5	5	5
M	250	500	1000	1500	2000	2500

Table 7.3.4.4: Sample sizes.

The Table 7.3.4.5 presents the results of the simulation experiment.

D	50	100	200	300	400	500
$\sigma_B^2 = 0.25$	0.9987	1	1	1	1	1
$\sigma_B^2 = 0.5$	0.9987	1	1	1	1	1
$\sigma_B^2 = 0.75$	0.2352	0.4131	0.6928	0.8652	0.9416	0.9746
$\sigma_B^2 = 1$	0.0517	0.0557	0.0491	0.0481	0.0508	0.0496
$\sigma_B^2 = 1.25$	0.1695	0.3096	0.5301	0.6968	0.8166	0.8926
$\sigma_B^2 = 1.5$	0.4619	0.7536	0.9671	0.9962	0.9996	1
$\sigma_B^2 = 1.75$	0.7423	0.9571	0.9997	1	1	1
$\sigma_B^2 = 2$	0.9006	0.9954	1	1	1	1

Table 7.3.4.5. Power results (α) of simulation experiment 3.

The Table 7.3.4.5 shows that, under $\sigma_B^2 = \sigma_A^2$, the value of α converge to 0.05 as D increases.

7.4 Partitioned Fay-Harriot model 2

7.4.1 The model

Let us consider the model (model 2)

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1,\dots,D = D_A + D_B, \quad t = 1,\dots,m_d, \tag{7.8}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We assume that the random vectors $(u_{d1},\dots,u_{dm_d})$, $d \leq D_A$, follow i.i.d. first order auto-regressive processes with variance and auto-correlation parameters σ_A^2 and ρ respectively; in short, $(u_{d1},\dots,u_{dm_d}) \sim_{iid} AR1(\sigma_A^2,\rho)$, $d \leq D_A$. We further assume that $(u_{d1},\dots,u_{dm_d}) \sim_{iid} AR1(\sigma_B^2,\rho)$, $d > D_A$, and that the errors e_{dt}'s are independent $N(0,\sigma_{dt}^2)$ with known σ_{dt}^2's. Finally we assume that the $(u_{d1},\dots,u_{dm_d})$'s and the e_{dt}'s are mutually independent.

In matrix notation the model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where vectors $\mathbf{y}$, $\mathbf{u}$ and $\mathbf{e}$ can be decomposed in the form $\mathbf{v} = (\mathbf{v}_A', \mathbf{v}_B')'$, with $\mathbf{v}_A = \underset{d\leq D_A}{\text{col}}(\mathbf{v}_d)$, $\mathbf{v}_B = \underset{d>D_A}{\text{col}}(\mathbf{v}_d)$ and $\mathbf{v}_d = \underset{1\leq t\leq m_d}{\text{col}}(v_{dt})$, matrix $\mathbf{X}$ can be similarly decomposed in the form $\mathbf{X} = (\mathbf{X}_A', \mathbf{X}_B')'$, with $\mathbf{X}_A = \underset{d\leq D_A}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_B = \underset{d>D_A}{\text{col}}(\mathbf{X}_d)$ and $\mathbf{X}_d = \underset{1\leq t\leq m_d}{\text{col}'}(\mathbf{x}_{dt})$, $\beta = \beta_{p\times 1}$, $\mathbf{Z} = \mathbf{I}_M$ and $M = \sum_{d=1}^D m_d$. In this notation, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \text{var}(\mathbf{u}) = \text{diag}(\sigma_A^2\Omega_A, \sigma_B^2\Omega_B), \quad \mathbf{V}_e = \text{var}(\mathbf{e}) = \underset{1\leq d\leq D}{\text{diag}}(\mathbf{V}_{ed})$$

where $\Omega_A = \underset{d \le D_A}{\text{diag}}(\Omega_d)$, $\Omega_B = \underset{d > D_A}{\text{diag}}(\Omega_d)$, $\mathbf{V}_{ed} = \underset{1 \le t \le m_d}{\text{diag}}(\sigma_{dt}^2)$ and

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \dots & \rho & 1 \end{pmatrix}_{m_d \times m_d}.$$

The covariance matrix of vector $\mathbf{y}$ is $\mathbf{V} = \text{var}(\mathbf{y}) = \text{diag}(\mathbf{V}_A, \mathbf{V}_B)$, where $\mathbf{V}_A = \underset{d \le D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_B = \underset{d > D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_d = \sigma_A^2 \Omega_d + \mathbf{V}_{ed}$ if $d \le D_A$ and $\mathbf{V}_d = \sigma_B^2 \Omega_d + \mathbf{V}_{ed}$ if $d > D_A$.

If $\sigma_A^2 > 0$, $\sigma_B^2 > 0$ and ρ are known, the best linear unbiased estimator (BLUE) of β is

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

and the best linear unbiased predictor (BLUP) of $\mathbf{u}$ is

$$\widehat{\mathbf{u}} = \mathbf{V}_u \mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}) = \text{diag}\left(\sigma_A^2 \underset{d \le D_A}{\text{diag}}(\Omega_d), \sigma_B^2 \underset{d > D_A}{\text{diag}}(\Omega_d)\right) \underset{1 \le d \le D}{\text{col}}(\mathbf{V}_d^{-1})(\mathbf{y} - \mathbf{X}\widehat{\beta}),$$

so that

$$\widehat{\mathbf{u}}_d = \begin{cases} \sigma_A^2 \Omega_d \mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta}), & d = 1, \dots, D_A, \\ \sigma_B^2 \Omega_d \mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta}), & d = D_A + 1, \dots, D. \end{cases}$$

The loglikelihood of the restricted (residual) maximum likelihood method is

$$\begin{aligned} l_{reml} &= l_{reml}(\sigma_A^2, \sigma_B^2, \rho) = -\frac{M-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}'\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}_A| - \frac{1}{2}\log|\mathbf{V}_B| \\ &- \frac{1}{2}\log|\mathbf{X}_A'\mathbf{V}_A^{-1}\mathbf{X}_A + \mathbf{X}_B'\mathbf{V}_B^{-1}\mathbf{X}_B| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}, \end{aligned}$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \quad \mathbf{PVP} = \mathbf{P}, \quad \mathbf{PX} = \mathbf{0}.$$

Let $\theta = (\theta_1, \theta_2, \theta_3) = (\sigma_A^2, \sigma_B^2, \rho)$, then

$$\begin{aligned} \mathbf{V}_1 &= \frac{\partial \mathbf{V}}{\partial \sigma_A^2} = \text{diag}\left(\underset{d \le D_A}{\text{diag}}(\Omega_d(\rho)), \underset{d > D_A}{\text{diag}}(\mathbf{0}_{m_d \times m_d})\right) \\ \mathbf{V}_2 &= \frac{\partial \mathbf{V}}{\partial \sigma_B^2} = \text{diag}\left(\underset{d \le D_A}{\text{diag}}(\mathbf{0}_{m_d \times m_d}), \underset{d > D_A}{\text{diag}}(\Omega_d(\rho))\right), \\ \mathbf{V}_3 &= \frac{\partial \mathbf{V}}{\partial \rho} = \text{diag}\left(\sigma_A^2 \underset{d \le D_A}{\text{diag}}(\dot{\Omega}_d(\rho)), \sigma_B^2 \underset{d > D_A}{\text{diag}}(\dot{\Omega}_d(\rho))\right), \end{aligned}$$

where $\dot{\Omega}(\rho) = \partial\Omega(\rho)/\partial\rho$. Then

$$\mathbf{P}_a = \frac{\partial \mathbf{P}}{\partial \theta_a} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_a}\mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P}, \quad a = 1,2,3.$$

By taking partial derivatives of l_{reml} with respect to θ_a, we get the scores

$$S_a = \frac{\partial l_{reml}}{\partial \theta_a} = -\frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_a) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y}, \quad a = 1,2,3.$$

By taking again partial derivatives with respect to θ_a and θ_b, taking expectations and changing the sign, we get the Fisher information matrix components

$$F_{ab} = \frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a,b = 1,2,3.$$

To calculate the REML estimate we apply the Fisher-scoring algorithm with the updating fórmula

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k),$$

where $\mathbf{S}$ and $\mathbf{F}$ are the column vector of scores and the Fisher information matrix respectively. As seeds we use $\rho^{(0)} = 0$, and $\sigma_A^{2(0)} = \sigma_B^{2(0)} = \widehat{\sigma}_{uH}^2$, where $\widehat{\sigma}_{uH}^2$ is the Henderson 3 estimator under model with $\rho = 0$ and $\sigma_A^2 = \sigma_B^2$. The REML estimator of β is

$$\widehat{\beta}_{reml} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of the REML estimators of θ and β are

$$\hat{\theta} \sim N_3(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}\nu_{aa}^{1/2},\ a = 1,2,3, \quad \hat{\beta}_j \pm z_{\alpha/2}q_{jj}^{1/2},\ j = 1,\dots,p,$$

where $\hat{\theta} = \theta^{\kappa}$, $\mathbf{F}^{-1}(\theta^{\kappa}) = (\nu_{ab})_{a,b=1,2,3}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^{\kappa})\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\dots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$). Observed $\hat{\beta}_j = \beta_0$, the p-value for testing the hypothesis $H_0 : \beta_j = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_j > |\beta_0|) = 2P(N(0,1) > \beta_0/\sqrt{q_{jj}}).$$

In what follows we present some matrix calculation that are useful to implement the Fisher-scoring algorithm. The target here is to avoid calculations of $M\times M$ matrices.

$$
\begin{aligned}
\mathbf{Q} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1},\\
\mathbf{P} &= \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}) - \underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}),\\
\mathbf{P}\mathbf{V}_a &= \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}),\\
\text{tr}(\mathbf{P}\mathbf{V}_a) &= \sum_{d=1}^{D}\text{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \sum_{d=1}^{D}\text{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q}),\\
\text{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b) &= \text{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\right\}\\
&\text{if}\quad a,b=1,2\quad\text{with}\quad a\neq b;\quad\text{otherwise}\\
\text{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b) &= \sum_{d=1}^{D}\text{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}) - 2\sum_{d=1}^{D}\text{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q})\\
&+ \text{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y} &= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'\\
&- \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d\right)\\
&+ \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'.
\end{aligned}
$$

Finally, the derivative of matrix $\Omega_d(\rho)$ with respect to ρ is

$$
\dot{\Omega}_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix}
0 & 1 & \dots & \dots & (m_d-1)\rho^{m_d-2}\\
1 & 0 & \ddots & & (m_d-2)\rho^{m_d-3}\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
(m_d-2)\rho^{m_d-3} & & \ddots & 0 & 1\\
(m_d-1)\rho^{m_d-2} & \dots & \dots & 1 & 0
\end{pmatrix} + \frac{2\rho\Omega_d(\rho)}{1-\rho^2}.
$$

7.4.2 The mean squared error of the EBLUP

We are interested in predicting the value of $\mu_{dt}=\mathbf{x}_{dt}\beta+u_{dt}$ by using the EBLUP $\widehat{\mu}_{dt}=\mathbf{x}_{dt}\widehat{\beta}+\widehat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt}=\mathbf{a}'\mathbf{y}$, where $\mathbf{a}=\underset{1\le\ell\le D}{\text{col}}(\underset{1\le k\le m_\ell}{\text{col}}(\delta_{d\ell}\delta_{tk}))$ is a vector having one 1 in the position $t+\sum_{\ell=1}^{d-1}m_\ell$ and 0's in the remaining cells. To estimate $\overline{Y}_{dt}$ we use $\widehat{\overline{Y}}_{dt}^{eblup} = \widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where $\theta = (\sigma_A^2, \sigma_B^2, \rho)$,

$$\begin{aligned} g_1(\theta) &= \mathbf{a}'\mathbf{ZTZ}'\mathbf{a}, \\ g_2(\theta) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{ZTZ}'\mathbf{a}], \\ g_3(\theta) &\approx \operatorname{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\} \end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}).$$

Calculation of $g_1(\theta)$

In the formula of $g_1(\theta) = \mathbf{a}'\mathbf{ZTZ}'\mathbf{a}$, we have that $\mathbf{Z} = \mathbf{I}_M$, and $\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \operatorname{diag}(\mathbf{T}_A, \mathbf{T}_B)$, where

$$\mathbf{T}_A = \sigma_A^2 \operatorname*{diag}_{d\le D_A}(\Omega_d) - \sigma_A^4 \operatorname*{diag}_{d\le D_A}(\Omega_d)\operatorname*{diag}_{d\le D_A}(\mathbf{V}_d^{-1})\operatorname*{diag}_{d\le D_A}(\Omega_d)$$

and

$$\mathbf{T}_B = \sigma_B^2 \operatorname*{diag}_{d> D_A}(\Omega_d) - \sigma_B^4 \operatorname*{diag}_{d> D_A}(\Omega_d)\operatorname*{diag}_{d> D_A}(\mathbf{V}_d^{-1})\operatorname*{diag}_{d> D_A}(\Omega_d)$$

Let us write $\mathbf{a}_d = \operatorname*{col}_{1\le k\le m_d}(\delta_{tk})$. Then, $g_1(\theta)$ can be expressed in the form

$$g_1(\theta) = \begin{cases} \sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{a}_d - \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d & \text{if } d \le D_A, \\ \sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{a}_d - \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d & \text{if } d > D_A. \end{cases}$$

Calculation of $g_2(\theta)$

We have that $g_2(\theta) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{ZTZ}'\mathbf{a}]$, where $\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} =$

$$\begin{aligned} \mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} &= \begin{pmatrix} \left[\sigma_A^2 \operatorname*{diag}_{d\le D_A}(\Omega_d) - \sigma_A^4 \operatorname*{diag}_{d\le D_A}(\Omega_d)\operatorname*{diag}_{d\le D_A}(\mathbf{V}_d^{-1})\operatorname*{diag}_{d\le D_A}(\Omega_d)\right]\operatorname*{diag}_{d\le D_A}(\mathbf{V}_{ed}^{-1})\operatorname*{col}_{d\le D_A}(\mathbf{X}_d) \\ \left[\sigma_B^2 \operatorname*{diag}_{d> D_A}(\Omega_d) - \sigma_B^4 \operatorname*{diag}_{d> D_A}(\Omega_d)\operatorname*{diag}_{d> D_A}(\mathbf{V}_d^{-1})\operatorname*{diag}_{d> D_A}(\Omega_d))\right]\operatorname*{diag}_{d> D_A}(\mathbf{V}_{ed}^{-1})\operatorname*{col}_{d> D_A}(\mathbf{X}_d) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_A^2 \operatorname*{col}_{d\le D_A}(\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_A^4 \operatorname*{col}_{d\le D_A}(\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \\ \sigma_B^2 \operatorname*{col}_{d> D_A}(\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_B^4 \operatorname*{col}_{d> D_A}(\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \end{pmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} g_2(\theta) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_A^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d + \sigma_A^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\right] \quad \text{if } d \le D_A, \\ &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_B^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d + \sigma_B^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\right] \quad \text{if } d > D_A. \end{aligned}$$

Calculation of $g_3(\theta)$

We have that

$$g_3(\theta) \approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where $\mathbf{b}'=\mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}=\mathbf{a}'\text{diag}\big(\sigma_A^2\underset{\ell\le D_A}{\text{diag}}(\Omega_\ell),\sigma_B^2\underset{\ell>D_A}{\text{diag}}(\Omega_\ell)\big)\underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell^{-1})=(\mathbf{b}_A',\mathbf{b}_B')$,

$$\mathbf{b}_A'=\sigma_A^2\underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}) \quad\text{and}\quad \mathbf{b}_B'=\sigma_B^2\underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}).$$

It holds that $\frac{\partial\mathbf{b}'}{\partial\sigma_A^2}=(\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2},\mathbf{0})$, $\frac{\partial\mathbf{b}'}{\partial\sigma_B^2}=(\mathbf{0},\frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2})$, $\frac{\partial\mathbf{b}'}{\partial\rho}=(\frac{\partial\mathbf{b}_A'}{\partial\rho},\frac{\partial\mathbf{b}_B'}{\partial\rho})$, where

$$\begin{aligned}
\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2} &= \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1})-\sigma_A^2\underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell A}\mathbf{V}_\ell^{-1}),\quad \mathbf{V}_{\ell A}=\frac{\partial\mathbf{V}_\ell}{\partial\sigma_A^2}=\Omega_\ell,\\
\frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2} &= \underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1})-\sigma_B^2\underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell B}\mathbf{V}_\ell^{-1}),\quad \mathbf{V}_{\ell B}=\frac{\partial\mathbf{V}_\ell}{\partial\sigma_B^2}=\Omega_\ell,\\
\frac{\partial\mathbf{b}_A'}{\partial\rho} &= \sigma_A^2\underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\dot{\Omega}_\ell\mathbf{V}_\ell^{-1})-\sigma_A^4\underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell\rho}\mathbf{V}_\ell^{-1}),\quad \mathbf{V}_{\ell\rho}=\frac{\partial\mathbf{V}_\ell}{\partial\rho}=\sigma_A^2\dot{\Omega}_\ell,\\
\frac{\partial\mathbf{b}_B'}{\partial\rho} &= \sigma_B^2\underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\dot{\Omega}_\ell\mathbf{V}_\ell^{-1})-\sigma_B^4\underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell\rho}\mathbf{V}_\ell^{-1}),\quad \mathbf{V}_{\ell\rho}=\frac{\partial\mathbf{V}_\ell}{\partial\rho}=\sigma_B^2\dot{\Omega}_\ell.
\end{aligned}$$

We define $q_{21}=q_{12}=0$, $q_{31}=q_{13}$, $q_{32}=q_{23}$,

$$\begin{aligned}
q_{11} &= \frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}\underset{\ell\le D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}\right)'=\big[\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d-2\sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\\
&+ \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big]I_{\{d\le D_A\}}(d),\\
q_{22} &= \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2}\underset{\ell>D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2}\right)'=\big[\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d-2\sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\\
&+ \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big]I_{\{d>D_A\}}(d),\\
q_{33} &= \frac{\partial\mathbf{b}'}{\partial\rho}\underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\rho}\right)'=\big[\sigma_A^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d-2\sigma_A^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d\\
&+ \sigma_A^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big]I_{\{d\le D_A\}}(d)\\
&+ \big[\sigma_B^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d-2\sigma_B^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d\\
&+ \sigma_B^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big]I_{\{d>D_A\}}(d).\\
q_{13} &= \frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}\underset{\ell\le D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_A'}{\partial\rho_A}\right)'=\big[\sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d-\sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\\
&- \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d+\sigma_A^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big]I_{\{d\le D_A\}}(d),\\
q_{23} &= \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2}\underset{\ell>D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_B'}{\partial\rho_B}\right)'=\big[\sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d-\sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\\
&- \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d+\sigma_B^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big]I_{\{d>D_A\}}(d),
\end{aligned}$$

Finally

$$g_3(\theta) = \begin{cases} \mathrm{tr}\Big\{ \begin{pmatrix} q_{11} & q_{13} \\ q_{31} & q_{33} \end{pmatrix} \begin{pmatrix} F_{11} & F_{13} \\ F_{31} & F_{33} \end{pmatrix}^{-1} \Big\}, & \text{if } d \le D_A \\ \mathrm{tr}\Big\{ \begin{pmatrix} q_{22} & q_{23} \\ q_{32} & q_{33} \end{pmatrix} \begin{pmatrix} F_{22} & F_{23} \\ F_{32} & F_{33} \end{pmatrix}^{-1} \Big\}, & \text{if } d > D_A, \end{cases}$$

where F_{ab} is the element of the REML Fisher information matrix.

7.4.3 testing for $H_0 : \rho = 0$

Let $\hat{\sigma}_A^2$, $\hat{\sigma}_B^2$ and $\hat{\rho}$ be the unrestricted REML estimators of σ_A^2, σ_B^2 and ρ respectively. Let $\tilde{\sigma}_A^2$ and $\tilde{\sigma}_B^2$ be the REML estimator of σ_A^2 and σ_B^2 under H_0. The REML likelihood ratio statistic (LRS) for testing $H_0 : \rho = 0$ is

$$\lambda = -2[l_{REML}(\tilde{\sigma}_A^2, \tilde{\sigma}_B^2) - l_{REML}(\hat{\sigma}_A^2, \hat{\sigma}_B^2, \hat{\rho})] = \log \frac{|\tilde{V}|}{|\hat{V}|} + \log \frac{|X'\tilde{V}^{-1}X|}{|X'\hat{V}^{-1}X|} + y'\tilde{P}y - y'\hat{P}y.$$

Asymptotic distribution of λ under H_0 is χ_1^2, so null hypothesis is rejected at the level α if $\lambda > \chi_{1,\alpha}^2$.

7.5 Partitioned Fay-Herriot model 3

7.5.1 The model

Let us consider the model (model 3)

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, \dots, D = D_A + D_B, \quad t = 1, \dots, m_d, \tag{7.9}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We assume that the random vectors $(u_{d1}, \dots, u_{dm_d})$, $d \le D_A$, follow i.i.d. first order auto-regressive processes with variance and auto-correlation parameters σ_A^2 and ρ_A respectively; in short, $(u_{d1}, \dots, u_{dm_d}) \sim_{iid} AR1(\sigma_A^2, \rho_A)$, $d \le D_A$. We further assume that $(u_{d1}, \dots, u_{dm_d}) \sim_{iid} AR1(\sigma_B^2, \rho_B)$, $d > D_A$, and that the errors e_{dt}'s are independent $N(0, \sigma_{dt}^2)$ with known σ_{dt}^2's. Finally we assume that the $(u_{d1}, \dots, u_{dm_d})$'s and the e_{dt}'s are mutually independent.

In matrix notation the model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where vectors $\mathbf{y}$, $\mathbf{u}$ and $\mathbf{e}$ can be decomposed in the form $\mathbf{v} = (\mathbf{v}_A', \mathbf{v}_B')'$, with $\mathbf{v}_A = \underset{d \le D_A}{\mathrm{col}}(\mathbf{v}_d)$, $\mathbf{v}_B = \underset{d > D_A}{\mathrm{col}}(\mathbf{v}_d)$ and $\mathbf{v}_d = \underset{1 \le t \le m_d}{\mathrm{col}}(v_{dt})$, matrix $\mathbf{X}$ can be similarly decomposed in the form $\mathbf{X} = (\mathbf{X}_A', \mathbf{X}_B')'$, with $\mathbf{X}_A = \underset{d \le D_A}{\mathrm{col}}(\mathbf{X}_d)$, $\mathbf{X}_B = \underset{d > D_A}{\mathrm{col}}(\mathbf{X}_d)$ and $\mathbf{X}_d = \underset{1 \le t \le m_d}{\mathrm{col}'}(\mathbf{x}_{dt})$, $\beta = \beta_{p \times 1}$, $\mathbf{Z} = \mathbf{I}_M$ and $M = \sum_{d=1}^D m_d$. In this notation, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \mathrm{var}(\mathbf{u}) = \mathrm{diag}(\sigma_A^2 \Omega_A, \sigma_B^2 \Omega_B), \quad \mathbf{V}_e = \mathrm{var}(\mathbf{e}) = \underset{1 \le d \le D}{\mathrm{diag}}(\mathbf{V}_{ed})$$

where $\Omega_A = \underset{d\le D_A}{\text{diag}}(\Omega_d)$, $\Omega_B = \underset{d>D_A}{\text{diag}}(\Omega_d)$, $\mathbf{V}_{ed} = \underset{1\le t\le m_d}{\text{diag}}(\sigma_{dt}^2)$ and

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \dots & \rho & 1 \end{pmatrix}_{m_d\times m_d}, \quad \rho = \rho_A, \rho_B.$$

The covariance matrix of vector $\mathbf{y}$ is $\mathbf{V} = \text{var}(\mathbf{y}) = \text{diag}(\mathbf{V}_A, \mathbf{V}_B)$, where $\mathbf{V}_A = \underset{d\le D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_B = \underset{d>D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_d = \sigma_A^2\Omega_d + \mathbf{V}_{ed}$ if $d \le D_A$ and $\mathbf{V}_d = \sigma_B^2\Omega_d + \mathbf{V}_{ed}$ if $d > D_A$.

If $\sigma_A^2 > 0$, ρ_A, $\sigma_B^2 > 0$ and ρ_B are known, the best linear unbiased estimator (BLUE) of β is

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

and the best linear unbiased predictor (BLUP) of $\mathbf{u}$ is

$$\widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\beta}) = \text{diag}\left(\sigma_A^2\underset{d\le D_A}{\text{diag}}(\Omega_d), \sigma_B^2\underset{d>D_A}{\text{diag}}(\Omega_d)\right)\underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1})(\mathbf{y}-\mathbf{X}\widehat{\beta}),$$

so that

$$\widehat{\mathbf{u}}_d = \begin{cases} \sigma_A^2\Omega_d\mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta}), & d = 1,\dots,D_A, \\ \sigma_B^2\Omega_d\mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta}), & d = D_A+1,\dots,D. \end{cases}$$

The loglikelihood of the restricted (residual) maximum likelihood method is

$$\begin{aligned} l_{reml} &= l_{reml}(\sigma_A^2, \rho_A, \sigma_B^2, \rho_B) = -\frac{M-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}'\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}_A| - \frac{1}{2}\log|\mathbf{V}_B| \\ &- \frac{1}{2}\log|\mathbf{X}_A'\mathbf{V}_A^{-1}\mathbf{X}_A + \mathbf{X}_B'\mathbf{V}_B^{-1}\mathbf{X}_B| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}, \end{aligned}$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \quad \mathbf{PVP} = \mathbf{P}, \quad \mathbf{PX} = \mathbf{0}.$$

Let $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\sigma_A^2, \rho_A, \sigma_B^2, \rho_B)$, then

$$\begin{aligned} \mathbf{V}_1 &= \frac{\partial\mathbf{V}}{\partial\sigma_A^2} = \text{diag}\left(\underset{d\le D_A}{\text{diag}}(\Omega_d(\rho_A)), \underset{d>D_A}{\text{diag}}(\mathbf{0}_{m_d\times m_d})\right) \\ \mathbf{V}_2 &= \frac{\partial\mathbf{V}}{\partial\rho_A} = \text{diag}\left(\sigma_A^2\underset{d\le D_A}{\text{diag}}(\dot{\Omega}_d(\rho_A)), \underset{d>D_A}{\text{diag}}(\mathbf{0}_{m_d\times m_d})\right), \\ \mathbf{V}_3 &= \frac{\partial\mathbf{V}}{\partial\sigma_B^2} = \text{diag}\left(\underset{d\le D_A}{\text{diag}}(\mathbf{0}_{m_d\times m_d}), \underset{d>D_A}{\text{diag}}(\Omega_d(\rho_B))\right), \\ \mathbf{V}_4 &= \frac{\partial\mathbf{V}}{\partial\rho_B} = \text{diag}\left(\underset{d\le D_A}{\text{diag}}(\mathbf{0}_{m_d\times m_d}), \sigma_B^2\underset{d>D_A}{\text{diag}}(\dot{\Omega}_d(\rho_B))\right). \end{aligned}$$

where $\dot{\Omega}(\rho) = \partial\Omega(\rho)/\partial\rho$. Then

$$\mathbf{P}_a = \frac{\partial \mathbf{P}}{\partial \theta_a} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_a}\mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P}, \quad a = 1,2,3,4.$$

By taking partial derivatives of l_{reml} with respect to θ_a, we get the scores

$$S_a = \frac{\partial l_{reml}}{\partial \theta_a} = -\frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_a) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y}, \quad a = 1,2,3,4.$$

By taking again partial derivatives with respect to θ_a and θ_b, taking expectations and changing the sign, we get the Fisher information matrix components

$$F_{ab} = \frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a,b = 1,2,3,4.$$

To calculate the REML estimate we apply the Fisher-scoring algorithm with the updating fórmula

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k),$$

where $\mathbf{S}$ and $\mathbf{F}$ are the column vector of scores and the Fisher information matrix respectively. As seeds we use $\rho_A^{(0)} = \rho_B^{(0)} = 0$, and $\sigma_A^{2(0)} = \sigma_B^{2(0)} = \widehat{\sigma}_{uH}^2$, where $\widehat{\sigma}_{uH}^2$ is the Henderson 3 estimator under model with $\rho_A = \rho_B = 0$ and $\sigma_A^2 = \sigma_B^2$. The REML estimator of β is

$$\widehat{\beta}_{reml} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of of the REML estimators of θ and β are

$$\hat{\theta} \sim N_4(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}\,\nu_{aa}^{1/2},\ a = 1,2,3,4, \quad \hat{\beta}_j \pm z_{\alpha/2}\,q_{jj}^{1/2},\ j = 1,\ldots,p,$$

where $\hat{\theta} = \theta^\kappa$, $\mathbf{F}^{-1}(\theta^\kappa) = (\nu_{ab})_{a,b=1,2,3,4}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^\kappa)\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\ldots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$). Observed $\hat{\beta}_j = \beta_0$, the p-value for testing the hypothesis $H_0 : \beta_j = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_j > |\beta_0|) = 2P(N(0,1) > \beta_0/\sqrt{q_{jj}}).$$

In what follows we present some matrix calculation that are useful to implement the Fisher-scoring algorithm. The target here is to avoid calculations of $M\times M$ matrices.

$$
\begin{aligned}
\mathbf{Q} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1},\\
\mathbf{P} &= \underset{1\le d\le D}{\operatorname{diag}}(\mathbf{V}_d^{-1}) - \underset{1\le d\le D}{\operatorname{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\operatorname{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}),\\
\mathbf{PV}_a &= \underset{1\le d\le D}{\operatorname{diag}}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \underset{1\le d\le D}{\operatorname{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\operatorname{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}),\\
\operatorname{tr}(\mathbf{PV}_a) &= \sum_{d=1}^{D}\operatorname{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \sum_{d=1}^{D}\operatorname{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q}),\\
\operatorname{tr}(\mathbf{PV}_a\mathbf{PV}_b) &= \operatorname{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\right\}\\
&\text{with}\quad a=1,2 \quad\text{and}\quad b=3,4, \quad\text{otherwise}
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{tr}(\mathbf{PV}_a\mathbf{PV}_b) &= \sum_{d=1}^{D}\operatorname{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}) - 2\sum_{d=1}^{D}\operatorname{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q})\\
&+ \operatorname{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\right\},\\
\mathbf{y}'\mathbf{PV}_a\mathbf{Py} &= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'\\
&- \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d\right)\\
&+ \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'.
\end{aligned}
$$

Finally, the derivative of matrix $\Omega_d(\rho)$ with respect to ρ is

$$
\dot{\Omega}_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix}
0 & 1 & \dots & \dots & (m_d-1)\rho^{m_d-2}\\
1 & 0 & \ddots & & (m_d-2)\rho^{m_d-3}\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
(m_d-2)\rho^{m_d-3} & & \ddots & 0 & 1\\
(m_d-1)\rho^{m_d-2} & \dots & \dots & 1 & 0
\end{pmatrix} + \frac{2\rho\Omega_d(\rho)}{1-\rho^2}.
$$

7.5.2 The mean squared error of the EBLUP

We are interested in predicting the value of $\mu_{dt}=\mathbf{x}_{dt}\beta+u_{dt}$ by using the EBLUP $\widehat{\mu}_{dt}=\mathbf{x}_{dt}\widehat{\beta}+\widehat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt}=\mathbf{a}'\mathbf{y}$, where $\mathbf{a}=\underset{1\le\ell\le D}{\operatorname{col}}(\underset{1\le k\le m_\ell}{\operatorname{col}}(\delta_{d\ell}\delta_{tk}))$ is a vector having one 1 in the position $t+\sum_{\ell=1}^{d-1}m_\ell$ and 0's in the remaining cells. To estimate $\overline{Y}_{dt}$ we use $\widehat{\overline{Y}}_{dt}^{eblup} = \widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where $\theta = (\sigma_A^2, \rho_A, \sigma_B^2, \rho_B)$,

$$\begin{aligned} g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\ g_2(\theta) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}], \\ g_3(\theta) &\approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\} \end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}).$$

Calculation of $g_1(\theta)$

In the formula of $g_1(\theta) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, we have that $\mathbf{Z} = \mathbf{I}_M$, and $\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \text{diag}(\mathbf{T}_A, \mathbf{T}_B)$, where

$$\mathbf{T}_A = \sigma_A^2 \underset{d\le D_A}{\text{diag}}(\Omega_d) - \sigma_A^4 \underset{d\le D_A}{\text{diag}}(\Omega_d) \underset{d\le D_A}{\text{diag}}(\mathbf{V}_d^{-1}) \underset{d\le D_A}{\text{diag}}(\Omega_d)$$

and

$$\mathbf{T}_B = \sigma_B^2 \underset{d> D_A}{\text{diag}}(\Omega_d) - \sigma_B^4 \underset{d> D_A}{\text{diag}}(\Omega_d) \underset{d> D_A}{\text{diag}}(\mathbf{V}_d^{-1}) \underset{d> D_A}{\text{diag}}(\Omega_d)$$

Let us write $\mathbf{a}_d = \underset{1\le k\le m_d}{\text{col}}(\delta_{tk})$. Then, $g_1(\theta)$ can be expressed in the form

$$g_1(\theta) = \begin{cases} \sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{a}_d - \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d & \text{if } d \le D_A, \\ \sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{a}_d - \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d & \text{if } d > D_A. \end{cases}$$

Calculation of $g_2(\theta)$

We have that $g_2(\theta) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}]$, where $\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X} =$

$$\begin{aligned} \mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X} &= \begin{pmatrix} \left[\sigma_A^2 \underset{d\le D_A}{\text{diag}}(\Omega_d) - \sigma_A^4 \underset{d\le D_A}{\text{diag}}(\Omega_d) \underset{d\le D_A}{\text{diag}}(\mathbf{V}_d^{-1}) \underset{d\le D_A}{\text{diag}}(\Omega_d)\right] \underset{d\le D_A}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{d\le D_A}{\text{col}}(\mathbf{X}_d) \\ \left[\sigma_B^2 \underset{d> D_A}{\text{diag}}(\Omega_d) - \sigma_B^4 \underset{d> D_A}{\text{diag}}(\Omega_d) \underset{d> D_A}{\text{diag}}(\mathbf{V}_d^{-1}) \underset{d> D_A}{\text{diag}}(\Omega_d))\right] \underset{d> D_A}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{d> D_A}{\text{col}}(\mathbf{X}_d) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_A^2 \underset{d\le D_A}{\text{col}}(\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_A^4 \underset{d\le D_A}{\text{col}}(\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \\ \sigma_B^2 \underset{d> D_A}{\text{col}}(\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_B^4 \underset{d> D_A}{\text{col}}(\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \end{pmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} g_2(\theta) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_A^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d + \sigma_A^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\right] \quad \text{if } d \le D_A, \\ &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_B^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d + \sigma_B^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\right] \quad \text{if } d > D_A. \end{aligned}$$

Calculation of $g_3(\theta)$

We have that

$$g_3(\theta) \approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where $\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \mathbf{a}'\text{diag}\big(\sigma_A^2 \underset{\ell\le D_A}{\text{diag}}(\Omega_\ell), \sigma_B^2 \underset{\ell>D_A}{\text{diag}}(\Omega_\ell)\big) \underset{1\le\ell\le D}{\text{diag}}(\mathbf{V}_\ell^{-1}) = (\mathbf{b}_A', \mathbf{b}_B')$,

$$\mathbf{b}_A' = \sigma_A^2 \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}) \quad \text{and} \quad \mathbf{b}_B' = \sigma_B^2 \underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}).$$

It holds that $\frac{\partial\mathbf{b}'}{\partial\sigma_A^2} = (\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}, \mathbf{0})$, $\frac{\partial\mathbf{b}'}{\partial\rho_A} = (\frac{\partial\mathbf{b}_A'}{\partial\rho_A}, \mathbf{0})$, $\frac{\partial\mathbf{b}'}{\partial\sigma_B^2} = (\mathbf{0}, \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2})$, $\frac{\partial\mathbf{b}'}{\partial\rho_B} = (\mathbf{0}, \frac{\partial\mathbf{b}_B'}{\partial\rho_B})$, where

$$\begin{aligned}
\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2} &= \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}) - \sigma_A^2 \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell A}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell A} = \frac{\partial\mathbf{V}_\ell}{\partial\sigma_A^2} = \Omega_\ell, \\
\frac{\partial\mathbf{b}_A'}{\partial\rho_A} &= \sigma_A^2 \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\dot{\Omega}_\ell\mathbf{V}_\ell^{-1}) - \sigma_A^4 \underset{\ell\le D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell\rho_A}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell\rho_A} = \frac{\partial\mathbf{V}_\ell}{\partial\rho_A} = \sigma_A^2\dot{\Omega}_\ell, \\
\frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2} &= \underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}) - \sigma_B^2 \underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell B}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell B} = \frac{\partial\mathbf{V}_\ell}{\partial\sigma_B^2} = \Omega_\ell, \\
\frac{\partial\mathbf{b}_B'}{\partial\rho_B} &= \sigma_B^2 \underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\dot{\Omega}_\ell\mathbf{V}_\ell^{-1}) - \sigma_B^4 \underset{\ell>D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell\rho_B}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell\rho_B} = \frac{\partial\mathbf{V}_\ell}{\partial\rho_B} = \sigma_B^2\dot{\Omega}_\ell.
\end{aligned}$$

We define $q_{21} = q_{12}$,

$$\begin{aligned}
q_{11} &= \frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2} \underset{\ell\le D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}\right)' = \big[\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d - 2\sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&+ \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big] I_{\{d\le D_A\}}(d), \\
q_{12} &= \frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2} \underset{\ell\le D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_A'}{\partial\rho_A}\right)' = \big[\sigma_A^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&- \sigma_A^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_A^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big] I_{\{d\le D_A\}}(d), \\
q_{22} &= \frac{\partial\mathbf{b}_A'}{\partial\rho_A} \underset{\ell\le D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_A'}{\partial\rho_A}\right)' = \big[\sigma_A^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - \sigma_A^6\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&- \sigma_A^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_A^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big] I_{\{d\le D_A\}}(d).
\end{aligned}$$

Similarly, we define $q_{43} = q_{34}$,

$$\begin{aligned}
q_{33} &= \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2} \underset{\ell>D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2}\right)' = \big[\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d - 2\sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&+ \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big] I_{\{d>D_A\}}(d), \\
q_{34} &= \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2} \underset{\ell>D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_B'}{\partial\rho_B}\right)' = \big[\sigma_B^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&- \sigma_B^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_B^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big] I_{\{d>D_A\}}(d), \\
q_{44} &= \frac{\partial\mathbf{b}_B'}{\partial\rho_B} \underset{\ell>D_A}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}_B'}{\partial\rho_B}\right)' = \sigma_B^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - \sigma_B^6\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&- \sigma_B^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_B^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\big] I_{\{d>D_A\}}(d).
\end{aligned}$$

Finally

$$g_3(\theta) = \begin{cases} \mathrm{tr}\Big\{ \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}^{-1} \Big\}, & \text{if } d \le D_A \\ \mathrm{tr}\Big\{ \begin{pmatrix} q_{33} & q_{34} \\ q_{43} & q_{44} \end{pmatrix} \begin{pmatrix} F_{33} & F_{34} \\ F_{43} & F_{44} \end{pmatrix}^{-1} \Big\}, & \text{if } d > D_A, \end{cases}$$

where F_{ab} is the element of the REML Fisher information matrix.

7.5.3 testing for $H_0 : \rho_A = \rho_B$

Let $\hat{\sigma}_A^2$, $\hat{\sigma}_B^2$, $\hat{\rho}_A$ and $\hat{\rho}_B$ be the unrestricted REML estimators of σ_A^2 and σ_B^2 ρ_A and ρ_B respectively. Let $\tilde{\sigma}_A^2$, $\tilde{\sigma}_B^2$ and $\tilde{\rho}$ be the REML estimator of σ_A^2, σ_B^2 and of the common value $\rho_A = \rho_B$ under H_0. The REML likelihood ratio statistic (LRS) for testing $H_0 : \rho_A = \rho_B$ is

$$\lambda = -2[l_{REML}(\tilde{\sigma}_A^2, \tilde{\sigma}_B^2, \tilde{\rho}) - l_{REML}(\hat{\sigma}_A^2, \hat{\sigma}_B^2, \hat{\rho}_A, \hat{\rho}_B)] = \log \frac{|\tilde{V}|}{|\hat{V}|} + \log \frac{|X'\tilde{V}^{-1}X|}{|X'\hat{V}^{-1}X|} + y'\tilde{P}y - y'\hat{P}y.$$

Asymptotic distribution of λ under H_0 is χ_1^2, so null hypothesis is rejected at the level α if $\lambda > \chi_{1,\alpha}^2$.

Chapter 8

Area-level time-space models

8.1 Model 1

8.1.1 Introduction

Let y_{dt} be a direct estimator of the target population parameter and let $\mathbf{x}_{dt}$ be a vector containing the aggregated values of p auxiliary variables. Subindexes d and t are used for domains and time instants respectively. Let us consider the model

$$y_{dt}=\mathbf{x}_{dt}\beta+u_{1d}+u_{2dt}+e_{dt},\quad d=1,\ldots,D,\quad t=1,\ldots,T, \tag{8.1}$$

where $\{u_{1d}\}$, $\{u_{2dt}\}$ y $\{e_{dt}\}$ are independent with distributions $\{u_{1d}\}_{d=1}^{D}\sim SAR(1)$, $\{u_{2dt}\}$ i.i.d $N(0,\sigma_2^2)$ and $e_{dt}\sim N(0,\sigma_{dt}^2)$. Model (8.1) can alternatively written as

$$\mathbf{y}=\mathbf{X}\beta+\mathbf{Z}_1\mathbf{u}_1+\mathbf{Z}_2\mathbf{u}_2+\mathbf{e}, \tag{8.2}$$

where

- $\mathbf{y}=\underset{1\leq d\leq D}{\text{col}}(\underset{1\leq t\leq T}{\text{col}}(y_{dt}))$, $\quad\mathbf{e}=\underset{1\leq d\leq D}{\text{col}}(\underset{1\leq t\leq T}{\text{col}}(e_{dt}))$,
- $\mathbf{u}_1=\underset{1\leq d\leq D}{\text{col}}(u_{1d})$, $\mathbf{u}_2=\underset{1\leq d\leq D}{\text{col}}(\mathbf{u}_{2d})$, $\mathbf{u}_{2d}=\underset{1\leq t\leq T}{\text{col}}(u_{2dt})$,
- $\mathbf{X}=\underset{1\leq d\leq D}{\text{col}}(\underset{1\leq t\leq T}{\text{col}}(\mathbf{x}_{dt}))$, $\mathbf{x}_{dt}=\underset{1\leq j\leq p}{\text{col}'}(x_{dtj})$, $\beta=\underset{1\leq j\leq p}{\text{col}}(\beta_j)$,
- $\mathbf{Z}_1=\underset{1\leq d\leq D}{\text{diag}}(\mathbf{1}_T)$, $\mathbf{Z}_2=\mathbf{I}_{M\times M}$, $M=DT$.

We assume that $\mathbf{u}_1\sim N(\mathbf{0},\mathbf{V}_{u_1})$, $\mathbf{u}_2\sim N(\mathbf{0},\mathbf{V}_{u_2})$ and $\mathbf{e}\sim N(\mathbf{0},\mathbf{V}_e)$ are independent with covariance matrices

$$\begin{aligned}
\mathbf{V}_{u_1} &= \sigma_1^2\Omega_1(\rho_1),\quad \Omega_1(\rho_1)=\left[(\mathbf{I}_D-\rho_1\mathbf{W})'(\mathbf{I}_D-\rho_1\mathbf{W})\right]^{-1}\triangleq\mathbf{C}^{-1}(\rho_1),\\
\mathbf{V}_{u_2} &= \sigma_2^2\mathbf{I}_{DT},\\
\mathbf{V}_e &= \underset{1\leq d\leq D}{\text{diag}}(\mathbf{V}_{ed}),\quad \mathbf{V}_{ed}=\underset{1\leq t\leq T}{\text{diag}}(\sigma_{dt}^2),
\end{aligned}$$

and known σ_{dt}^2's. We assume that the rows of the proximity matrix $\mathbf{W}$ are stochastic vectors, i.e. with components summing up to one. The vector $\mathbf{u}_1$ is distributed according to a stochastic process SAR(1) and the variables $\mathbf{u}_{2dt}$ are i.i.d. normal.

The variance of $\mathbf{y}$ is

$$\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}_1\mathbf{V}_{u_1}\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{V}_{u_2}\mathbf{Z}_2' + \mathbf{V}_e = \mathbf{Z}_1\mathbf{V}_{u_1}\mathbf{Z}_1' + \underset{1\le d\le D}{\mathrm{diag}}(\sigma_2^2\mathbf{I}_T + \mathbf{V}_{ed}).$$

Its inverse can be calculated with the formula

$$(A+CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1},$$

with $A = \underset{1\le d\le D}{\mathrm{diag}}(\sigma_2^2\mathbf{I}_T + \mathbf{V}_{ed})$, $B = \mathbf{V}_{u_1}$, $C = \mathbf{Z}_1$ and $D = \mathbf{Z}_1'$. Then

$$\mathbf{V}^{-1} = A^{-1} - A^{-1}\mathbf{Z}_1(\mathbf{V}_{u_1}^{-1} + \mathbf{Z}_1'A^{-1}\mathbf{Z}_1)^{-1}\mathbf{Z}_1'A^{-1},$$

where $A^{-1} = \underset{1\le d\le D}{\mathrm{diag}}(A_d^{-1})$ and $A_d = \sigma_2^2\mathbf{I}_T + \mathbf{V}_{ed}$. Observe that by applying the above formula we avoid inverting an $M\times M$ matrix, instead we have to invert D matrices of order $T\times T$.

Let us define the parameter $\theta = (\sigma_1^2, \rho_1, \sigma_2^2)$. The formula

$$\frac{\partial \mathbf{C}^{-1}}{\partial \rho_1} = -\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \rho_1}\mathbf{C}^{-1}$$

is used for calculating the partial derivatives of $\mathbf{V}$ with respect to the components of θ, i.e.

$$\begin{aligned}
\mathbf{V}_1 &= \frac{\partial \mathbf{V}}{\partial \sigma_1^2} = \mathbf{Z}_1\frac{\partial \mathbf{V}_{u_1}}{\partial \sigma_1^2}\mathbf{Z}_1' = \mathbf{Z}_1\Omega_1(\rho_1)\mathbf{Z}_1',\\
\mathbf{V}_2 &= \frac{\partial \mathbf{V}}{\partial \rho_1} = \mathbf{Z}_1\frac{\partial \mathbf{V}_{u_1}}{\partial \rho_1}\mathbf{Z}_1' = -\sigma_1^2\mathbf{Z}_1\Omega_1(\rho_1)\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1}\Omega_1(\rho_1)\mathbf{Z}_1',\\
\mathbf{V}_3 &= \frac{\partial \mathbf{V}}{\partial \sigma_2^2} = \mathbf{I}_{DT},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1} &= \frac{\partial \mathbf{C}}{\partial \rho_1} = \frac{\partial}{\partial \rho_1}\left\{(\mathbf{I}_D - \rho_1\mathbf{W})'(\mathbf{I}_D - \rho_1\mathbf{W})\right\}\\
&= -\mathbf{W}' + \rho_1\mathbf{W}'\mathbf{W} - \mathbf{W} + \rho_1\mathbf{W}'\mathbf{W} = -\mathbf{W} - \mathbf{W}' + 2\rho_1\mathbf{W}'\mathbf{W}.
\end{aligned}$$

8.1.2 BLUP

The BLU estimators and predictors of β and $\mathbf{u}$ are

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{y} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}),$$

where $\mathbf{V}_u = \mathrm{diag}(\mathbf{V}_{u_1}, \mathbf{V}_{u_2})$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$. To calculate $\widehat{\mathbf{u}}$ we apply the formula

$$\widehat{\mathbf{u}} = \begin{pmatrix} \mathbf{V}_{u_1}\mathbf{Z}_1' \\ \mathbf{V}_{u_2}\mathbf{Z}_2' \end{pmatrix}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}) = \begin{pmatrix} \sigma_1^2\Omega_1(\rho_1)\mathbf{Z}_1'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}) \\ \sigma_2^2\mathbf{Z}_2'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}) \end{pmatrix}.$$

The BLUP predictor of μ_{dt} is

$$\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{1d} + \widehat{u}_{2dt}.$$

8.1.3 Residual maximum likelihood estimation

For the residual maximum likelihood (REML) estimation method, the log-likelihood is

$$l_{REML}(\theta) = -\frac{M-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}'\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where $\theta = (\theta_1, \theta_2, \theta_3) = (\sigma_1^2, \rho_1, \sigma_2^2)$ and

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \quad \mathbf{PVP} = \mathbf{P}, \quad \mathbf{PX} = \mathbf{0}.$$

Then

$$\mathbf{P}_a = \frac{\partial \mathbf{P}}{\partial \theta_a} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_a}\mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P}, \quad a = 1,2,3.$$

By taking derivatives of l_{REML} with respect to θ_a, we get

$$S_a = \frac{\partial l_{REML}}{\partial \theta_a} = -\frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_a) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y}, \quad a = 1,2,3.$$

By taking again derivatives wit respect to θ_a and θ_b, taking expectations and changing the sign, we have

$$F_{ab} = \frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a,b = 1,2,3.$$

The updating formula of the Fisher-scoring algorithm is

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k).$$

This algorithm requires starting values of θ (seeds). We may obtain seeds by considering the model without $\mathbf{u}_1$ and with $\rho_2 = 0$. For this last model we might consider the Henderson 3 estimator $\widehat{\sigma}^2_{u_2H}$ of the only remaining variance σ_2^2. Therefore, we might propose the following seeds: $\sigma_1^{2(0)} = \sigma_2^{2(0)} = \frac{1}{2}\widehat{\sigma}^2_{u_2H}$, $\rho_1^{(0)} = 0.3$.

The REML estimator of β is

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

The asymptotic distributions of the REML estimators θ and β are

$$\hat{\theta} \sim N_2(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}\nu_{aa}^{1/2},\ a = 1,\dots,3, \quad \hat{\beta}_j \pm z_{\alpha/2}q_{jj}^{1/2},\ j = 1,\dots,p,$$

where $\hat{\theta} = \theta^\kappa$, $\mathbf{F}^{-1}(\theta^\kappa) = (\nu_{ab})_{a,b=1,\dots,3}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^\kappa)\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\dots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantil of the standard normal distribution $N(0,1)$. Observed $\hat{\beta}_j = \beta_0$, the p-value for testing the test of hypothesis $H_0 : \beta_j = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_j > |\beta_0|) = 2P(N(0,1) > \beta_0/\sqrt{q_{jj}}).$$

8.1.4 Simulations

For $d = 1,\dots,D$, $t = 1,\dots,T$, the explanatory and target variables are

$$\begin{aligned} x_{dt} &= (b_{dt}-a_{dt})U_{dt}+a_{dt},\ U_{dt}=\frac{t}{T+1},\ a_{dt}=1,\ b_{dt}=1+\frac{1}{D}\left(T(d-1)+t\right), \\ y_{dt} &= \beta_1+\beta_2 x_{dt}+u_{1d}+u_{2dt}+e_{dt},\ \beta_1=0,\ \beta_2=1, \end{aligned}$$

where $u_{2dt} \sim N(0,\sigma_2^2)$ and $e_{dt} \sim N(0,\sigma_{dt}^2)$ are independent with $\sigma_2^2 = 1$ and

$$\sigma_{dt}^2 = \frac{(\alpha_1-\alpha_0)\left(T(d-1)+t-1\right)}{M-1}+\alpha_0, \quad \alpha_0 = 0.8,\ \alpha_1 = 1.2.$$

The vector $\mathbf{u}_1 = \underset{1\le d\le D}{\text{col}}(u_{1d})$ is generated from the distribution $N_D(0,\sigma_1^2\Omega_1(\rho_1))$, using the proximity matrix

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1/2 & 0 & 1/2 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 1/2 & 0 & 1/2 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}_{D\times D}, \quad \sigma_1^2 = 1, \rho_1 = 0.5 \tag{8.3}$$

Simulation 1a

The steps of the simulation experiment are

1. Do $\beta_1 = 0$, $\beta_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho_1 = 0.5$, definir W según (8.3) and generate σ_{dt}^2, x_{dt}, $d = 1,\dots,D$, $t = 1,\dots,T$.
2. Repeat $K = 4000$ times ($k = 1,\dots,K$)
   2.1. Generate $y_{dt}^{(k)}$ and calculate $\mu_{dt}^{(k)} = \beta_1+\beta_2 x_{dt}+u_{1d}^{(k)}+u_{2dt}^{(k)}$, $d = 1,\dots,D$, $t = 1,\dots,T$.
   2.2. Calculate $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)},\hat{\beta}_2^{(k)},\hat{\sigma}_1^{2(k)},\hat{\rho}_1^{(k)},\hat{\sigma}_2^{2(k)}\}$ and $\hat{\mu}_{dt}^{(k)} = \hat{\beta}_1^{(k)}+\hat{\beta}_2^{(k)}x_{dt}+\hat{u}_{1d}^{(k)}+\hat{u}_{2dt}^{(k)}$, by using the REML method.
3. For each $\tau \in \{\beta_1,\beta_2,\sigma_1^2,\rho_1,\sigma_2^2\}$ and for each $\hat{\mu}_{dt}$, $d = 1,\dots,D$, $t = 1,\dots,T$, calculate

$$BIAS(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)}-\tau), \quad MSE(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)}-\tau)^2.$$

$$BIAS_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)}-\mu_{dt}^{(k)}), \quad MSE_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)}-\mu_{dt}^{(k)})^2,$$

$$BIAS = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T}BIAS_{dt}, \quad MSE = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T}MSE_{dt}.$$

The simulation experiment is repeated for each of the 6 combinations of sample sizes appearing in the Table 8.1.4.1.

D	50	100	200	300	400	500
T	5	5	5	5	5	5
M	250	500	1000	1500	2000	2500

Table 8.1.4.1: Sample sizes.

The results of the simulation experiments are presented in the Table 8.1.4.2.

D	50	100	200	300	400	500
$BIAS(\hat{\beta}_1)$	-0.0046	0.0047	0.0014	-0.0012	-0.0002	-0.0009
$MSE(\hat{\beta}_1)$	0.1595	0.0851	0.0435	0.0278	0.0219	0.0169
$BIAS(\hat{\beta}_2)$	0.0012	-0.0005	-0.0011	0.0003	-0.0003	-0.0002
$MSE(\hat{\beta}_2)$	0.0159	0.0083	0.0042	0.0028	0.0020	0.0016
$BIAS(\hat{\sigma}_1^2)$	-0.0350	-0.0138	-0.0064	-0.0065	-0.0036	-0.0053
$MSE(\hat{\sigma}_1^2)$	0.1101	0.0547	0.0285	0.0192	0.0141	0.0116
$BIAS(\hat{\rho}_1)$	-0.0169	-0.0073	-0.0034	-0.0029	-0.0014	-0.0009
$MSE(\hat{\rho}_1)$	0.0267	0.0115	0.0059	0.0038	0.0030	0.0023
$BIAS(\hat{\sigma}_2^2)$	0.0006	-0.0012	0.0010	-0.0005	-0.0009	0.0001
$MSE(\hat{\sigma}_2^2)$	0.0388	0.0191	0.0100	0.0064	0.0049	0.0040
$BIAS$	-0.0003	0.0005	-0.0002	-0.0003	0.0001	0.0001
MSE	0.5784	0.5739	0.5715	0.5716	0.5712	0.5710

Table 8.1.4.2. Results of the simulation experiments.

The Table 8.1.4.2 shows that bias is always close to zero and that MSE decreases as the number of domains increases, so that the estimators are empirically consistent.

Simulation 1b

The steps of the simulation experiment are

1. Do $\beta_1 = 0$, $\beta_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho_1 = 0.5$, define W according to (8.3), generate σ_{dt}^2 y x_{dt} and read MSE_{dt}, $d = 1,\ldots,D$, $t = 1,\ldots,T$.

2. Repeat $K = 200$ timess ($k = 1,\ldots,K$)

    2.1. Generate $y_{dt}^{(k)}$ and calculate $\mu_{dt}^{(k)} = \beta_1 + \beta_2 x_{dt} + u_{1d}^{(k)} + u_{2dt}^{(k)}$, $d = 1,\ldots,D$, $t = 1,\ldots,T$.

    2.2. Calculate $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\sigma}_1^{2(k)}, \hat{\rho}_1^{(k)}, \hat{\sigma}_2^{2(k)}\}$ by using REML method.

    2.3. Repeat $B = 100$ times ($b = 1,\ldots,B$)

        2.3.1. Generate $y_{dt}^{(kb)}$ with the parameters $\{\hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\sigma}_1^{2(k)}, \hat{\rho}_1^{(k)}, \hat{\sigma}_2^{2(k)}\}$ from step 2.2. Generar $\mu_{dt}^{(kb)} = \hat{\beta}_1^{(k)} + \hat{\beta}_2^{(k)} x_{dt} + u_{1d}^{(kb)} + u_{2dt}^{(kb)}$.

        2.3.2. Calculate $\hat{\tau}^{(kb)} \in \{\hat{\beta}_1^{(kb)}, \hat{\beta}_2^{(kb)}, \hat{\sigma}_1^{2(kb)}, \hat{\rho}_1^{(kb)}, \hat{\sigma}_2^{2(kb)}\}$ and $\hat{\mu}_{dt}^{(kb)} = \hat{\beta}_1^{(kb)} + \hat{\beta}_2^{(kb)} x_{dt} + \hat{u}_{1d}^{(kb)} + \hat{u}_{2dt}^{(kb)}$, by using REML method.

    2.4. Calculate

$$mse_{dt}^{(k)} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\mu}_{dt}^{(kb)} - \mu_{dt}^{(kb)})^2.$$

3. For $d=1,\dots,D$, $t=1,\dots,T$, calculate

$$B_{dt}=\frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)}-MSE_{dt}),\quad E_{dt}=\frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)}-MSE_{dt})^2.$$

$$B=\frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T}B_{dt},\quad E=\frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T}E_{dt}.$$

The simulation experiment is repeated for all 6 combinations of sample sizes appearing in the Table 8.1.4.1. The simulation results are presented in the Table 8.1.4.3.

D	50	100	200	400
B	-0.0032	-0.0052	-0.0020	-0.0025
E	0.0082	0.0075	0.0071	0.0069

Table 8.1.4.3. Results of simulation 1b.

The Table 8.1.4.3 shows that bias B is always close to zero and that the MSE E decreases as the number of domains increases, so that the estimators *mse* are empirically consistent.

8.2 Model 2

8.2.1 Introduction

Let y_{dt} be a direct estimator of the characteristic of interest and let $\mathbf{x}_{dt}$ be a vector containing the aggregated values of p of auxiliary variables. The subindex d is used for domains and the subindex t for time instants. Let us consider the model

$$y_{dt}=\mathbf{x}_{dt}\beta+u_{1d}+u_{2dt}+e_{dt},\quad d=1,\dots,D,\quad t=1,\dots,T, \tag{8.4}$$

where $\{u_{1d}\}$, $\{u_{2dt}\}$ and $\{e_{dt}\}$ are independent with distributions $\{u_{1d}\}_{d=1}^{D}\sim SAR(1)$, $\{u_{2dt}\}_{t=1}^{T}$ i.i.d $AR(1)$ and $e_{dt}\sim N(0,\sigma_{dt}^2)$.

The model (8.4) can be alternatively written in the form

$$\mathbf{y}=\mathbf{X}\beta+\mathbf{Z}_1\mathbf{u}_1+\mathbf{Z}_2\mathbf{u}_2+\mathbf{e}, \tag{8.5}$$

where

- $\mathbf{y}=\underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(y_{dt}))$, $\quad\mathbf{e}=\underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(e_{dt}))$,
- $\mathbf{u}_1=\underset{1\le d\le D}{\text{col}}(u_{1d})$, $\mathbf{u}_2=\underset{1\le d\le D}{\text{col}}(\mathbf{u}_{2d})$, $\mathbf{u}_{2d}=\underset{1\le t\le T}{\text{col}}(u_{2dt})$,
- $\mathbf{X}=\underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(\mathbf{x}_{dt}))$, $\mathbf{x}_{dt}=\underset{1\le j\le p}{\text{col}'}(x_{dtj})$, $\beta=\underset{1\le j\le p}{\text{col}}(\beta_j)$,
- $\mathbf{Z}_1=\underset{1\le d\le D}{\text{diag}}(\mathbf{1}_T)$, $\mathbf{Z}_2=\mathbf{I}_{M\times M}$, $M=DT$.

We assume that $\mathbf{u}_1 \sim N(\mathbf{0}, \mathbf{V}_{u_1})$, $\mathbf{u}_2 \sim N(\mathbf{0}, \mathbf{V}_{u_2})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with covariance matrices

$$\begin{aligned}
\mathbf{V}_{u_1} &= \sigma_1^2 \Omega_1(\rho_1), \quad \Omega_1(\rho_1) = \left[(\mathbf{I}_D - \rho_1 \mathbf{W})'(\mathbf{I}_D - \rho_1 \mathbf{W})\right]^{-1} \triangleq \mathbf{C}^{-1}(\rho_1), \\
\mathbf{V}_{u_2} &= \sigma_2^2 \Omega_2(\rho_2), \quad \Omega_2(\rho_2) = \underset{1 \le d \le D}{\text{diag}} (\Omega_{2d}(\rho_2)), \\
\mathbf{V}_e &= \underset{1 \le d \le D}{\text{diag}} (\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1 \le t \le T}{\text{diag}} (\sigma_{dt}^2), \\
\Omega_{2d} &= \Omega_{2d}(\rho_2) = \frac{1}{1-\rho_2^2} \begin{pmatrix} 1 & \rho_2 & \dots & \rho_2^{T-2} & \rho_2^{T-1} \\ \rho_2 & 1 & \ddots & & \rho_2^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho_2^{T-2} & & \ddots & 1 & \rho_2 \\ \rho_2^{T-1} & \rho_2^{T-2} & \dots & \rho_2 & 1 \end{pmatrix}_{T \times T},
\end{aligned}$$

where the σ_{dt}^2's are known. We assume that the rows of matrix $\mathbf{W}$ are stochastic vectors, i.e. their components sum up to one. The vector $\mathbf{u}_1$ is distributed as a SAR(1) stochastic process and the vectors $\mathbf{u}_{2d}$ are independent with homogeneous AR(1) distributions (they all have the same variance and auto-correlation parameters).

The variance of $\mathbf{y}$ is

$$\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}_1 \mathbf{V}_{u_1} \mathbf{Z}_1' + \mathbf{Z}_2 \mathbf{V}_{u_2} \mathbf{Z}_2' + \mathbf{V}_e = \mathbf{Z}_1 \mathbf{V}_{u_1} \mathbf{Z}_1' + \underset{1 \le d \le D}{\text{diag}} (\sigma_2^2 \Omega_{2d}(\rho_2) + \mathbf{V}_{ed}).$$

Its inverse can be calculated by applying the formula

$$(A + CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}$$

with $A = \underset{1 \le d \le D}{\text{diag}} (\sigma_2^2 \Omega_{2d}(\rho_2) + \mathbf{V}_{ed})$, $B = \mathbf{V}_{u_1}$, $C = \mathbf{Z}_1$ and $D = \mathbf{Z}_1'$. Then

$$\mathbf{V}^{-1} = A^{-1} - A^{-1}\mathbf{Z}_1(\mathbf{V}_{u_1}^{-1} + \mathbf{Z}_1' A^{-1} \mathbf{Z}_1)^{-1} \mathbf{Z}_1' A^{-1},$$

where $A^{-1} = \underset{1 \le d \le D}{\text{diag}} (A_d^{-1})$ and $A_d = \sigma_2^2 \Omega_{2d}(\rho_2) + \mathbf{V}_{ed}$. Observe that by applying this formula we substitute the inversion of one matrix of order $M \times M$ by the inversion of D matrices of order $T \times T$.

Let us define the parameter $\theta = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)$. We apply the formula

$$\frac{\partial \mathbf{C}^{-1}}{\partial \rho_1} = -\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \rho_1} \mathbf{C}^{-1},$$

to obtain the partial derivatives of $\mathbf{V}$ with respect to θ.

$$\begin{aligned}
\mathbf{V}_1 &= \frac{\partial \mathbf{V}}{\partial \sigma_1^2} = \mathbf{Z}_1 \frac{\partial \mathbf{V}_{u_1}}{\partial \sigma_1^2}\mathbf{Z}_1' = \mathbf{Z}_1\Omega_1(\rho_1)\mathbf{Z}_1',\\
\mathbf{V}_2 &= \frac{\partial \mathbf{V}}{\partial \rho_1} = \mathbf{Z}_1 \frac{\partial \mathbf{V}_{u_1}}{\partial \rho_1}\mathbf{Z}_1' = -\sigma_1^2\mathbf{Z}_1\Omega_1(\rho_1)\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1}\Omega_1(\rho_1)\mathbf{Z}_1',\\
\mathbf{V}_3 &= \frac{\partial \mathbf{V}}{\partial \sigma_2^2} = \underset{1\le d\le D}{\text{diag}}\,(\Omega_{2d}(\rho_2)),\\
\mathbf{V}_4 &= \frac{\partial \mathbf{V}}{\partial \rho_2} = \sigma_2^2\,\underset{1\le d\le D}{\text{diag}}\left(\frac{\partial \Omega_{2d}(\rho_2)}{\partial \rho_2}\right),
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1} &= \frac{\partial \mathbf{C}}{\partial \rho_1} = \frac{\partial}{\partial \rho_1}\left\{(\mathbf{I}_D-\rho_1\mathbf{W})'(\mathbf{I}_D-\rho_1\mathbf{W})\right\}\\
&= -\mathbf{W}'+\rho_1\mathbf{W}'\mathbf{W}-\mathbf{W}+\rho_1\mathbf{W}'\mathbf{W} = -\mathbf{W}-\mathbf{W}'+2\rho_1\mathbf{W}'\mathbf{W},
\end{aligned}$$

and

$$\frac{\partial \Omega_{2d}(\rho_2)}{\partial \rho_2} = \frac{1}{1-\rho_2^2}\begin{pmatrix}
0 & 1 & \dots & \dots & (T-1)\rho_2^{T-2}\\
1 & 0 & \ddots & & (T-2)\rho_2^{T-3}\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
(T-2)\rho_2^{T-3} & & \ddots & 0 & 1\\
(T-1)\rho_2^{T-2} & \dots & \dots & 1 & 0
\end{pmatrix} + \frac{2\rho_2\Omega_{2d}(\rho_2)}{1-\rho_2^2}.$$

8.2.2 BLUP

The BLU estimator and predictor of β and $\mathbf{u}$ are

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{y} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\beta}),$$

where $\mathbf{V}_u = \text{diag}(\mathbf{V}_{u_1},\mathbf{V}_{u_2})$ and $\mathbf{Z}=(\mathbf{Z}_1,\mathbf{Z}_2)$. For calculating $\widehat{\mathbf{u}}$ we apply the formula

$$\widehat{\mathbf{u}} = \begin{pmatrix}\mathbf{V}_{u_1}\mathbf{Z}_1'\\ \mathbf{V}_{u_2}\mathbf{Z}_2'\end{pmatrix}\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\beta}) = \begin{pmatrix}\sigma_1^2\Omega_1(\rho_1)\mathbf{Z}_1'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\beta})\\ \sigma_2^2\Omega_2(\rho_2)\mathbf{Z}_2'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\beta})\end{pmatrix}.$$

The BLUP of μ_{dt} is

$$\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta}+\widehat{u}_{1d}+\widehat{u}_{2dt}.$$

8.2.3 Residual maximum likelihood estimation

The log-likelihood of the residual maximum likelihood estimation method is

$$l_{REML}(\theta) = -\frac{M-p}{2}\log 2\pi+\frac{1}{2}\log|\mathbf{X}'\mathbf{X}|-\frac{1}{2}\log|\mathbf{V}|-\frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)$ and

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \quad \mathbf{PVP} = \mathbf{P}, \quad \mathbf{PX} = \mathbf{0}.$$

Then

$$\mathbf{P}_a = \frac{\partial \mathbf{P}}{\partial \theta_a} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_a}\mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P}, \quad a = 1, \ldots, 4.$$

By taking derivatives on l_{REML} with respect to θ_a we get

$$S_a = \frac{\partial l_{REML}}{\partial \theta_a} = -\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y}, \quad a = 1, \ldots, 4.$$

By taking again derivatives with respect to θ_a and θ_b, taking expectations and changing the sign, we obtain

$$F_{ab} = \frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a, b = 1, \ldots, 4.$$

The updating formula of the Fisher-scoring algorithm is

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k).$$

We can take the reduced model without $\mathbf{u}_1$ and with $\rho_2 = 0$ as a reference for obtaining seeds for the Fisher-scoring algorithm. For the mentioned reduced model, it is easy to calculate the Henderson 3 estimator $\widehat{\sigma}_{u_2H}^2$ of the only remaining variance σ_2^2. Therefore, a possible set of algorithm seeds is $\sigma_1^{2(0)} = \sigma_2^{2(0)} = \frac{1}{2}\widehat{\sigma}_{u_2H}^2$, $\rho_1^{(0)} = \rho_2^{(0)} = 0.3$.

The REML estimator of β is

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

The asymptotic distributions of the REML estimators of θ and β are

$$\hat{\theta} \sim N_2(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1 - \alpha$ for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}\nu_{aa}^{1/2}, \ a = 1, \ldots, 4, \quad \hat{\beta}_j \pm z_{\alpha/2}q_{jj}^{1/2}, \ j = 1, \ldots, p,$$

where $\hat{\theta} = \theta^\kappa$, $\mathbf{F}^{-1}(\theta^\kappa) = (\nu_{ab})_{a,b=1,\ldots,4}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^\kappa)\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\ldots,p}$, κ is the last iteration in the Fisher-scoring algorithm and z_α is the α-quantil of the standard normal distribution $N(0,1)$. If we observe $\hat{\beta}_j = \beta_0$, the p-value for testing $H_0 : \beta_j = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_j > |\beta_0|) = 2P(N(0,1) > \beta_0/\sqrt{q_{jj}}).$$

8.2.4 Simulations

For $d = 1, \ldots, D$, $t = 1, \ldots, T$, the explanatory and target variables are

$$\begin{aligned} x_{dt} &= (b_{dt} - a_{dt})U_{dt} + a_{dt}, \ U_{dt} = \frac{t}{T+1}, \ a_{dt} = 1, \ b_{dt} = 1 + \frac{1}{D}(T(d-1) + t), \\ y_{dt} &= \beta_1 + \beta_2 x_{dt} + u_{1d} + u_{2dt} + e_{dt}, \ \beta_1 = 0, \ \beta_2 = 1, \end{aligned}$$

where $e_{dt} \sim N(0,\sigma_{dt}^2)$ and

$$\sigma_{dt}^2 = \frac{(\alpha_1-\alpha_0)\,(T(d-1)+t-1)}{M-1} + \alpha_0, \quad \alpha_0 = 0.8,\, \alpha_1 = 1.2.$$

The vector $\mathbf{u}_1 = \underset{1\leq d\leq D}{\mathrm{col}}(u_{1d})$ is generated from the distribution $N_D(0,\sigma_1^2\Omega_1(\rho_1))$, with $\sigma_1^2 = 1$, $\rho_1 = 0.5$ and proximity matrix (8.3). For $d = 1,\ldots,D$, the random effects u_{2dt} are generated as follows:

$$u_{2d1} = (1-\rho_2^2)^{-1/2}\varepsilon_{d1}, \qquad u_{2dt} = \rho_2 u_{2dt-1} + \varepsilon_{dt},\ t = 2,\ldots,T,$$

where $\varepsilon_{dt} \sim N(0,\sigma_2^2)$, $d = 1,\ldots,D$, $t = 1,\ldots,T$, and $\rho_2 = 0.5$.

Simulation 2a

The steps of the simulation experiment 2a are

1. Do $\beta_1 = 0$, $\beta_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho_1 = 0.5$, define W according to (8.3) and generate σ_{dt}^2, x_{dt}, $d = 1,\ldots,D$, $t = 1,\ldots,T$.

2. Repeat $K = 2000$ times ($k = 1,\ldots,K$)

   2.1. Generate $y_{dt}^{(k)}$ and calculate $\mu_{dt}^{(k)} = \beta_1 + \beta_2 x_{dt} + u_{1d}^{(k)} + u_{2dt}^{(k)}$, $d = 1,\ldots,D$, $t = 1,\ldots,T$.

   2.2. Calculate $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)},\hat{\beta}_2^{(k)},\hat{\sigma}_1^{2(k)},\widehat{\rho}_1^{(k)},\hat{\sigma}_2^{2(k)},\widehat{\rho}_2^{(k)}\}$ and $\hat{\mu}_{dt}^{(k)} = \hat{\beta}_1^{(k)} + \hat{\beta}_2^{(k)} x_{dt} + \hat{u}_{1d}^{(k)} + \hat{u}_{2dt}^{(k)}$, by using the REML method.

3. For each $\tau \in \{\beta_1,\beta_2,\sigma_1^2,\rho_1,\sigma_2^2,\rho_2\}$ and for $\hat{\mu}_{dt}$, $d = 1,\ldots,D$, $t = 1,\ldots,T$, calculate

$$BIAS(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)}-\tau), \quad MSE(\hat{\tau}) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\tau}^{(k)}-\tau)^2.$$

$$BIAS_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)}-\mu_{dt}^{(k)}), \quad MSE_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{\mu}_{dt}^{(k)}-\mu_{dt}^{(k)})^2,$$

$$BIAS = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T} BIAS_{dt}, \quad MSE = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T} MSE_{dt}.$$

The simulation experiment is repeated for the 6 combinations of sample sizes appearing in the Table 8.2.4.1.

D	50	100	200	300	400	500
T	5	5	5	5	5	5
M	250	500	1000	1500	2000	2500

Table 8.2.4.1: Sample sizes.

The Tabla 8.2.4.2 presents the simulation results.

D	50	100	200	300	400	500
$BIAS(\hat{\beta}_1)$	0.0021	0.0059	0.0015	0.0023	0.0027	-0.0001
$MSE(\hat{\beta}_1)$	0.1966	0.1050	0.0515	0.0344	0.0257	0.0199
$BIAS(\hat{\beta}_2)$	0.0004	-0.0016	-0.0019	-0.0012	-0.0025	-0.0011
$MSE(\hat{\beta}_2)$	0.0212	0.0109	0.0053	0.0036	0.0025	0.0020
$BIAS(\hat{\sigma}_1^2)$	0.0413	-0.0248	-0.0308	-0.0337	-0.0269	-0.0226
$MSE(\hat{\sigma}_1^2)$	0.2519	0.1603	0.1025	0.0706	0.0550	0.0449
$BIAS(\hat{\rho}_1)$	-0.0245	0.0061	0.0083	0.0108	0.0075	0.0053
$MSE(\hat{\rho}_1)$	0.0455	0.0241	0.0139	0.0086	0.0069	0.0054
$BIAS(\hat{\sigma}_2^2)$	-0.0174	-0.0082	-0.0039	-0.0050	-0.0013	-0.0017
$MSE(\hat{\sigma}_2^2)$	0.0430	0.0228	0.0109	0.0072	0.0056	0.0044
$BIAS(\hat{\rho}_2)$	-0.0896	-0.0293	-0.0113	-0.0030	-0.0023	-0.0011
$MSE(\hat{\rho}_2)$	0.0412	0.0191	0.0112	0.0074	0.0056	0.0044
$BIAS$	-0.0016	0.0000	-0.0006	-0.0008	-0.0002	-0.0001
MSE	0.5500	0.5446	0.5424	0.5418	0.5413	0.5409

Table 8.2.4.2. Resultados del experimento de simulación 2a.

The Table 8.2.4.2 shows that the bias is always close to zero and that the MSE decreases as the number of domains increases, so that the REML estimators are empirically consistent.

Simulation 2b

The steps of the simulation experiment 2b are

1. Do $\beta_1 = 0$, $\beta_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho_1 = 0.5$, $\rho_2 = 0.5$, define W according to (8.3), generate σ_{dt}^2 and x_{dt} and read MSE_{dt}, $d = 1,\dots,D$, $t = 1,\dots,T$.
2. Repeat $K = 200$ times ($k = 1,\dots,K$)
   - 2.1. Generate $y_{dt}^{(k)}$ and calculate $\mu_{dt}^{(k)} = \beta_1 + \beta_2 x_{dt} + u_{1d}^{(k)} + u_{2dt}^{(k)}$, $d = 1,\dots,D$, $t = 1,\dots,T$.
   - 2.2. Calculate $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\sigma}_1^{2(k)}, \widehat{\rho}_1^{(k)}, \hat{\sigma}_2^{2(k)}, \widehat{\rho}_2^{(k)}\}$ by using the REML method.
   - 2.3. Repeat $B = 100$ times ($b = 1,\dots,B$)
     - 2.3.1. Generate $y_{dt}^{(kb)}$ with the parameters $\{\hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\sigma}_1^{2(k)}, \widehat{\rho}_1^{(k)}, \hat{\sigma}_2^{2(k)}, \widehat{\rho}_2^{(k)}\}$ obtained in step 2.2. Generate $\mu_{dt}^{(kb)} = \hat{\beta}_1^{(k)} + \hat{\beta}_2^{(k)} x_{dt} + u_{1d}^{(kb)} + u_{2dt}^{(kb)}$.
     - 2.3.2. Calculate $\hat{\tau}^{(kb)} \in \{\hat{\beta}_1^{(kb)}, \hat{\beta}_2^{(kb)}, \hat{\sigma}_1^{2(kb)}, \widehat{\rho}_1^{(kb)}, \hat{\sigma}_2^{2(kb)}, \widehat{\rho}_2^{(kb)}\}$ and $\hat{\mu}_{dt}^{(kb)} = \hat{\beta}_1^{(kb)} + \hat{\beta}_2^{(kb)} x_{dt} + \hat{u}_{1d}^{(kb)} + \hat{u}_{2dt}^{(kb)}$, by using the REML method.
   - 2.4. Calculate
$$mse_{dt}^{(k)} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\mu}_{dt}^{(kb)} - \mu_{dt}^{(kb)})^2.$$

3. For $d = 1,\dots,D$, $t = 1,\dots,T$, calculate

$$B_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt}), \quad E_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt})^2.$$

$$B = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T}B_{dt}, \quad E = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{T}E_{dt}.$$

The simulation experiment is repeated for the 6 combinations of sample sizes appearing in the Table 8.1.4.1. The Tabla 8.1.4.3 presents the simulation results.

D	50	100	200	400
B	0.0009	-0.0032	0.0010	
E	0.0086	0.0072	0.0067	

Table 8.2.4.3. Results of simulation 2b.

The Table 8.2.4.3 shows that the bias B is always close to zero and that the MSE E decreases as the number of domains increases, so that the estimators *mse* are empirically consistent.

Chapter 9

Unit-level time models

9.1 Unit-level model with correlated time effects

9.1.1 Introduction

Let us consider a version of mixed model (1.1) with two nested random factors, where the first factor has D levels and, for each level d ($d = 1,\dots,D$) of this factor, the second factor has m_d levels. More concretely, let us consider the model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{W}^{-1/2}\mathbf{e}, \tag{9.1}$$

where $\mathbf{u}_1 = \mathbf{u}_{1,D\times 1} \sim N(0,\sigma_1^2\mathbf{I}_D)$, $\mathbf{u}_2 = \mathbf{u}_{2,M\times 1} \sim N(0,\sigma_2^2\Omega(\rho))$ and $\mathbf{e} = \mathbf{e}_{n\times 1} \sim N(0,\sigma_0^2\mathbf{I}_n)$ are independent, $\mathbf{y} = \mathbf{y}_{n\times 1}$, $\mathbf{X} = \mathbf{X}_{n\times p}$ with $\mathrm{r}(\mathbf{X}) = p$, $\beta = \beta_{p\times 1}$, $\mathbf{Z}_1 = \underset{1\le d\le D}{\mathrm{diag}}(\mathbf{1}_{n_d})_{n\times D}$, $\mathbf{Z}_2 = \underset{1\le d\le D}{\mathrm{diag}}(\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}))_{n\times M}$, $M = \sum_{d=1}^D m_d$, $n = \sum_{d=1}^D n_d$, $n_d = \sum_{t=1}^{m_d} n_{dt}$, $\mathbf{I}_a$ is the $a\times a$ identity matrix, $\mathbf{1}_a$ is the $a\times 1$ vector with all its elements equal to 1, $\mathbf{W} = \underset{1\le d\le D}{\mathrm{diag}}(\mathbf{W}_d)$, $\mathbf{W}_d = \underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{W}_{dt})$, $\mathbf{W}_{dt} = \underset{1\le j\le n_{dt}}{\mathrm{diag}}(w_{dtj})_{n\times n}$ with known $w_{dtj} > 0$, $d = 1,\dots,D$, $t = 1,\dots,m_d$, $j = 1,\dots,n_{dt}$, $\Omega(\rho) = \underset{1\le d\le D}{\mathrm{diag}}(\Omega_d)$ and

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \dots & \rho & 1 \end{pmatrix}_{m_d\times m_d}.$$

Model (9.1) can alternatively be written in the form

$$y_{dtj} = \mathbf{x}_{dtj}\beta + u_{1,d} + u_{2,dt} + w_{dtj}^{-1/2}e_{dtj}, \quad d = 1,\dots,D, t = 1,\dots,m_d, j = 1,\dots,n_{dt}, \tag{9.2}$$

where y_{dtj} is the target variable for the sample unit j, time t and domain d, and $\mathbf{x}_{dtj}$ is the row (d,t,j) of matrix $\mathbf{X}$. The random vectors $(u_{2d1},\dots,u_{2dm_d})$, $d = 1,\dots,D$, are i.i.d. AR(1).

In what follows we use the alternative parameters

$$\sigma^2=\sigma_0^2,\quad \varphi_1=\frac{\sigma_1^2}{\sigma_0^2},\quad \varphi_2=\frac{\sigma_2^2}{\sigma_0^2},\quad \rho=\rho.$$

Let $\sigma=(\sigma^2,\varphi_1,\varphi_2,\rho)$ be the vector of variance components, with $\sigma^2>0$, $\varphi_1>0$, $\varphi_2>0$ and $-1<\rho<1$. If σ is known, the BLUE of $\beta=(\beta_1,\ldots,\beta_p)'$ and the BLUP of $\mathbf{u}=(\mathbf{u}_1',\mathbf{u}_2')'$ are

$$\widehat{\beta}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}}=\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\widehat{\beta}\right). \tag{9.3}$$

Formulas (9.3) are not computationally efficient because they require the inversion of the $n\times n$ matrix $\mathbf{V}$. By calculating the inversion of $\mathbf{V}$ new formulas are obtained.

Under model (9.1), we have $\text{var}(\mathbf{u}_1)=\sigma^2\varphi_1\mathbf{I}_D$, $\text{var}(\mathbf{u}_2)=\sigma^2\varphi_2\Omega(\rho)$, $\text{var}(\mathbf{e})=\sigma^2\mathbf{I}_n$ and

$$\mathbf{V}=\text{var}(\mathbf{y})=\mathbf{Z}_1\text{var}(\mathbf{u}_1){\mathbf{Z}_1}'+\mathbf{Z}_2\text{var}(\mathbf{u}_2){\mathbf{Z}_2}'+\sigma^2\mathbf{W}^{-1}=\sigma^2\Sigma=\sigma^2\text{diag}(\Sigma_1,\ldots,\Sigma_D),$$

where

$$\Sigma_d=\varphi_1\mathbf{1}_{n_d}\mathbf{1}_{n_d}'+\varphi_2\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')+\mathbf{W}_d^{-1}=\varphi_1\mathbf{1}_{n_d}\mathbf{1}_{n_d}'+\mathbf{L}_d,\ d=1,\ldots,D.$$

To calculate $\mathbf{L}_d^{-1}$ we use the formula

$$(A+CBD)^{-1}=A^{-1}-A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1}$$

with $A=\mathbf{W}_d^{-1}$, $C=\varphi_2\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})$, $B=\Omega_d$ and $D=\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')$. we obtain

$$\begin{aligned}\mathbf{L}_d^{-1} &= \mathbf{W}_d-\varphi_2\mathbf{W}_d\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\left[\Omega_d^{-1}(\rho)+\varphi_2\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\mathbf{W}_d\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\right]^{-1}\\ &\cdot\ \underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\mathbf{W}_d.\end{aligned}$$

To calculate Σ_d^{-1} we use the formula

$$(\mathbf{A}+\mathbf{u}\mathbf{v}')^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}'\mathbf{A}^{-1}}{1+\mathbf{v}'\mathbf{A}^{-1}\mathbf{u}}$$

with $A=\mathbf{L}_d$, $\mathbf{u}=\varphi_1\mathbf{1}_{n_d}$, $\mathbf{v}'=\mathbf{1}_{n_d}'$. We obtain

$$\Sigma_d^{-1}=\mathbf{L}_d^{-1}-\frac{\varphi_1}{1+\varphi_1\mathbf{1}_{n_d}'\mathbf{L}_d^{-1}\mathbf{1}_{n_d}}\mathbf{L}_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\mathbf{L}_d^{-1}.$$

The final formula for $\widehat{\beta}$ is

$$\widehat{\beta}=\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{X}_d\right)^{-1}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right) \tag{9.4}$$

where $\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$ and $\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$. The final formula for $\widehat{\mathbf{u}}$ is

$$\begin{aligned}
\widehat{\mathbf{u}} &= \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\widehat{\beta}\right) = \begin{pmatrix} \sigma_1^2\mathbf{I}_D & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\Omega(\rho) \end{pmatrix} \begin{bmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_2' \end{bmatrix} \underset{1\le d\le D}{\text{diag}}\left(\mathbf{V}_d^{-1}\right) \underset{1\le d\le D}{\text{col}}\left[\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right] \\
&= \begin{bmatrix} \varphi_1 \underset{1\le d\le D}{\text{diag}}\left(\mathbf{1}_{n_d}'\right) \underset{1\le d\le D}{\text{diag}}\left(\Sigma_d^{-1}\right) \underset{1\le d\le D}{\text{col}}\left[\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right] \\ \varphi_2 \underset{1\le d\le D}{\text{diag}}\left(\Omega_d(\rho)\right) \underset{1\le d\le D}{\text{diag}}\left(\underset{1\le t\le m_d}{\text{diag}}\left(\mathbf{1}_{n_{dt}}'\right)\right) \underset{1\le d\le D}{\text{diag}}\left(\Sigma_d^{-1}\right) \underset{1\le d\le D}{\text{col}}\left[\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right] \end{bmatrix} \\
&= \begin{bmatrix} \varphi_1 \underset{1\le d\le D}{\text{col}}\left[\mathbf{1}_{n_d}'\Sigma_d^{-1}\left(\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right)\right] \\ \varphi_2 \underset{1\le d\le D}{\text{col}}\left[\Omega_d(\rho) \underset{1\le t\le m_d}{\text{diag}}\left(\mathbf{1}_{n_{dt}}'\right)\Sigma_d^{-1}\left(\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right)\right] \end{bmatrix}.
\end{aligned}$$

9.1.2 REML estimators of model parameters

The restricted log-likelihood is

$$l_{reml}(\sigma) = -\frac{1}{2}(n-p)\log 2\pi - \frac{1}{2}(n-p)\log\sigma^2 - \frac{1}{2}\log|\mathbf{K}'\Sigma\mathbf{K}| - \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where

$$\mathbf{P} = \mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}' = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1},\ \mathbf{K} = \mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}.$$

Let us denote the derivatives of $\Omega(\rho)$ by $\Omega'(\rho) = \frac{\partial\Omega(\rho)}{\partial\rho}$ and $\Omega''(\rho) = \frac{\partial^2\Omega(\rho)}{\partial\rho^2}$. By taking partial derivatives with respect to σ^2, φ_1^2, φ_2^2 and ρ we obtain the components of the score vector $S(\sigma)$.

$$\begin{aligned}
S_{\sigma^2} &= -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{y}, \\
S_{\varphi_1} &= -\frac{1}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\} + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y}, \\
S_{\varphi_2} &= -\frac{1}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\} + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}, \\
S_{\rho} &= -\frac{\varphi_2}{2}\text{tr}\left\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\right\} + \frac{\varphi_2}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y},
\end{aligned}$$

The second partial derivatives of the restricted log-likelihood function are

$$\begin{aligned}
H_{\sigma^2\sigma^2} &= \frac{n-p}{2\sigma^4} - \frac{1}{\sigma^6}\mathbf{y}'\mathbf{P}\mathbf{y}, \quad H_{\sigma^2\varphi_1} = -\frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y}, \\
H_{\sigma^2\varphi_2} &= -\frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}, \quad H_{\sigma^2\rho} = -\frac{\varphi_2}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}, \\
H_{\varphi_1\varphi_1} &= \frac{1}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\} - \frac{1}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y}, \\
H_{\varphi_1\varphi_2} &= \frac{1}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\} - \frac{1}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}, \\
H_{\varphi_1\rho} &= \frac{\varphi_2}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\} - \frac{\varphi_2}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}, \\
H_{\varphi_2\varphi_2} &= \frac{1}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\} - \frac{1}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y},
\end{aligned}$$

$$\begin{aligned}
H_{\varphi_2\rho} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}+\frac{\varphi_2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}\\
&+ \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}-\frac{\varphi_2}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y},\\
H_{\rho\rho} &= \frac{\varphi_2^2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}-\frac{\varphi_2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega''(\rho)\mathbf{Z}_2'\}\\
&- \frac{\varphi_2^2}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}+\frac{\varphi_2}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega''(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y}.
\end{aligned}$$

By taking expectations, changing the sign and taking into account that $\mathbf{PX}=\mathbf{0}$ and $\mathbf{P\Sigma P}=\mathbf{P}$, we get the elements of the Fisher information matrix,

$$\begin{aligned}
F_{\sigma^2\sigma^2} &= -\frac{n-p}{2\sigma^4}+\frac{1}{\sigma^4}\operatorname{tr}\{\mathbf{P}\Sigma\}=\frac{n-p}{2\sigma^4},\quad F_{\sigma^2\varphi_1}=\frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\},\\
F_{\sigma^2\varphi_2} &= \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\},\quad F_{\sigma^2\rho}=\frac{\varphi_2}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\},\\
F_{\varphi_1\varphi_1} &= \frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\},\quad F_{\varphi_1\varphi_2}=\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\}\\
F_{\varphi_1\rho} &= \frac{\varphi_2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\},\quad F_{\varphi_2\varphi_2}=\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\},\\
F_{\varphi_2\rho} &= \frac{\varphi_2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\},\quad F_{\rho\rho}=\frac{\varphi_2^2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}.
\end{aligned}$$

The updating formula of the Fisher-scoring algorithm is

$$\sigma^{k+1}=\sigma^k+\mathbf{F}^{-1}(\sigma^k)\mathbf{S}(\sigma^k).$$

As algorithm seeds we may use $\rho^{(0)}=0$ and the Henderson 3 estimators $\sigma_0^{2(0)}$, $\sigma_1^{2(0)}$, $\sigma_2^{2(0)}$, under the model with $\rho=0$. The REML estimator $\widehat{\beta}_{reml}$ is calculated by using the formula (9.4).

Observation 9.1.1. From equation $S_{\sigma^2}=\mathbf{0}$, we get

$$\widehat{\sigma}^2=\frac{1}{n-p}\mathbf{y}'\mathbf{P}\mathbf{y}, \tag{9.5}$$

which can be used to introduce an algorithm updating σ^2 with (9.5) and $\varphi=(\varphi_1,\varphi_2,\rho)'$ with

$$\varphi^{k+1}=\varphi^k+F^{-1}(\varphi^k)S(\varphi^k).$$

Observation 9.1.2. It holds that

$$\dot{\Omega}_d(\rho)=\frac{1}{1-\rho^2}\begin{pmatrix}
0 & 1 & \dots & \dots & (m_d-1)\rho^{m_d-2}\\
1 & 0 & \ddots & & (m_d-2)\rho^{m_d-3}\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
(m_d-2)\rho^{m_d-3} & & \ddots & 0 & 1\\
(m_d-1)\rho^{m_d-2} & \dots & \dots & 1 & 0
\end{pmatrix}+\frac{2\rho\Omega_d(\rho)}{1-\rho^2},$$

$$\Omega_d^{-1}(\rho)=\begin{pmatrix} 1 & -\rho & 0 & \dots & \dots & 0\\ -\rho & 1+\rho^2 & -\rho & 0 & & 0\\ 0 & -\rho & 1+\rho^2 & -\rho & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & \ddots & 0\\ 0 & & \ddots & -\rho & 1+\rho^2 & -\rho\\ 0 & \dots & \dots & 0 & -\rho & 1 \end{pmatrix}=\mathbf{I}_{m_d}+\rho^2\mathbf{E}-\rho\mathbf{F},$$

where $\mathbf{E}$ is a diagonal matrix with diagonal elements $0,1,\dots,1,0$, and $\mathbf{F}$ is a matrix whose elements in the diagonals immediately above and below the principal diagonal are equal to -1 and whose remaining elements are equal to 0.

Matrix calculations

In what follows we presents computationally efficient formulas for the scores and the Fisher information components. These formulas avoid the construction of $n\times n$ matrices. Let us define

$$\Sigma=\underset{1\le d\le D}{\text{diag}}(\Sigma_d),\ \mathbf{X}=\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d),\ \mathbf{y}=\underset{1\le d\le D}{\text{col}}(\mathbf{y}_d),\ \mathbf{R}=(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}=\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{X}_d\right)^{-1}$$

so that

$$\mathbf{P}=\Sigma^{-1}-\Sigma^{-1}\mathbf{X}\mathbf{R}\mathbf{X}'\Sigma^{-1}=\underset{1\le d\le D}{\text{diag}}(\Sigma_d^{-1})-\underset{1\le d\le D}{\text{col}}(\Sigma_d^{-1}\mathbf{X}_d)\mathbf{R}\underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\Sigma_d^{-1})$$

The scores are

$$\begin{aligned} S_{\sigma^2} &= -\frac{n-p}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{y}_d-\frac{1}{2\sigma^4}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right),\\ S_{\varphi_1} &= -\frac{1}{2}\text{tr}\{\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\}+\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y}=-\frac{1}{2}\sum_{d=1}^{D}\mathbf{1}_{n_d}'[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}]\mathbf{1}_{n_d}\\ &+ \frac{1}{2\sigma^2}\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{y}_d-\frac{1}{\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right)\\ &+ \frac{1}{2\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right), \end{aligned}$$

$$
\begin{aligned}
S_{\varphi_2} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\} + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y} \\
&= -\frac{1}{2}\sum_{d=1}^{D}\operatorname{tr}\left\{\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}]\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\right\} \\
&+ \frac{1}{2\sigma^2}\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{y}_d \\
&- \frac{1}{\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right) \\
&+ \frac{1}{2\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right) \\
&\cdot\ \mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right), \\
S_{\rho} &= -\frac{\varphi_2}{2}\operatorname{tr}\left\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\right\} + \frac{\varphi_2}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y} \\
&= -\frac{\varphi_2}{2}\sum_{d=1}^{D}\operatorname{tr}\left\{\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}]\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\right\} \\
&+ \frac{\varphi_2}{2\sigma^2}\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{y}_d \\
&- \frac{\varphi_2}{\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right) \\
&+ \frac{\varphi_2}{2\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right) \\
&\cdot\ \mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right).
\end{aligned}
$$

The elements of the REML Fisher information matrix are

$$
\begin{aligned}
F_{\sigma^2\sigma^2} &= \frac{n-p}{2\sigma^2} \\
F_{\sigma^2\varphi_1} &= \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\} = \frac{1}{2\sigma^2}\sum_{d=1}^{D}\mathbf{1}_{n_d}'\left[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\right]\mathbf{1}_{n_d} \\
F_{\sigma^2\varphi_2} &= \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\} \\
&= \frac{1}{2\sigma^2}\sum_{d=1}^{D}\operatorname{tr}\{\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\left[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\right]\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\}
\end{aligned}
$$

$$
\begin{aligned}
F_{\sigma^2\rho} &= \frac{\varphi_2}{2\sigma^2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\left[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\right]\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega'_d(\rho)\}\\
F_{\varphi_1\varphi_1} &= \frac{1}{2}\mathrm{tr}\{\mathbf{Z}'_1\mathbf{P}\mathbf{Z}_1\mathbf{Z}'_1\mathbf{P}\mathbf{Z}_1\} = \frac{1}{2}\sum_{d=1}^{D}(\mathbf{1}'_{n_d}\Sigma_d^{-1}\mathbf{1}_{n_d})^2-\sum_{d=1}^{D}\mathbf{1}'_{n_d}\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{1}_{n_d}\\
&+ \frac{1}{2}\sum_{d=1}^{D}\mathbf{1}'_{n_d}\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{1}_{n_d},\\
F_{\varphi_1\varphi_2} &= \frac{1}{2}\mathrm{tr}\{\mathbf{Z}'_2\mathbf{P}\mathbf{Z}_1\mathbf{Z}'_1\mathbf{P}\mathbf{Z}_2\Omega(\rho)\} = \frac{1}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\}\\
&- \sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\}\\
&+ \frac{1}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\mathbf{X}_d\right)\\
&\cdot\ \mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\},\\
F_{\varphi_1\rho} &= \frac{\varphi_2}{2}\mathrm{tr}\{\mathbf{Z}'_2\mathbf{P}\mathbf{Z}_1\mathbf{Z}'_1\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\}\\
&= \frac{\varphi_2}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega'_d(\rho)\}\\
&- \varphi_2\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega'_d(\rho)\}\\
&+ \frac{\varphi_2}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\Sigma_d^{-1}\mathbf{X}_d\right)\\
&\cdot\ \mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega'_d(\rho)\},\\
F_{\varphi_2\varphi_2} &= \frac{1}{2}\mathrm{tr}\{\mathbf{Z}'_2\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}'_2\mathbf{P}\mathbf{Z}_2\Omega(\rho)\}\\
&= \frac{1}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\}\\
&- \sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\\
&\cdot\ \Omega_d(\rho)\}\\
&+ \frac{1}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}'_d\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dt}})\Sigma_d^{-1}\mathbf{X}_d\right)\\
&\cdot\ \mathbf{R}\mathbf{X}'_d\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\},
\end{aligned}
$$

$$
\begin{aligned}
F_{\varphi_2\rho} &= \frac{\varphi_2}{2}\,\mathrm{tr}\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\}\\
&= \frac{\varphi_2}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\}\\
&- \varphi_2\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\\
&\cdot\ \Omega_d'(\rho)\}\\
&+ \frac{\varphi_2}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right)\\
&\cdot\ \mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\},\\
F_{\rho\rho} &= \frac{\varphi_2^2}{2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}\\
&= \frac{\varphi_2^2}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\}\\
&- \varphi_2^2\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\\
&\cdot\ \Omega_d'(\rho)\}\\
&+ \frac{\varphi_2^2}{2}\sum_{d=1}^{D}\mathrm{tr}\{\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right)\\
&\cdot\ \mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\Omega_d'(\rho)\}.
\end{aligned}
$$

9.1.3 The EBLUP of the domain mean

The EBLUP of the linear parameter $\eta = \mathbf{a}'\mathbf{y} = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\mathbf{y}_r$ is

$$\widehat{\eta} = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta} + \widehat{\mathbf{V}}_{rs}\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta})\right]$$

As $\mathbf{V}_{ers} = \mathbf{0}$, $\mathbf{V}_{rs} = \mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s' + \mathbf{V}_{ers} = \mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'$ and $\widehat{\mathbf{u}} = \Sigma_u\mathbf{Z}_s'\mathbf{V}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta})$, we get

$$
\begin{aligned}
\widehat{\eta} &= \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta} + \mathbf{Z}_r\Sigma_u\mathbf{Z}_s'\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta})\right] = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta} + \mathbf{Z}_r\widehat{\mathbf{u}}\right]\\
&= \mathbf{a}'\left[\mathbf{X}\widehat{\beta} + \mathbf{Z}_1\widehat{\mathbf{u}}_1 + \mathbf{Z}_2\widehat{\mathbf{u}}_2\right] + \mathbf{a}_s'\left[\mathbf{y}_s - \mathbf{X}_s\widehat{\beta} - \mathbf{Z}_{s1}\widehat{\mathbf{u}}_1 - \mathbf{Z}_{s2}\widehat{\mathbf{u}}_2\right].
\end{aligned}
$$

Under model (9.2), $\overline{Y}_{dt} = \frac{1}{N_{dt}}\sum_{j=1}^{N_{dt}} y_{dtj}$ can be written as a linear parameter $\eta = \mathbf{a}'\mathbf{y}$, where

$$
\begin{aligned}
\mathbf{a}' &= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}', \ldots, \mathbf{0}_{N_{d-1}}', \mathbf{0}_{N_{d1}}', \ldots, \mathbf{0}_{N_{d(i-1)}}', \mathbf{1}_{N_{dt}}', \mathbf{0}_{N_{d(i+1)}}', \ldots, \mathbf{0}_{N_{dm_d}}', \mathbf{0}_{N_{d+1}}', \ldots, \mathbf{0}_{N_D}')\\
&= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}', \ldots, \mathbf{0}_{N_{d-1}}', \underset{1\le k\le m_d}{\mathrm{col}'}\left[\delta_{tk}\mathbf{1}_{N_{dk}}'\right], \mathbf{0}_{N_{d+1}}', \ldots, \mathbf{0}_{N_D}') = \frac{1}{N_{dt}}\underset{1\le \ell\le D}{\mathrm{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\mathrm{col}'}[\delta_{tk}\mathbf{1}_{N_{\ell k}}']\}
\end{aligned}
$$

with $\delta_{ab} = 1$ si $a = b$ and $\delta_{ab} = 0$ si $a \neq b$. It holds that $\mathbf{a}'\mathbf{X} = \overline{X}_{dt}$,

$$\begin{aligned}\mathbf{a}'\mathbf{Z}_1 &= \frac{1}{N_{dt}} \underset{1\le\ell\le D}{\text{col}'}\{\delta_{d\ell} \underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}}]\} \underset{1\le\ell\le D}{\text{diag}}(\mathbf{1}_{N_\ell}) = \underset{1\le\ell\le D}{\text{col}'}\{\delta_{d\ell}\} = \overline{\mathbf{Z}}_{1,dt},\\ \mathbf{a}'\mathbf{Z}_2 &= \frac{1}{N_{dt}} \underset{1\le\ell\le D}{\text{col}'}\{\delta_{d\ell} \underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}}]\} \underset{1\le\ell\le D}{\text{diag}}(\underset{1\le k\le m_\ell}{\text{diag}}(\mathbf{1}_{N_{\ell k}})) = \underset{1\le\ell\le D}{\text{col}'}\{\underset{1\le k\le m_\ell}{\text{col}'}\{\delta_{d\ell}\delta_{tk}\}\} = \overline{\mathbf{Z}}_{2,dt}.\end{aligned}$$

If $n_{dt} > 0$, the EBLUP of $\overline{Y}_{dt}$ is

$$\widehat{\overline{Y}}_{dt}^{eblup} = \overline{\mathbf{X}}_{dt}\widehat{\beta} + \overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1 + \overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2 + f_{dt}\left[\overline{\mathbf{y}}_{s,dt} - \overline{\mathbf{X}}_{s,dt}\widehat{\beta} - \overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1 - \overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2\right],$$

where $\overline{\mathbf{y}}_{s,dt} = \frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}} y_{dtj}$, $\overline{\mathbf{X}}_{s,dt} = \frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}} \mathbf{x}_{dtj}$ and $f_{dt} = \frac{n_{dt}}{N_{dt}}$. If $n_{dt} = 0$, the EBLUP of $\overline{Y}_{dt}$ is the synthetic part

$$\widehat{\overline{Y}}_{dt}^{eblup} = \overline{\mathbf{X}}_{dt}\widehat{\beta} + \overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1 + \overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2.$$

9.1.4 Mean squared error of the EBLUP

Let $\theta = (\sigma_0^2, \varphi_1, \varphi_2, \rho)$ be the vector of variance components. A second order approximation to the mean squared error of the EBLUP is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta) + g_4(\theta),$$

where

$$\begin{aligned}g_1(\theta) &= \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r,\\ g_2(\theta) &= [\mathbf{a}_r'\mathbf{X}_r - \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r - \mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r],\\ g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}_s(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},\\ g_4(\theta) &= \mathbf{a}_r'\mathbf{V}_{er}\mathbf{a}_r.\end{aligned}$$

Calculation of $g_1(\theta)$

The elements of formula $g_1(\theta) = \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r$ are

$$\begin{aligned}\mathbf{a}_r' &= \frac{1}{N_{dt}}\left(\mathbf{0}'_{N_1-n_1},\ldots,\mathbf{0}'_{N_{d-1}-n_{d-1}}, \underset{1\le k\le m_d}{\text{col}'}\left[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}\right], \mathbf{0}'_{N_{d+1}-n_{d+1}},\ldots,\mathbf{0}'_{N_D-n_D}\right),\\ \mathbf{Z}_r &= [\mathbf{Z}_{1r}\,\mathbf{Z}_{2r}], \quad \mathbf{T}_s = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{V}_u = \begin{pmatrix}\mathbf{T}_{11s} & \mathbf{T}_{12s}\\ \mathbf{T}_{21s} & \mathbf{T}_{22s}\end{pmatrix},\\ \mathbf{V}_u &= \begin{pmatrix}\sigma_1^2\mathbf{I}_D & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\Omega(\rho)\end{pmatrix}, \quad \mathbf{Z}_s = [\mathbf{Z}_{1s}\,\mathbf{Z}_{2s}], \quad \mathbf{V}_s^{-1} = \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}.\end{aligned}$$

It holds that

$$\begin{aligned}
\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{V}_u &= \begin{pmatrix} \sigma_1^2\mathbf{Z}_{1s}' \\ \sigma_2^2\Omega(\rho)\mathbf{Z}_{2s}' \end{pmatrix} \underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\left[\sigma_1^2\mathbf{Z}_{1s}, \sigma_2^2\mathbf{Z}_{2s}\Omega(\rho)\right] \\
&= \begin{pmatrix} \sigma_1^4\mathbf{Z}_{1s}'\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{1s} & \sigma_1^2\sigma_2^2\mathbf{Z}_{1s}'\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s}\Omega(\rho) \\ \sigma_1^2\sigma_2^2\Omega(\rho)\mathbf{Z}_{2s}'\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{1s} & \sigma_2^4\Omega(\rho)\mathbf{Z}_{2s}'\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s}\Omega(\rho) \end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{Z}_{1s}'\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{1s} &= \underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{1}_{n_d}'\}\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{1}_{n_d}\} = \underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{1}_{nd}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{nd}\}, \\
\mathbf{Z}_{1s}'\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s}\Omega(\rho) &= \underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{1}_{n_d}'\}\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\underset{1\le d\le D}{\operatorname{diag}}\{\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\}\Omega(\rho) \\
&= \underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{1}_{nd}'\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\}, \\
\Omega(\rho)\mathbf{Z}_{2s}'\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s}\Omega(\rho) &= \underset{1\le d\le D}{\operatorname{diag}}\{\Omega_d(\rho)\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\}.
\end{aligned}$$

The blocks of matrix $\mathbf{T}_s$ are

$$\begin{aligned}
\mathbf{T}_{11s} &= \sigma_1^2\underset{1\le d\le D}{\operatorname{diag}}\{1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}\}, \\
\mathbf{T}_{12s} &= -\sigma_1^2\sigma_2^2\underset{1\le d\le D}{\operatorname{diag}}\{\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\},\ \mathbf{T}_{21s} = (\mathbf{T}_{12s})', \\
\mathbf{T}_{22s} &= \sigma_2^2\underset{1\le d\le D}{\operatorname{diag}}\{\Omega_d(\rho)-\sigma_2^2\Omega_d(\rho)\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\}.
\end{aligned}$$

The product $\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'$ is calculated as follows.

$$\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r' = [\mathbf{Z}_{1r}\,\mathbf{Z}_{2r}]\,\mathbf{T}_s\left[\mathbf{Z}_{1r}'\,\mathbf{Z}_{2r}'\right]' = \mathbf{Z}_{1r}\mathbf{T}_{11s}\mathbf{Z}_{1r}' + \mathbf{Z}_{1r}\mathbf{T}_{12s}\mathbf{Z}_{2r}' + \mathbf{Z}_{2r}\mathbf{T}_{21s}\mathbf{Z}_{1r}' + \mathbf{Z}_{2r}\mathbf{T}_{22s}\mathbf{Z}_{2r}'.$$

It holds that

$$
\begin{aligned}
\mathbf{M}_{11}^{rr} &= \mathbf{Z}_{1r}\mathbf{T}_{11s}\mathbf{Z}_{1r}' = \sigma_1^2 \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{N_d-n_d}\} \underset{1\le d\le D}{\text{diag}} \{1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}\} \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{N_d-n_d}'\} \\
&= \sigma_1^2 \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{N_d-n_d}[1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}]\mathbf{1}_{N_d-n_d}'\}, \\
\mathbf{M}_{12}^{rr} &= \mathbf{Z}_{1r}\mathbf{T}_{12s}\mathbf{Z}_{2r}' = -\sigma_1^2\sigma_2^2 \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{N_d-n_d}\} \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}})\Omega_d(\rho)\} \\
&\cdot \underset{1\le d\le D}{\text{diag}} \{ \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}}')\} \\
&= -\sigma_1^2\sigma_2^2 \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{N_d-n_d}\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}})\Omega_d(\rho) \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}}')\}, \\
\mathbf{M}_{21}^{rr} &= (\mathbf{M}_{12}^{rr})', \\
\mathbf{M}_{22}^{rr} &= \mathbf{Z}_{2r}\mathbf{T}_{22s}\mathbf{Z}_{2r}' = \sigma_2^2 \underset{1\le d\le D}{\text{diag}} \{ \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}})\} \\
&\cdot \underset{1\le d\le D}{\text{diag}} \{\Omega_d(\rho)-\sigma_2^2\Omega_d(\rho) \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}})\Omega_d(\rho)\} \underset{1\le d\le D}{\text{diag}} \{ \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}}')\} \\
&= \sigma_2^2 \underset{1\le d\le D}{\text{diag}} \{ \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho) \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}}')\} \\
&- \sigma_2^4 \underset{1\le d\le D}{\text{diag}} \{ \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho) \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}}) \\
&\cdot \Omega_d(\rho) \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}}')\}.
\end{aligned}
$$

As

$$
\mathbf{a}_r' = \frac{1}{N_{dt}} \underset{1\le \ell\le D}{\text{col}'} \left[\delta_{d\ell} \underset{1\le k\le m_\ell}{\text{col}'} [\delta_{tk}\mathbf{1}_{N_{\ell k}-n_{\ell k}}']\right] \quad \text{y} \quad f_{dt} = \frac{n_{dt}}{N_{dt}},
$$

We obtain

$$
\begin{aligned}
\mathbf{a}_r'\mathbf{M}_{11}^{rr}\mathbf{a}_r &= \sigma_1^2\mathbf{a}_r' \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{N_d-n_d}[1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}]\mathbf{1}_{N_d-n_d}'\}\mathbf{a}_r \\
&= \sigma_1^2(1-f_{dt})^2[1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}] = \sigma^2\varphi_1(1-f_{dt})^2[1-\varphi_1\mathbf{1}_{n_d}'\Sigma_{ds}^{-1}\mathbf{1}_{n_d}], \\
\mathbf{a}_r'\mathbf{M}_{12}^{rr}\mathbf{a}_r &= -\sigma_1^2\sigma_2^2\mathbf{a}_r' \underset{1\le d\le D}{\text{diag}} \{\mathbf{1}_{N_d-n_d}\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}})\Omega_d(\rho) \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{N_{dk}-n_{dk}}')\}\mathbf{a}_r, \\
&= -\sigma^2\varphi_1\varphi_2(1-f_{dt})\mathbf{1}_{n_d}'\Sigma_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}})\Omega_d(\rho) \underset{1\le k\le m_d}{\text{col}} [\delta_{tk}(1-f_{dk})] \\
&= -\sigma^2\varphi_1\varphi_2(1-f_{dt})^2\mathbf{1}_{n_d}'\Sigma_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}} (\mathbf{1}_{n_{dk}})\Omega_d(\rho) \underset{1\le k\le m_d}{\text{col}} (\delta_{tk})
\end{aligned}
$$

$$\begin{aligned}
\mathbf{a}_r'\mathbf{M}_{22}^{rr}\mathbf{a}_r &= \sigma_2^2\mathbf{a}_r' \operatorname*{diag}_{1\le d\le D}\{\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho)\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{N_{dk}-n_{dk}}')\}\mathbf{a}_r \\
&- \sigma_2^4\mathbf{a}_r' \operatorname*{diag}_{1\le d\le D}\{\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho)\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1}\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{n_{dk}}) \\
&\cdot \Omega_d(\rho)\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{N_{dk}-n_{dk}}')\}\mathbf{a}_r \\
&= \sigma^2\varphi_2\operatorname*{col'}_{1\le k\le m_d}[(1-f_{dk})\delta_{tk}]\Omega_d(\rho)\operatorname*{col}_{1\le k\le m_d}[(1-f_{dk})\delta_{tk}] - \sigma^2\varphi_2^2\operatorname*{col'}_{1\le k\le m_d}[(1-f_{dk})\delta_{tk}]\Omega_d(\rho) \\
&\cdot \operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{n_{dk}}')\Sigma_{ds}^{-1}\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\operatorname*{col}_{1\le k\le m_d}[(1-f_{dk})\delta_{tk}] \\
&= \sigma^2\varphi_2(1-f_{dt})^2\operatorname*{col'}_{1\le k\le m_d}(\delta_{tk})\Omega_d(\rho)\operatorname*{col}_{1\le k\le m_d}(\delta_{tk}) - \sigma^2\varphi_2^2(1-f_{dt})^2\operatorname*{col'}_{1\le k\le m_d}(\delta_{tk})\Omega_d(\rho) \\
&\cdot \operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{n_{dk}}')\Sigma_{ds}^{-1}\operatorname*{diag}_{1\le k\le m_d}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\operatorname*{col}_{1\le k\le m_d}(\delta_{tk}).
\end{aligned}$$

Finally,

$$g_1(\theta) = \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r = \mathbf{a}_r'\mathbf{M}_{11}^{rr}\mathbf{a}_r + 2\mathbf{a}_r'\mathbf{M}_{12}^{rr}\mathbf{a}_r + \mathbf{a}_r'\mathbf{M}_{22}^{rr}\mathbf{a}_r.$$

Calculation of $g_2(\theta)$

The formula for $g_2(\theta)$ is

$$g_2(\theta) = [\mathbf{a}_r'\mathbf{X}_r - \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r - \mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r] = [\mathbf{a}_{21}' - \mathbf{a}_{22}']\mathbf{Q}_s[\mathbf{a}_{21} - \mathbf{a}_{22}],$$

where $\mathbf{Q}_s = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1} = \sigma^2\left(\sum_{d=1}^D \mathbf{X}_{ds}'\Sigma_{ds}^{-1}\mathbf{X}_{ds}\right)^{-1}$ and $\Sigma_{es}^{-1} = \sigma^{-2}\mathbf{W}_s$. On the one hand

$$\mathbf{a}_{21}' = \mathbf{a}_r'\mathbf{X}_r = \frac{1}{N_{dt}}\mathbf{1}_{N_{dt}-n_{dt}}'\mathbf{X}_{dt,r} = \frac{1}{N_{dt}}\sum_{j\in r}\mathbf{x}_{dtj} = (1-f_{dt})\overline{\mathbf{X}}_{dt}^*,\ \text{where}\ \overline{\mathbf{X}}_{dt}^* = \frac{1}{N_{dt}-n_{dt}}\sum_{j\in r}\mathbf{x}_{dtj}.$$

On the other hand

$$\mathbf{a}_{22}' = \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\Sigma_{es}^{-1}\mathbf{X}_s = \sigma^{-2}\mathbf{a}_r'(\mathbf{M}_{11}^{rs} + \mathbf{M}_{12}^{rs} + \mathbf{M}_{21}^{rs} + \mathbf{M}_{22}^{rs})\mathbf{W}_s\mathbf{X}_s = \mathbf{G}_{11} + \mathbf{G}_{12} + \mathbf{G}_{21} + \mathbf{G}_{22},$$

where

$$\begin{aligned}
&\mathbf{M}_{11}^{rs} = \mathbf{Z}_{1r}\mathbf{T}_{11s}\mathbf{Z}_{1s}',\ \mathbf{M}_{12}^{rs} = \mathbf{Z}_{1r}\mathbf{T}_{12s}\mathbf{Z}_{2s}' \\
&\mathbf{M}_{21}^{rs} = \mathbf{Z}_{2r}\mathbf{T}_{21s}\mathbf{Z}_{1s}' = (\mathbf{M}_{12}^{sr})',\ \mathbf{M}_{22}^{rs} = \mathbf{Z}_{2r}\mathbf{T}_{22s}\mathbf{Z}_{2s}'.
\end{aligned}$$

let us define $\mathbf{w}'_{n_{dk}} = (w_{dk1}, \dots, w_{dkn_{dk}})$. It holds that

$$\begin{aligned}
\mathbf{G}_{11} &= \sigma^{-2}\mathbf{a}'_r\mathbf{M}_{11}^{rs}\mathbf{W}_s\mathbf{X}_s = \frac{\sigma_1^2}{\sigma^2 N_{dt}} \underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}]\mathbf{1}_{N_d-n_d}[1-\sigma_1^2\mathbf{1}'_{n_d}\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}]\mathbf{1}'_{n_d}\mathbf{W}_{ds}\mathbf{X}_{ds}\\
&= \varphi_1(1-f_{dt})[1-\varphi_1\mathbf{1}'_{n_d}\Sigma_{ds}^{-1}\mathbf{1}_{n_d}]\sum_{k=1}^{m_d}\mathbf{w}'_{n_{dk}}\mathbf{X}_{dk,s},\\
\mathbf{G}_{12} &= \sigma^{-2}\mathbf{a}'_r\mathbf{M}_{12}^{rs}\mathbf{W}_s\mathbf{X}_s\\
&= -\frac{\sigma_1^2\sigma_2^2}{\sigma^2 N_{dt}} \underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}]\mathbf{1}_{N_d-n_d}\mathbf{1}'_{n_d}\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{W}_{ds}\mathbf{X}_{ds}\\
&= -\varphi_1\varphi_2(1-f_{dt})\mathbf{1}'_{n_d}\Sigma_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\underset{1\le k\le m_d}{\text{col}}(\mathbf{w}'_{n_{dj}}\mathbf{X}_{dk,s}),\\
\mathbf{G}_{21} &= \sigma^{-2}\mathbf{a}'_r\mathbf{M}_{21}^{rs}\mathbf{W}_s\mathbf{X}_s\\
&= -\frac{\sigma_1^2\sigma_2^2}{\sigma^2 N_{dt}} \underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}]\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{W}_{ds}\mathbf{X}_{ds}\\
&= -\varphi_1\varphi_2(1-f_{dt})\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}]\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\Sigma_{ds}^{-1}\mathbf{1}_{n_d}\sum_{k=1}^{m_d}\mathbf{w}'_{n_{dk}}\mathbf{X}_{dk,s},\\
\mathbf{G}_{22} &= \sigma^{-2}\mathbf{a}'_r\mathbf{M}_{22}^{rs}\mathbf{W}_s\mathbf{X}_s\\
&= \frac{\sigma_2^2}{\sigma^2 N_{dt}} \underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}]\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\\
&- \sigma_2^2\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\}\mathbf{W}_{ds}\mathbf{X}_{ds}\\
&= \varphi_2(1-f_{dt})\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}]\Omega_d(\rho)\left[\mathbf{I}_{m_d}-\varphi_2\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\Sigma_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\Omega_d(\rho)\right]\\
&\cdot \underset{1\le k\le m_d}{\text{col}}(\mathbf{w}'_{n_{dk}}\mathbf{X}_{dk,s}).
\end{aligned}$$

Calculation of $g_3(\theta)$

The formula for $g_3(\theta)$ is

$$g_3(\theta) \approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}_s(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where

$$\begin{aligned}
\mathbf{b}' &= \mathbf{a}'_r\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}'_s\mathbf{V}_s^{-1} = \mathbf{a}'_r[\mathbf{Z}_{1r},\mathbf{Z}_{2r}]\,\text{diag}\left\{\sigma_1^2\mathbf{I}_D,\sigma_2^2\Omega(\rho)\right\}[\mathbf{Z}'_{1s},\mathbf{Z}'_{2s}]'\mathbf{V}_s^{-1}\\
&= \mathbf{a}'_r\left[\sigma_1^2\mathbf{Z}_{1r}\mathbf{Z}'_{1s}+\sigma_2^2\mathbf{Z}_{2r}\Omega(\rho)\mathbf{Z}'_{2s}\right]\mathbf{V}_s^{-1} = \sigma_1^2\mathbf{a}'_r\mathbf{Z}_{1r}\mathbf{Z}'_{1s}\mathbf{V}_s^{-1}+\sigma_2^2\mathbf{a}'_r\mathbf{Z}_{2r}\Omega(\rho)\mathbf{Z}'_{2s}\mathbf{V}_s^{-1}\\
&= \mathbf{b}'_1+\mathbf{b}'_2 = \underset{1\le\ell\le D}{\text{col}'}[\delta_{d\ell}\mathbf{b}'_{1\ell}]+\underset{1\le\ell\le D}{\text{col}'}[\delta_{d\ell}\mathbf{b}'_{2\ell}],
\end{aligned}$$

$$\begin{aligned}
\mathbf{b}'_{1d} &= \frac{\sigma^2\varphi_1}{N_{dt}} \underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}]\mathbf{1}_{N_d-n_d}\mathbf{1}'_{n_d}\mathbf{V}_{ds}^{-1} = \varphi_1(1-f_{dt})\mathbf{1}'_{n_d}\Sigma_{ds}^{-1},\\
\mathbf{b}'_{2d} &= \frac{\sigma^2\varphi_2}{N_{dt}} \underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}]\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{V}_{ds}^{-1},\\
&= \varphi_2(1-f_{dt})\underset{1\le d\le D}{\text{col}'}[\delta_{tk}]\,\Omega_d(\rho)\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\Sigma_{ds}^{-1}.
\end{aligned}$$

Let us define $\mathbf{A}_{ds} = \Omega_d^{-1}(\rho) + \varphi_2 \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{W}_{ds}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})$. Then

$$\begin{aligned}
\Sigma_{ds}^{-1} &= \mathbf{L}_{ds}^{-1} - \frac{\varphi_1}{1+\varphi_1\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1},\\
\mathbf{L}_{ds}^{-1} &= \mathbf{W}_{ds} - \varphi_2\mathbf{W}_{ds}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{W}_{ds}.
\end{aligned}$$

By applying the formula $\frac{\partial \mathbf{A}^{-1}}{\partial\gamma} = -\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial\gamma}\mathbf{A}^{-1}$, we calculate the partial derivatives of $\mathbf{L}_{ds}^{-1}$.

$$\begin{aligned}
\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\sigma^2} &= \mathbf{0}_{n_d\times n_d}, \quad \frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_1} = \mathbf{0}_{n_d\times n_d},\\
\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2} &= -\mathbf{W}_{ds}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{W}_{ds} + \varphi_2\mathbf{W}_{ds}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\\
&\cdot \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{W}_{ds}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{W}_{ds},\\
\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\rho} &= -\varphi_2\mathbf{W}_{ds}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\Omega_d^{-1}(\rho)\Omega'(\rho)\Omega_d^{-1}(\rho)\mathbf{A}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}'_{n_{dk}})\mathbf{W}_{ds}.
\end{aligned}$$

The partial derivatives of Σ_{ds}^{-1} are

$$\begin{aligned}
\frac{\partial\Sigma_{ds}^{-1}}{\partial\sigma^2} &= \mathbf{0}_{n_d\times n_d}\\
\frac{\partial\Sigma_{ds}^{-1}}{\partial\varphi_1} &= -\frac{1}{[1+\varphi_1\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}]^2}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1},\\
\frac{\partial\Sigma_{ds}^{-1}}{\partial\varphi_2} &= \frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2} + \frac{\varphi_1^2\mathbf{1}'_{n_d}\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2}\mathbf{1}_{n_d}}{[1+\varphi_1\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}]^2}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\\
&- \frac{\varphi_1}{1+\varphi_1\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}}\left[\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1} + \mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2}\right],\\
\frac{\partial\Sigma_{ds}^{-1}}{\partial\rho} &= \frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\rho} + \frac{\varphi_1^2\mathbf{1}'_{n_d}\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\rho}\mathbf{1}_{n_d}}{[1+\varphi_1\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}]^2}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\\
&- \frac{\varphi_1}{1+\varphi_1\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}}\left[\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\rho}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{L}_{ds}^{-1} + \mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\rho}\right].
\end{aligned}$$

Let us define $\theta=(\theta_1,\theta_2,\theta_3,\theta_4)=(\sigma^2,\varphi_1,\varphi_2,\rho)$. The partial derivatives of $\mathbf{b}'_{1d}$ y $\mathbf{b}'_{2d}$ are

$$
\begin{aligned}
\frac{\partial \mathbf{b}'_{1d}}{\partial \sigma^2} &= \mathbf{0}_{1\times n_d},\\
\frac{\partial \mathbf{b}'_{1d}}{\partial \varphi_1} &= (1-f_{dt})\mathbf{1}'_{n_d}\left[\Sigma_{ds}^{-1}+\varphi_1\frac{\partial \Sigma_{ds}^{-1}}{\partial \varphi_1}\right],\\
\frac{\partial \mathbf{b}'_{1d}}{\partial \theta_\ell} &= \varphi_1(1-f_{dt})\mathbf{1}'_{n_d}\frac{\partial \Sigma_{ds}^{-1}}{\partial \theta_\ell},\ \ell=3,4,\\
\frac{\partial \mathbf{b}'_{2d}}{\partial \sigma^2} &= \mathbf{0}_{1\times n_d},\\
\frac{\partial \mathbf{b}'_{2d}}{\partial \varphi_1} &= \varphi_2(1-f_{dt})\underset{1\le k\le m_d}{\mathrm{col}'}[\delta_{tk}]\Omega_d(\rho)\underset{1\le k\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dk}})\frac{\partial \Sigma_{ds}^{-1}}{\partial \varphi_1},\\
\frac{\partial \mathbf{b}'_{2d}}{\partial \varphi_2} &= (1-f_{dt})\underset{1\le k\le m_d}{\mathrm{col}'}[\delta_{tk}]\Omega_d(\rho)\underset{1\le k\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dk}})\left[\Sigma_{ds}^{-1}+\varphi_2\frac{\partial \Sigma_{ds}^{-1}}{\partial \varphi_2}\right],\\
\frac{\partial \mathbf{b}'_{2d}}{\partial \rho} &= \varphi_2(1-f_{dt})\underset{1\le k\le m_d}{\mathrm{col}'}[\delta_{tk}]\left[\Omega'_d(\rho)\underset{1\le k\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dk}})\Sigma_{ds}^{-1}+\Omega_d(\rho)\underset{1\le k\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dk}})\frac{\partial \Sigma_{ds}^{-1}}{\partial \rho}\right].
\end{aligned}
$$

Let $\mathbf{Q}=(q_{ab})_{a,b=1,\dots,4}$ be the matrix with elements

$$
q_{ab}=\left(\frac{\partial \mathbf{b}'_{1d}}{\partial \theta_a}+\frac{\partial \mathbf{b}'_{2d}}{\partial \theta_a}\right)\sigma^2\Sigma_{ds}\left(\frac{\partial \mathbf{b}'_{1d}}{\partial \theta_b}+\frac{\partial \mathbf{b}'_{2d}}{\partial \theta_b}\right)',\quad a,b=1,2,3,4,
$$

and F_{θ_a,θ_b}'s be the elements of the REML Fisher information matrix. Then

$$
\begin{aligned}
g_3(\theta) &\approx \mathrm{tr}\left\{\mathbf{Q}E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}\\
&\approx \mathrm{tr}\left\{\begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14}\\ q_{21} & q_{22} & q_{23} & q_{24}\\ q_{31} & q_{32} & q_{33} & q_{34}\\ q_{41} & q_{42} & q_{43} & q_{44}\end{pmatrix}\begin{pmatrix} F_{\sigma^2\sigma^2} & F_{\sigma^2\varphi_1} & F_{\sigma^2\varphi_2} & F_{\sigma^2\rho}\\ F_{\varphi_1\sigma^2} & F_{\varphi_1\varphi_1} & F_{\varphi_1\varphi_2} & F_{\varphi_1\rho}\\ F_{\varphi_2\sigma^2} & F_{\varphi_2\varphi_1} & F_{\varphi_2\varphi_2} & F_{\varphi_2\rho}\\ F_{\rho\sigma^2} & F_{\rho\varphi_1} & F_{\rho\varphi_2} & F_{\rho\rho}\end{pmatrix}^{-1}\right\}
\end{aligned}
$$

Calculation of $g_4(\theta)$

We recall that $g_4(\theta)=\mathbf{a}'_r\mathbf{V}_{er}\mathbf{a}_r$, where

$$
\begin{aligned}
\mathbf{a}'_r &= \frac{1}{N_{dt}}\underset{1\le \ell\le D}{\mathrm{col}'}\left[\delta_{d\ell}\underset{1\le k\le m_\ell}{\mathrm{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}-n_{\ell k}}]\right],\\
\mathbf{V}_{er}^{-1} &= \sigma^{-2}\mathbf{W}_r=\sigma^{-2}\underset{1\le d\le D}{\mathrm{diag}}\{\mathbf{W}_{dr}\}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
g_4(\theta) &= \frac{1}{N_{dt}}\underset{1\le \ell\le D}{\mathrm{col}'}\left[\delta_{d\ell}\underset{1\le k\le m_\ell}{\mathrm{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}-n_{\ell k}}]\right]\sigma^2\underset{1\le d\le D}{\mathrm{diag}}\{\mathbf{W}_{dr}^{-1}\}\frac{1}{N_{dt}}\underset{1\le \ell\le D}{\mathrm{col}}\left[\delta_{d\ell}\underset{1\le k\le m_\ell}{\mathrm{col}}[\delta_{tk}\mathbf{1}_{N_{\ell k}-n_{\ell k}}]\right]\\
&= \frac{\sigma^2}{N_{dt}^2}\mathbf{1}'_{N_{dt}-n_{dt}}\underset{j\in r}{\mathrm{diag}}\{w_{dtj}^{-1}\}\mathbf{1}_{N_{dt}-n_{dt}}=\frac{\sigma^2}{N_{dt}^2}\sum_{j\in r_{dt}}\frac{1}{w_{dtj}}.
\end{aligned}
$$

9.2 Unit-level model with independent time effects

9.2.1 Introduction

Let us consider a version of linear mixed model (1.1) having two nested random factors. Assume that the first factor has D levels and, for each of these levels d ($d = 1,\dots,D$), the second one has m_d levels. The model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{W}^{-1/2}\mathbf{e}, \qquad (9.6)$$

where $\mathbf{u}_1 = \mathbf{u}_{1,D\times 1} \sim N(0,\sigma_1^2\mathbf{I}_D)$, $\mathbf{u}_2 = \mathbf{u}_{2,M\times 1} \sim N(0,\sigma_2^2\mathbf{I}_M)$ and $\mathbf{e} = \mathbf{e}_{n\times 1} \sim N(0,\sigma_0^2\mathbf{I}_n)$ are independent, $\mathbf{y} = \mathbf{y}_{n\times 1}$, $\mathbf{X} = \mathbf{X}_{n\times p}$ with $\mathrm{r}(\mathbf{X}) = p$, $\beta = \beta_{p\times 1}$, $\mathbf{Z}_1 = \underset{1\le d\le D}{\mathrm{diag}}\,(\mathbf{1}_{n_d})_{n\times D}$, $\mathbf{Z}_2 = \underset{1\le d\le D}{\mathrm{diag}}\,(\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}}))_{n\times M}$, $M = \sum_{d=1}^D m_d$, $n = \sum_{d=1}^D n_d$, $n_d = \sum_{t=1}^{m_d} n_{dt}$, $\mathbf{I}_a$ is the $a\times a$ identity matrix, $\mathbf{1}_a$ is the $a\times 1$ vector with all its elements equal to 1, $\mathbf{W} = \underset{1\le d\le D}{\mathrm{diag}}\,(\mathbf{W}_d)$, $\mathbf{W}_d = \underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{W}_{dt})$, $\mathbf{W}_{dt} = \underset{1\le j\le n_{dt}}{\mathrm{diag}}\,(w_{dtj})_{n\times n}$ with known $w_{dtj} > 0$, $d = 1,\dots,D$, $t = 1,\dots,m_d$, $j = 1,\dots,n_{dt}$. Model (9.6) can alternatively be written in the form

$$y_{dtj} = \mathbf{x}_{dtj}\beta + u_{1,d} + u_{2,dt} + w_{dtj}^{-1/2}e_{dtj}, \quad d = 1,\dots,D, t = 1,\dots,m_d, j = 1,\dots,n_{dt}, \qquad (9.7)$$

where y_{dtj} is the target variable for the sample unit j, time t and domain d, and $\mathbf{x}_{dtj}$ is the row (d,t,j) of matrix $\mathbf{X}$. In what follows we use the alternative parameters

$$\sigma^2 = \sigma_0^2, \quad \varphi_1 = \frac{\sigma_1^2}{\sigma_0^2}, \quad \varphi_2 = \frac{\sigma_2^2}{\sigma_0^2}.$$

Let $\sigma = (\sigma^2,\varphi_1,\varphi_2)$ be the vector of variance components, with $\sigma^2 > 0$, $\varphi_1 > 0$ and $\varphi_2 > 0$. If σ is known, the BLUE of $\beta = (\beta_1,\dots,\beta_p)'$ and the BLUP of $\mathbf{u} = (\mathbf{u}_1',\mathbf{u}_2')'$ are

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{y} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\beta}\right). \qquad (9.8)$$

Formulas (9.8) are not computationally efficient because they require the inversion of the $n\times n$ matrix $\mathbf{V}$. By calculating the inversion of $\mathbf{V}$ new formulas are obtained. Under model (9.6), we have $\mathrm{var}(\mathbf{u}_1) = \sigma^2\varphi_1\mathbf{I}_D$, $\mathrm{var}(\mathbf{u}_2) = \sigma^2\varphi_2\mathbf{I}_M$, $\mathrm{var}(\mathbf{e}) = \sigma^2\mathbf{I}_n$ and

$$\mathbf{V} = \mathrm{var}(\mathbf{y}) = \mathbf{Z}_1\mathrm{var}(\mathbf{u}_1)\mathbf{Z}_1{}' + \mathbf{Z}_2\mathrm{var}(\mathbf{u}_2)\mathbf{Z}_2{}' + \sigma^2\mathbf{W}^{-1} = \sigma^2\Sigma = \sigma^2\mathrm{diag}(\Sigma_1,\dots,\Sigma_D),$$

where

$$\Sigma_d = \varphi_1\mathbf{1}_{n_d}\mathbf{1}_{n_d}' + \varphi_2\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}})\mathbf{I}_{m_d}\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}}') + \mathbf{W}_d^{-1} = \varphi_1\mathbf{1}_{n_d}\mathbf{1}_{n_d}' + \mathbf{L}_d, \; d = 1,\dots,D.$$

To calculate$\mathbf{L}_d^{-1}$ we use the formula

$$(A + CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}$$

with $A = \mathbf{W}_d^{-1}$, $C = \varphi_2\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}})$, $B = \mathbf{I}_{m_d}$ y $D = \underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}}')$. We obtain

$$\begin{aligned}\mathbf{L}_d^{-1} \;&=\; \mathbf{W}_d - \varphi_2\mathbf{W}_d\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}})\left[\mathbf{I}_{m_d} + \varphi_2\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}}')\mathbf{W}_d\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}})\right]^{-1}\\ &\cdot\; \underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}}')\mathbf{W}_d.\end{aligned}$$

To calculate Σ_d^{-1} we use the formula

$$(\mathbf{A}+\mathbf{u}\mathbf{v}')^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}'\mathbf{A}^{-1}}{1+\mathbf{v}'\mathbf{A}^{-1}\mathbf{u}}$$

with $A=\mathbf{L}_d$, $\mathbf{u}=\varphi_1\mathbf{1}_{n_d}$, $\mathbf{v}'=\mathbf{1}'_{n_d}$. We obtain

$$\Sigma_d^{-1}=\mathbf{L}_d^{-1}-\frac{\varphi_1}{1+\varphi_1\mathbf{1}'_{n_d}\mathbf{L}_d^{-1}\mathbf{1}_{n_d}}\mathbf{L}_d^{-1}\mathbf{1}_{n_d}\mathbf{1}'_{n_d}\mathbf{L}_d^{-1}.$$

The formula for $\widehat{\beta}$ is

$$\widehat{\beta}=\left(\sum_{d=1}^{D}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{X}_d\right)^{-1}\left(\sum_{d=1}^{D}\mathbf{X}'_d\Sigma_d^{-1}\mathbf{y}_d\right) \tag{9.9}$$

where $\mathbf{X}=\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$, and $\mathbf{y}=\underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$. The formula for $\widehat{\mathbf{u}}$ is

$$\begin{aligned}
\widehat{\mathbf{u}} &= \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\widehat{\beta}\right)=\begin{pmatrix}\sigma_1^2\mathbf{I}_D & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}_M\end{pmatrix}\begin{bmatrix}\mathbf{Z}'_1\\ \mathbf{Z}'_2\end{bmatrix}\underset{1\le d\le D}{\text{diag}}\left(\mathbf{V}_d^{-1}\right)\underset{1\le d\le D}{\text{col}}\left[\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right]\\
&= \begin{bmatrix}\varphi_1\underset{1\le d\le D}{\text{diag}}\left(\mathbf{1}'_{n_d}\right)\underset{1\le d\le D}{\text{diag}}\left(\Sigma_d^{-1}\right)\underset{1\le d\le D}{\text{col}}\left[\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right]\\ \varphi_2\underset{1\le d\le D}{\text{diag}}\left(\underset{1\le t\le m_d}{\text{diag}}\left(\mathbf{1}'_{n_{dt}}\right)\right)\underset{1\le d\le D}{\text{diag}}\left(\Sigma_d^{-1}\right)\underset{1\le d\le D}{\text{col}}\left[\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right]\end{bmatrix}\\
&= \begin{bmatrix}\varphi_1\underset{1\le d\le D}{\text{col}}\left[\mathbf{1}'_{n_d}\Sigma_d^{-1}\left(\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right)\right]\\ \varphi_2\underset{1\le d\le D}{\text{col}}\left[\underset{1\le t\le m_d}{\text{diag}}\left(\mathbf{1}'_{n_{dt}}\right)\Sigma_d^{-1}\left(\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}\right)\right]\end{bmatrix}.
\end{aligned}$$

9.2.2 REML estimators of model parameters

The REML log-likelihood function is

$$l_{reml}(\sigma)=-\frac{1}{2}(n-p)\log 2\pi-\frac{1}{2}(n-p)\log\sigma^2-\frac{1}{2}\log|\mathbf{K}'\Sigma\mathbf{K}|-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where

$$\mathbf{P}=\mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}'=\Sigma^{-1}-\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1},\ \mathbf{K}=\mathbf{W}-\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}.$$

By taking partial derivatives with respect to σ^2, φ_1^2 and φ_2^2 we get the components of the vectors of scores $S(\sigma)$.

$$\begin{aligned}
S_{\sigma^2} &= -\frac{n-p}{2\sigma^2}+\frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{y},\\
S_{\varphi_1} &= -\frac{1}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}'_1\}+\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}'_1\mathbf{P}\mathbf{y},\\
S_{\varphi_2} &= -\frac{1}{2}\text{tr}\{\mathbf{P}\mathbf{Z}_2\mathbf{Z}'_2\}+\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\mathbf{Z}'_2\mathbf{P}\mathbf{y},
\end{aligned}$$

Second partial derivatives of the REML log-likelihood function are

$$\begin{aligned}
H_{\sigma^2\sigma^2} &= \frac{n-p}{2\sigma^4} - \frac{1}{\sigma^6}\mathbf{y}'\mathbf{P}\mathbf{y}, \quad H_{\sigma^2\varphi_1} = -\frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y},\\
H_{\sigma^2\varphi_2} &= -\frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{y},\\
H_{\varphi_1\varphi_1} &= \frac{1}{2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\} - \frac{1}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y},\\
H_{\varphi_1\varphi_2} &= \frac{1}{2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\} - \frac{1}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{y},\\
H_{\varphi_2\varphi_2} &= \frac{1}{2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\} - \frac{1}{\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{y}.
\end{aligned}$$

By taking expectations, changing the sign and taking into account that $\mathbf{P}\mathbf{X}=\mathbf{0}$ and $\mathbf{P}\Sigma\mathbf{P}=\mathbf{P}$, we obtain the elements of the Fisher information matrix

$$\begin{aligned}
F_{\sigma^2\sigma^2} &= -\frac{n-p}{2\sigma^4} + \frac{1}{\sigma^4}\,\mathrm{tr}\{\mathbf{P}\Sigma\} = \frac{n-p}{2\sigma^4}, \quad F_{\sigma^2\varphi_1} = \frac{1}{2\sigma^2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\},\\
F_{\sigma^2\varphi_2} &= \frac{1}{2\sigma^2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\}, \quad F_{\varphi_1\varphi_1} = \frac{1}{2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\},\\
F_{\varphi_1\varphi_2} &= \frac{1}{2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\}, \quad F_{\varphi_2\varphi_2} = \frac{1}{2}\,\mathrm{tr}\{\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\}.
\end{aligned}$$

The updating formula of the Fisher-scoring algorithm is

$$\sigma^{k+1} = \sigma^k + \mathbf{F}^{-1}(\sigma^k)\mathbf{S}(\sigma^k).$$

As algorithm seeds we can use the Henderson 3 estimators $\sigma_0^{2(0)}$, $\sigma_1^{2(0)}$ and $\sigma_2^{2(0)}$. Estimator $\widehat{\beta}_{reml}$ is calculated by applying the formula (9.9).

Observation 9.2.1. From equation $S_{\sigma^2}=\mathbf{0}$, we get

$$\widehat{\sigma}^2 = \frac{1}{n-p}\mathbf{y}'\mathbf{P}\mathbf{y}, \tag{9.10}$$

and we can introduce an algorithm updating σ^2 with (9.10) and $\varphi=(\varphi_1,\varphi_2)'$ with

$$\varphi^{i+1} = \varphi^i + F(\varphi^i)^{-1}S(\varphi^i).$$

Matrix calculations

In what follows we presents computationally efficient formulas for the scores and the Fisher information components. These formulas avoid the construction of $n\times n$ matrices. Let us define

$$\Sigma = \underset{1\le d\le D}{\mathrm{diag}}(\Sigma_d),\ \mathbf{X} = \underset{1\le d\le D}{\mathrm{col}}(\mathbf{X}_d),\ \mathbf{y} = \underset{1\le d\le D}{\mathrm{col}}(\mathbf{y}_d),\ \mathbf{R} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{X}_d\right)^{-1}$$

so that

$$\mathbf{P} = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}\mathbf{R}\mathbf{X}'\Sigma^{-1} = \underset{1\le d\le D}{\mathrm{diag}}(\Sigma_d^{-1}) - \underset{1\le d\le D}{\mathrm{col}}(\Sigma_d^{-1}\mathbf{X}_d)\mathbf{R}\underset{1\le d\le D}{\mathrm{col}'}(\mathbf{X}_d'\Sigma_d^{-1})$$

The components of the score vector are

$$
\begin{aligned}
S_{\sigma^2} &= -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{y}_d - \frac{1}{2\sigma^4}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right),\\
S_{\varphi_1} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\} + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y} = -\frac{1}{2}\sum_{d=1}^{D}\mathbf{1}_{n_d}'[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}]\mathbf{1}_{n_d}\\
&+ \frac{1}{2\sigma^2}\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{y}_d - \frac{1}{\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right)\\
&+ \frac{1}{2\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right),\\
S_{\varphi_2} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\} + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{y}\\
&= -\frac{1}{2}\sum_{d=1}^{D}\operatorname{tr}\left\{\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}]\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\right\}\\
&+ \frac{1}{2\sigma^2}\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{y}_d\\
&- \frac{1}{\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right)\\
&+ \frac{1}{2\sigma^2}\left(\sum_{d=1}^{D}\mathbf{y}_d'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right)\\
&\cdot\ \mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{y}_d\right),
\end{aligned}
$$

The elements of the Fisher iformation matrix are

$$
\begin{aligned}
F_{\sigma^2\sigma^2} &= \frac{n-p}{2\sigma^2}\\
F_{\sigma^2\varphi_1} &= \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\} = \frac{1}{2\sigma^2}\sum_{d=1}^{D}\mathbf{1}_{n_d}'\left[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\right]\mathbf{1}_{n_d}\\
F_{\sigma^2\varphi_2} &= \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\} = \frac{1}{2\sigma^2}\sum_{d=1}^{D}\operatorname{tr}\{\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}')\left[\Sigma_d^{-1}-\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\right]\underset{1\le t\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dt}})\}
\end{aligned}
$$

$$
\begin{aligned}
F_{\varphi_1\varphi_1} &= \frac{1}{2}\operatorname{tr}\{\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\} = \frac{1}{2}\sum_{d=1}^{D}(\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{1}_{n_d})^2 - \sum_{d=1}^{D}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\\
&+ \frac{1}{2}\sum_{d=1}^{D}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{1}_{n_d},
\end{aligned}
$$

$$\begin{aligned}
F_{\varphi_1\varphi_2} &= \frac{1}{2}\text{tr}\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\} = \frac{1}{2}\sum_{d=1}^{D}\text{tr}\{\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\} \\
&- \sum_{d=1}^{D}\text{tr}\{\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\} \\
&+ \frac{1}{2}\sum_{d=1}^{D}\text{tr}\{\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\Sigma_d^{-1}\mathbf{X}_d\right)\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\},
\end{aligned}$$

$$\begin{aligned}
F_{\varphi_2\varphi_2} &= \frac{1}{2}\text{tr}\{\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\} \\
&= \frac{1}{2}\sum_{d=1}^{D}\text{tr}\{\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\} \\
&- \sum_{d=1}^{D}\text{tr}\{\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\} \\
&+ \frac{1}{2}\sum_{d=1}^{D}\text{tr}\{\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\mathbf{R}\left(\sum_{d=1}^{D}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}')\Sigma_d^{-1}\mathbf{X}_d\right) \\
&\cdot \mathbf{R}\mathbf{X}_d'\Sigma_d^{-1}\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}})\},
\end{aligned}$$

9.2.3 Henderson 3 estimators of model parameters

In this section we present the *fitting constants method* to estimate the variance components. This method is also known as *Henderson 3* (H3) since it was introduced by Henderson (1953). To apply the H3 method, we treat the factors $\mathbf{u}_1$ and $\mathbf{u}_2$ as fixed and we fit the model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{W}^{-1/2}\mathbf{e} \tag{9.11}$$

by using the weighted least squared method. To avoid collinearity between the fixed effects of model (9.11) we equate to zero the parameters of the last time instants within domains, i.e. $u_{2,dm_d} = 0$, $d = 1,\ldots,D$. This is equivalent to delete columns $\sum_{j=1}^{d} m_j$, $d = 1,\ldots,D$, from matrix $\mathbf{Z}_2 = \underset{1\le d\le D}{\text{diag}}\ (\underset{1\le t\le m_d}{\text{diag}}\ (\mathbf{1}_{n_{dt}}))$. Therefore, we use the following incidence matrices

$$\mathbf{Z}_1 = \underset{1\le d\le D}{\text{diag}}\ (\mathbf{1}_{n_d})_{n\times D} \quad \text{and} \quad \widetilde{\mathbf{Z}}_2 = \underset{1\le d\le D}{\text{diag}}\left(\text{col}\{\underset{1\le t\le m_d-1}{\text{diag}}\ (\mathbf{1}_{n_{dt}}), \mathbf{0}_{n_{dm_d}\times(m_d-1)}\}\right).$$

The H3 estimators are

$$\begin{aligned}
\widehat{\sigma}_0^2 &= \frac{\mathbf{y}'\mathbf{M}_3\mathbf{y}}{n-\text{r}(\mathbf{X}^{(3)})} = \frac{\mathbf{y}'\mathbf{M}_3\mathbf{y}}{n-p-M},\\
\widehat{\sigma}_2^2 &= \frac{\mathbf{y}'\mathbf{M}_2\mathbf{y}-\mathbf{y}'\mathbf{M}_3\mathbf{y}-\widehat{\sigma}_0^2\left[\text{r}(\mathbf{X}^{(3)})-\text{r}(\mathbf{X}^{(2)})\right]}{\text{tr}\{\mathbf{L}_2\}} = \frac{\mathbf{y}'\mathbf{M}_2\mathbf{y}-\mathbf{y}'\mathbf{M}_3\mathbf{y}-(M-D)\widehat{\sigma}_0^2}{\text{tr}\{\mathbf{L}_2\}},\\
\widehat{\sigma}_1^2 &= \frac{\mathbf{y}'\mathbf{M}_1\mathbf{y}-\mathbf{y}'\mathbf{M}_3\mathbf{y}-\widehat{\sigma}_0^2\left[\text{r}(\mathbf{X}^{(3)})-\text{r}(\mathbf{X}^{(1)})\right]-\widehat{\sigma}_2^2\text{tr}\{\mathbf{L}_2\}}{\text{tr}\{\mathbf{L}_1\}}\\
&= \frac{\mathbf{y}'\mathbf{M}_1\mathbf{y}-\mathbf{y}'\mathbf{M}_3\mathbf{y}-M\widehat{\sigma}_0^2-\widehat{\sigma}_2^2\text{tr}\{\mathbf{L}_2\}}{\text{tr}\{\mathbf{L}_1\}},
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{X}^{(1)} &= \mathbf{X}, \qquad \mathbf{X}^{(2)} = (\mathbf{X},\mathbf{Z}_1), \qquad \mathbf{X}^{(3)} = (\mathbf{X},\mathbf{Z}_1,\widetilde{\mathbf{Z}}_2)\\
\mathbf{M}_1 &= \mathbf{W}-\mathbf{W}\mathbf{X}^{(1)}(\mathbf{X}^{(1)t}\mathbf{W}\mathbf{X}^{(1)})^{-1}\mathbf{X}^{(1)t}\mathbf{W}, \quad \mathbf{L}_1 = \mathbf{Z}_1'\mathbf{M}_1\mathbf{Z}_1,\\
\mathbf{M}_2 &= \mathbf{W}-\mathbf{W}\mathbf{X}^{(2)}(\mathbf{X}^{(2)t}\mathbf{W}\mathbf{X}^{(2)})^{-1}\mathbf{X}^{(2)t}\mathbf{W}, \quad \mathbf{L}_2 = \widetilde{\mathbf{Z}}_2'\mathbf{M}_2\widetilde{\mathbf{Z}}_2\\
\mathbf{M}_3 &= \mathbf{W}-\mathbf{W}\mathbf{X}^{(3)}(\mathbf{X}^{(3)t}\mathbf{W}\mathbf{X}^{(3)})^{-1}\mathbf{X}^{(3)t}\mathbf{W}
\end{aligned}$$

The above formulas are not computationally efficient because they require the inversion of $p+M$ matrices and M is in general quite large. In what follows more efficient formulas are given.

Calculations for $\mathbf{M}_1$

Let us define $\mathbf{C} = (\mathbf{X}^{(1)t}\mathbf{W}\mathbf{X}^{(1)})^{-1} = \left(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{X}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\right)^{-1}$. Then

$$\begin{aligned}
\mathbf{y}'\mathbf{M}_1\mathbf{y} &= \mathbf{y}'\mathbf{W}\mathbf{y} - \mathbf{y}'\mathbf{W}\mathbf{X}\mathbf{C}\mathbf{X}'\mathbf{W}\mathbf{y}\\
&= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{W}_d\mathbf{y}_d - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{W}_d\mathbf{X}_d\right)\mathbf{C}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{W}_d\mathbf{X}_d\right)'\\
&= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{y}_{dt} - \left(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\right)\mathbf{C}\left(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\right)'\\
\mathbf{L}_1 &= \mathbf{Z}_1'\mathbf{M}_1\mathbf{Z}_1 = \underset{1\le d\le D}{\text{diag}}(\mathbf{1}_{n_d}'\mathbf{W}_d\mathbf{1}_{n_d}) - \underset{1\le d\le D}{\text{col}}(\mathbf{1}_{n_d}'\mathbf{W}_d\mathbf{X}_d)\mathbf{C}\underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{W}_d\mathbf{1}_{n_d})\\
\text{tr}(\mathbf{L}_1) &= w_{...} - \sum_{d=1}^{D}\text{tr}\left\{\mathbf{1}_{n_d}'\mathbf{W}_d\mathbf{X}_d\mathbf{C}\mathbf{X}_d'\mathbf{W}_d\mathbf{1}_{n_d}\right\},\\
&= w_{...} - \sum_{d=1}^{D}\text{tr}\left\{\left(\sum_{t=1}^{m_d}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt}\right)\mathbf{C}\left(\sum_{t=1}^{m_d}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt}\right)'\right\},\quad w_{...} = \sum_{d=1}^{D}\sum_{t=1}^{m_d}\sum_{j=1}^{n_{dt}}w_{dtj}.
\end{aligned}$$

where $\mathbf{w}_{n_d}' = \mathbf{1}_{n_d}'\mathbf{W}_d$ and $\mathbf{w}_{n_{dt}}' = \mathbf{1}_{n_{dt}}'\mathbf{W}_{dt}$.

Calculations for $\mathbf{M}_2$

Let us define $\mathbf{G}_1 = (\mathbf{Z}_1'\mathbf{W}\mathbf{Z}_1)^{-1} = \underset{1\le d\le D}{\text{diag}}(w_{d..}^{-1})$, $\mathbf{P}_1 = \mathbf{W} - \mathbf{W}\mathbf{Z}_1\mathbf{G}_1\mathbf{Z}_1'\mathbf{W}$ and

$$\mathbf{B} = \left(\mathbf{X}^{(2)t}\mathbf{W}\mathbf{X}^{(2)}\right)^{-1} = \begin{pmatrix}\mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z}_1\\ \mathbf{Z}_1'\mathbf{W}\mathbf{X} & \mathbf{Z}_1'\mathbf{W}\mathbf{Z}_1\end{pmatrix}^{-1} = \begin{pmatrix}\mathbf{B}^{11} & \mathbf{B}^{12}\\ \mathbf{B}^{21} & \mathbf{B}^{22}\end{pmatrix},$$

where $w_{d..} = \sum_{t=1}^{m_d}\sum_{j=1}^{n_{dt}}w_{dtj}$. Then

$$\begin{aligned}
\mathbf{B}^{11} &= (\mathbf{X}'\mathbf{P}_1\mathbf{X})^{-1},\\
\mathbf{B}^{12} &= -\mathbf{B}^{11}\mathbf{X}'\mathbf{W}\mathbf{Z}_1\mathbf{G}_1 = -\mathbf{B}^{11}\underset{1\le d\le D}{\text{col}'}(w_{d..}^{-1}\mathbf{X}_d'\mathbf{W}_d\mathbf{1}_{n_d}),\quad \mathbf{B}^{21} = \mathbf{B}^{12\prime},\\
\mathbf{B}^{22} &= \mathbf{G}_1 + \mathbf{G}_1\mathbf{Z}_1'\mathbf{W}\mathbf{X}\mathbf{B}^{11}\mathbf{X}'\mathbf{W}\mathbf{Z}_1\mathbf{G}_1\\
&= \underset{1\le d\le D}{\text{diag}}(w_{d..}^{-1}) + \underset{1\le d\le D}{\text{col}}(w_{d..}^{-1}\mathbf{w}_{n_d}'\mathbf{X}_d)\mathbf{B}^{11}\underset{1\le d\le D}{\text{col}'}(w_{d..}^{-1}\mathbf{X}_d'\mathbf{w}_{n_d})\\
&= \underset{1\le d\le D}{\text{diag}}(w_{d..}^{-1}) + \underset{1\le d\le D}{\text{col}}(w_{d..}^{-1}\sum_{t=1}^{m_d}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt})\mathbf{B}^{11}\left[\underset{1\le d\le D}{\text{col}}(w_{d..}^{-1}\sum_{t=1}^{m_d}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt})\right]',
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{X}'\mathbf{P}_1\mathbf{X} &= \mathbf{X}'\mathbf{W}\mathbf{X} - \mathbf{X}'\mathbf{W}\mathbf{Z}_1\mathbf{G}_1\mathbf{Z}_1'\mathbf{W}\mathbf{X} = \sum_{d=1}^{D}\mathbf{X}_d'\mathbf{W}_d\mathbf{X}_d - \sum_{d=1}^{D}\mathbf{X}_d'\mathbf{W}_d\mathbf{1}_{n_d}w_{d..}^{-1}\mathbf{1}_{n_d}'\mathbf{W}_d\mathbf{X}_d\\
&= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{X}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt} - \sum_{d=1}^{D}w_{d..}^{-1}\left(\sum_{t=1}^{m_d}\mathbf{X}_{dt}'\mathbf{W}_{dt}\mathbf{1}_{n_{dt}}\right)\left(\sum_{t=1}^{m_d}\mathbf{X}_{dt}'\mathbf{W}_{dt}\mathbf{1}_{n_{dt}}\right)'.
\end{aligned}$$

The quadratic form is

$$\begin{aligned}
\mathbf{y}'\mathbf{M}_2\mathbf{y} &= \mathbf{y}'\mathbf{W}\mathbf{y} - \mathbf{y}'\mathbf{W}[\mathbf{X},\mathbf{Z}_1]\mathbf{B}[\mathbf{X}',\mathbf{Z}_1']'\mathbf{W}\mathbf{y} \\
&= \mathbf{y}'\mathbf{W}\mathbf{y} - \left[\mathbf{y}'\mathbf{W}\mathbf{X}\mathbf{B}^{11}\mathbf{X}'\mathbf{W}\mathbf{y} + \mathbf{y}'\mathbf{W}\mathbf{Z}_1\mathbf{B}^{22}\mathbf{Z}_1'\mathbf{W}\mathbf{y} + 2\mathbf{y}'\mathbf{W}\mathbf{X}\mathbf{B}^{12}\mathbf{Z}_1'\mathbf{W}\mathbf{y}\right] \\
&= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{y}_{dt} - \left(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\right)\mathbf{B}^{11}\left(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\right)' \\
&- \left[\underset{1\le d\le D}{\mathrm{col}'}\left(\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}}\right)\right]\mathbf{B}^{22}\left[\underset{1\le d\le D}{\mathrm{col}'}\left(\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}}\right)\right]' \\
&- 2\left(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\right)\mathbf{B}^{12}\left[\underset{1\le d\le D}{\mathrm{col}'}\left(\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}}\right)\right]',
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}_2 &= \widetilde{\mathbf{Z}}_2'\mathbf{M}_2\widetilde{\mathbf{Z}}_2 = \widetilde{\mathbf{Z}}_2'\mathbf{W}\widetilde{\mathbf{Z}}_2 - \widetilde{\mathbf{Z}}_2'\mathbf{W}[\mathbf{X},\mathbf{Z}_1]\mathbf{B}[\mathbf{X}',\mathbf{Z}_1']'\mathbf{W}\widetilde{\mathbf{Z}}_2 \\
&= \widetilde{\mathbf{Z}}_2'\mathbf{W}\widetilde{\mathbf{Z}}_2 - \widetilde{\mathbf{Z}}_2'\mathbf{W}\left[\mathbf{X}\mathbf{B}^{11}\mathbf{X}' - \mathbf{Z}_1\mathbf{B}^{22}\mathbf{Z}_1' - \mathbf{X}\mathbf{B}^{12}\mathbf{Z}_1' - \mathbf{Z}_1\mathbf{B}^{21}\mathbf{X}'\right]\mathbf{W}\widetilde{\mathbf{Z}}_2 \\
&= (\widetilde{\mathbf{Z}}_2'\mathbf{W}\widetilde{\mathbf{Z}}_2) - (\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X})\mathbf{B}^{11}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X})' - (\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1)\mathbf{B}^{22}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1)' \\
&- (\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X})\mathbf{B}^{12}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1)' - (\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1)\mathbf{B}^{21}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X})',
\end{aligned}$$

To calculate the trace of $\mathbf{L}_2$ we need some previous calculations.

$$\begin{aligned}
\widetilde{\mathbf{Z}}_2'\mathbf{W}\widetilde{\mathbf{Z}}_2 &= \underset{1\le d\le D}{\mathrm{diag}}\left\{\left[\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}'\right]\begin{pmatrix}\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{W}_{dt}) & \mathbf{0}\\ \mathbf{0} & \mathbf{W}_{dm_d}\end{pmatrix}\begin{bmatrix}\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{1}_{n_{dt}})\\ \mathbf{0}\end{bmatrix}\right\} \\
&= \underset{1\le d\le D}{\mathrm{diag}}\Big(\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}'\mathbf{W}_{dt}\mathbf{1}_{n_{dt}})\Big) = \underset{1\le d\le D}{\mathrm{diag}}\Big(\underset{1\le t\le m_d-1}{\mathrm{diag}}(w_{dt.})\Big), \\
\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1 &= \underset{1\le d\le D}{\mathrm{diag}}\left\{\left[\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}'\right]\begin{pmatrix}\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{W}_{dt}) & \mathbf{0}\\ \mathbf{0} & \mathbf{W}_{dm_d}\end{pmatrix}\begin{bmatrix}\underset{1\le t\le m_d-1}{\mathrm{col}}(\mathbf{1}_{n_{dt}})\\ \mathbf{1}_{n_{dm_d}}\end{bmatrix}\right\} \\
&= \underset{1\le d\le D}{\mathrm{diag}}\Big(\underset{1\le t\le m_d-1}{\mathrm{col}}(\mathbf{1}_{n_{dt}}'\mathbf{W}_{dt}\mathbf{1}_{n_{dt}})\Big) = \underset{1\le d\le D}{\mathrm{diag}}\Big(\underset{1\le t\le m_d-1}{\mathrm{col}}(w_{dt.})\Big), \\
\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X} &= \underset{1\le d\le D}{\mathrm{col}}\left\{\left[\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}'\right]\begin{pmatrix}\underset{1\le t\le m_d-1}{\mathrm{diag}}(\mathbf{W}_{dt}) & \mathbf{0}\\ \mathbf{0} & \mathbf{W}_{dm_d}\end{pmatrix}\begin{bmatrix}\underset{1\le t\le m_d-1}{\mathrm{col}}(\mathbf{X}_{dt})\\ \mathbf{X}_{dm_d}\end{bmatrix}\right\} \\
&= \underset{1\le d\le D}{\mathrm{col}}\Big(\underset{1\le t\le m_d-1}{\mathrm{col}}(\mathbf{1}_{n_{dt}}'\mathbf{W}_{dt}\mathbf{X}_{dt})\Big).
\end{aligned}$$

Finally, the trace of $\mathbf{L}_2$is

$$\begin{aligned}
\mathrm{tr}(\mathbf{L}_2) &= \mathrm{tr}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\widetilde{\mathbf{Z}}_2) - \mathrm{tr}((\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X})\mathbf{B}^{11}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X})') - \mathrm{tr}((\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1)\mathbf{B}^{22}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1)') \\
&- 2\mathrm{tr}((\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X})\mathbf{B}^{12}(\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{Z}_1)') = \left(w_{...} - \sum_{d=1}^{D} w_{dm_d.}\right) - t_{11} - t_{22} - 2t_{12},
\end{aligned}$$

where

$$
\begin{aligned}
t_{11} &= \sum_{d=1}^{D} \text{tr}\left\{ \underset{1\le t\le m_d-1}{\text{col}} (\mathbf{1}'_{n_{dt}} \mathbf{W}_{dt}\mathbf{X}_{dt})\mathbf{B}^{11} \underset{1\le t\le m_d-1}{\text{col}'} (\mathbf{X}'_{dt}\mathbf{W}_{dt}\mathbf{1}_{n_{dt}}) \right\} \\
&= \sum_{d=1}^{D}\sum_{t=1}^{m_d-1} \text{tr}\left\{\mathbf{1}'_{n_{dt}}\mathbf{W}_{dt}\mathbf{X}_{dt}\mathbf{B}^{11}\mathbf{X}'_{dt}\mathbf{W}_{dt}\mathbf{1}_{n_{dt}}\right\} = \sum_{d=1}^{D}\sum_{t=1}^{m_d-1} \mathbf{w}'_{n_{dt}}\mathbf{X}_{dt}\mathbf{B}^{11}\mathbf{X}'_{dt}\mathbf{w}_{n_{dt}}, \\
t_{22} &= \text{tr}\left\{ \underset{1\le d\le D}{\text{diag}} \big(\underset{1\le t\le m_d-1}{\text{col}} (w_{dt.})\big) \left[\underset{1\le d\le D}{\text{diag}} (w_{d..}^{-1}) + \underset{1\le d\le D}{\text{col}} (w_{d..}^{-1}\mathbf{w}'_{n_d}\mathbf{X}_d)\mathbf{B}^{11} \underset{1\le d\le D}{\text{col}'} (w_{d..}^{-1}\mathbf{X}'_d\mathbf{w}_{n_d}) \right] \right. \\
&\cdot \left. \underset{1\le d\le D}{\text{diag}} \big(\underset{1\le t\le m_d-1}{\text{col}'} (w_{dt.})\big) \right\} = \sum_{d=1}^{D} w_{d..}^{-1} \text{tr}\left\{ \underset{1\le t\le m_d-1}{\text{col}} (w_{dt.}) \underset{1\le t\le m_d-1}{\text{col}'} (w_{dt.}) \right\} \\
&+ \text{tr}\left\{ \underset{1\le d\le D}{\text{col}} (w_{d..}^{-1} \underset{1\le t\le m_d-1}{\text{col}} (w_{dt.})\mathbf{w}'_{n_d}\mathbf{X}_d)\mathbf{B}^{11} \underset{1\le d\le D}{\text{col}'} (w_{d..}^{-1}\mathbf{X}'_d\mathbf{w}_{n_d} \underset{1\le t\le m_d-1}{\text{col}'} (w_{dt.})) \right\} \\
&= \sum_{d=1}^{D} w_{d..}^{-1} \sum_{t=1}^{m_d-1} w_{dt.}^2 + \sum_{d=1}^{D} w_{d..}^{-2} \text{tr}\left\{ \underset{1\le t\le m_d-1}{\text{col}} (w_{dt.})\mathbf{w}'_{n_d}\mathbf{X}_d\mathbf{B}^{11}\mathbf{X}'_d\mathbf{w}_{n_d} \underset{1\le t\le m_d-1}{\text{col}'} (w_{dt.}) \right\} \\
&= \sum_{d=1}^{D} w_{d..}^{-1} \sum_{t=1}^{m_d-1} w_{dt.}^2 + \sum_{d=1}^{D} w_{d..}^{-2} (\mathbf{w}'_{n_d}\mathbf{X}_d\mathbf{B}^{11}\mathbf{X}'_d\mathbf{w}_{n_d}) \sum_{t=1}^{m_d-1} w_{dt.}^2 \\
&= \sum_{d=1}^{D} w_{d..}^{-1} \left\{ \left[1 + w_{d..}^{-1} \big(\sum_{t=1}^{m_d} \mathbf{w}'_{n_{dt}}\mathbf{X}_{dt}\big)\mathbf{B}^{11}\big(\sum_{t=1}^{m_d} \mathbf{w}'_{n_{dt}}\mathbf{X}_{dt}\big)' \right] \sum_{t=1}^{m_d-1} w_{dt.}^2 \right\}, \\
t_{12} &= -\text{tr}\left\{ \underset{1\le d\le D}{\text{col}} \big(\underset{1\le t\le m_d-1}{\text{col}} (\mathbf{1}'_{n_{dt}}\mathbf{W}_{dt}\mathbf{X}_{dt})\big)\mathbf{B}^{11} \underset{1\le d\le D}{\text{col}'} (w_{d..}^{-1}\mathbf{X}'_d\mathbf{W}_d\mathbf{1}_{n_d}) \underset{1\le d\le D}{\text{diag}} \big(\underset{1\le t\le m_d-1}{\text{col}'} (w_{dt.})\big) \right\} \\
&= -\text{tr}\left\{ \underset{1\le d\le D}{\text{col}} \big(\underset{1\le t\le m_d-1}{\text{col}} (\mathbf{1}'_{n_{dt}}\mathbf{W}_{dt}\mathbf{X}_{dt})\big)\mathbf{B}^{11} \underset{1\le d\le D}{\text{col}'} (w_{d..}^{-1}\mathbf{X}'_d\mathbf{W}_d\mathbf{1}_{n_d} \underset{1\le t\le m_d-1}{\text{col}'} (w_{dt.})) \right\} \\
&= -\sum_{d=1}^{D} w_{d..}^{-1} \text{tr}\left\{ \underset{1\le t\le m_d-1}{\text{col}} (\mathbf{1}'_{n_{dt}}\mathbf{W}_{dt}\mathbf{X}_{dt})\mathbf{B}^{11}\mathbf{X}'_d\mathbf{W}_d\mathbf{1}_{n_d} \underset{1\le t\le m_d-1}{\text{col}'} (w_{dt.}) \right\} \\
&= -\sum_{d=1}^{D} w_{d..}^{-1} \sum_{t=1}^{m_d-1} \mathbf{1}'_{n_{dt}}\mathbf{W}_{dt}\mathbf{X}_{dt}\mathbf{B}^{11}\mathbf{X}'_d\mathbf{W}_d\mathbf{1}_{n_d} w_{dt.} \\
&= -\sum_{d=1}^{D} w_{d..}^{-1} \left(\sum_{t=1}^{m_d-1} w_{dt.}\mathbf{1}'_{n_{dt}}\mathbf{W}_{dt}\mathbf{X}_{dt} \right) \mathbf{B}^{11} \sum_{t=1}^{m_d} \mathbf{X}'_{dt}\mathbf{w}_{n_{dt}}.
\end{aligned}
$$

Calculations for $\mathbf{M}_3$

The target of this section is to obtain a computationally efficient formula for $\mathbf{y}'\mathbf{M}_3\mathbf{y}$, where $\mathbf{M}_3 = \mathbf{W} - \mathbf{W}\mathbf{X}^{(3)}(\mathbf{X}^{(3)t}\mathbf{W}\mathbf{X}^{(3)})^{-1}\mathbf{X}^{(3)t}\mathbf{W}$, $\mathbf{X}^{(3)} = (\mathbf{X}, \mathbf{Z}_1, \widetilde{\mathbf{Z}}_2)$ and $\mathbf{Z} = (\mathbf{Z}_1, \widetilde{\mathbf{Z}}_2)$. We start by calculating the inverse of $\mathbf{X}^{(3)t}\mathbf{W}\mathbf{X}^{(3)}$. We have

$$
\mathbf{A} = (\mathbf{X}^{(3)t}\mathbf{W}\mathbf{X}^{(3)})^{-1} = \begin{pmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} \\ \mathbf{Z}'\mathbf{W}\mathbf{X} & \mathbf{Z}'\mathbf{W}\mathbf{Z} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{D}^{11} & \mathbf{D}^{12} \\ \mathbf{D}^{21} & \mathbf{D}^{22} \end{pmatrix},
$$

$\mathbf{D}^{11}=(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}$, $\mathbf{D}^{12}=-\mathbf{D}^{11}\mathbf{X}'\mathbf{W}\mathbf{Z}\mathbf{G}$, $\mathbf{D}^{21}=(\mathbf{D}^{12})'$, $\mathbf{D}^{22}=\mathbf{G}+\mathbf{G}\mathbf{Z}'\mathbf{W}\mathbf{X}\mathbf{D}^{11}\mathbf{X}'\mathbf{W}\mathbf{Z}\mathbf{G}$, $\mathbf{G}=(\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1}$ and $\mathbf{P}=\mathbf{W}-\mathbf{W}\mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{W}$.

The elements of matrix $\mathbf{G}^{-1}=\mathbf{Z}'\mathbf{W}\mathbf{Z}$ are

$$\begin{aligned}
\mathbf{G}^{11} &= \mathbf{Z}_1'\mathbf{W}\mathbf{Z}_1=\underset{1\le d\le D}{\text{diag}}\,(\mathbf{1}_{n_d}')\,\underset{1\le d\le D}{\text{diag}}\,(\mathbf{W}_d)\,\underset{1\le d\le D}{\text{diag}}\,(\mathbf{1}_{n_d})=\underset{1\le d\le D}{\text{diag}}\,(\mathbf{1}_{n_d}'\mathbf{W}_d\mathbf{1}_{n_d})=\underset{1\le d\le D}{\text{diag}}\,(w_{d..}),\\
\mathbf{G}^{12} &= \mathbf{Z}_1'\mathbf{W}\widetilde{\mathbf{Z}}_2=\underset{1\le d\le D}{\text{diag}}\left\{\underset{1\le t\le m_d-1}{\text{col}'}(w_{dt.})\right\},\quad \mathbf{G}^{21}=(\mathbf{G}^{12})',\\
\mathbf{G}^{22} &= \widetilde{\mathbf{Z}}_2'\mathbf{W}\widetilde{\mathbf{Z}}_2=\underset{1\le d\le D}{\text{diag}}\left\{\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.})\right\}.
\end{aligned}$$

The elements of matrix $\mathbf{G}$ are

$$\begin{aligned}
\mathbf{G}_{11} &= [\mathbf{G}^{11}-\mathbf{G}^{12}(\mathbf{G}^{22})^{-1}\mathbf{G}^{21}]^{-1}\\
&= \left[\underset{1\le d\le D}{\text{diag}}\,(w_{d..})-\underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{col}'}(w_{dt.})\Big)\,\underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.}^{-1})\Big)\,\underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{col}}(w_{dt.})\Big)\right]^{-1}\\
&= \left[\underset{1\le d\le D}{\text{diag}}\Big(w_{d..}-\sum_{t=1}^{m_d-1}w_{dt.}\Big)\right]^{-1}=\underset{1\le d\le D}{\text{diag}}\,(w_{dm_d.}^{-1}),\\
\mathbf{G}_{12} &= -\mathbf{G}_{11}\mathbf{G}^{12}(\mathbf{G}^{22})^{-1}=-\underset{1\le d\le D}{\text{diag}}\,(w_{dm_d.}^{-1})\,\underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{col}'}(w_{dt.})\Big)\,\underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.}^{-1})\Big)\\
&= -\underset{1\le d\le D}{\text{diag}}\Big(w_{dm_d.}^{-1}\underset{1\le t\le m_d-1}{\text{col}'}(w_{dt.})\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.}^{-1})\Big)=-\underset{1\le d\le D}{\text{diag}}\,(w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}'),\\
\mathbf{G}_{22} &= (\mathbf{G}^{22})^{-1}+(\mathbf{G}^{22})^{-1}\mathbf{G}^{21}\mathbf{G}_{11}\mathbf{G}^{12}(\mathbf{G}^{22})^{-1}=\underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.}^{-1})\Big)\\
&+ \underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.}^{-1})\underset{1\le t\le m_d-1}{\text{col}}(w_{dt.})\,w_{dm_d.}^{-1}\underset{1\le t\le m_d-1}{\text{col}'}(w_{dt.})\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.}^{-1})\Big)\\
&= \underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{diag}}\,(w_{dt.}^{-1})\Big)+\underset{1\le d\le D}{\text{diag}}\Big(w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}\mathbf{1}_{m_d-1}'\Big).
\end{aligned}$$

To obtain a computationally efficient formula for $\mathbf{P}=\mathbf{W}-\mathbf{W}\mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{W}$, we need some previous calculations.

$$\begin{aligned}
\mathbf{W}\mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{W} &= \underset{1\le d\le D}{\text{diag}}\,(\mathbf{W}_d)[\mathbf{Z}_1,\widetilde{\mathbf{Z}}_2]\begin{pmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{pmatrix}\begin{bmatrix}\mathbf{Z}_1'\\ \widetilde{\mathbf{Z}}_2'\end{bmatrix}\underset{1\le d\le D}{\text{diag}}\,(\mathbf{W}_d)\\
&= \underset{1\le d\le D}{\text{diag}}\,(\mathbf{W}_d)\big[\mathbf{Z}_1\mathbf{G}_{11}\mathbf{Z}_1'+\mathbf{Z}_1\mathbf{G}_{12}\widetilde{\mathbf{Z}}_2'+\widetilde{\mathbf{Z}}_2\mathbf{G}_{21}\mathbf{Z}_1'+\widetilde{\mathbf{Z}}_2\mathbf{G}_{22}\widetilde{\mathbf{Z}}_2'\big]\underset{1\le d\le D}{\text{diag}}\,(\mathbf{W}_d)\\
&= \mathbf{Z}_{11}+\mathbf{Z}_{12}+\mathbf{Z}_{21}+\mathbf{Z}_{22}.
\end{aligned}$$

We have

$$
\begin{aligned}
\mathbf{Z}_{11} &= \underset{1\le d\le D}{\operatorname{diag}}\,(\mathbf{W}_d\mathbf{1}_{n_d}w_{dm_d.}^{-1}\mathbf{1}_{n_d}'\mathbf{W}_d) = \underset{1\le d\le D}{\operatorname{diag}}\,(w_{dm_d.}^{-1}\mathbf{w}_{n_d}\mathbf{w}_{n_d}'),\\
\mathbf{Z}_{12} &= -\underset{1\le d\le D}{\operatorname{diag}}\Big(\mathbf{W}_d\mathbf{1}_{n_d}w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}'\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}\big]\mathbf{W}_d\Big)\\
&= -\underset{1\le d\le D}{\operatorname{diag}}\Big(w_{dm_d.}^{-1}\mathbf{w}_{n_d}\mathbf{1}_{m_d-1}'\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\\
&= -\underset{1\le d\le D}{\operatorname{diag}}\Big(w_{dm_d.}^{-1}\mathbf{w}_{n_d}\big[\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\\
&= -\underset{1\le d\le D}{\operatorname{diag}}\Big(w_{dm_d.}^{-1}\operatorname{diag}\big[\underset{1\le t\le m_d-1}{\operatorname{col}}(\mathbf{w}_{n_{dt}})\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big),\\
\mathbf{Z}_{22} &= \underset{1\le d\le D}{\operatorname{diag}}\Big(\mathbf{W}_d\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}),\mathbf{0}\big]'\underset{1\le t\le m_d-1}{\operatorname{diag}}(w_{dt.}^{-1})\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}\big]\mathbf{W}_d\Big)\\
&+ \underset{1\le d\le D}{\operatorname{diag}}\Big(\mathbf{W}_d\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}),\mathbf{0}\big]'w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}\mathbf{1}_{m_d-1}'\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}\big]\mathbf{W}_d\Big)\\
&= \underset{1\le d\le D}{\operatorname{diag}}\Big(\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{w}_{n_{dt}}),\mathbf{0}\big]'\underset{1\le t\le m_d-1}{\operatorname{diag}}(w_{dt.}^{-1})\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\\
&+ \underset{1\le d\le D}{\operatorname{diag}}\Big(w_{dm_d.}^{-1}\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{w}_{n_{dt}}),\mathbf{0}\big]'\mathbf{1}_{m_d-1}\mathbf{1}_{m_d-1}'\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\\
&= \underset{1\le d\le D}{\operatorname{diag}}\Big(\operatorname{diag}\big[\underset{1\le t\le m_d-1}{\operatorname{diag}}(w_{dt.}^{-1}\mathbf{w}_{n_{dt}}\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\\
&+ \underset{1\le d\le D}{\operatorname{diag}}\Big(w_{dm_d.}^{-1}\big[\underset{1\le t\le m_d-1}{\operatorname{col}}(\mathbf{w}_{n_{dt}}),\mathbf{0}\big]'\big[\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big).
\end{aligned}
$$

To calculate the quadratic form $\mathbf{X}'\mathbf{P}\mathbf{X}$ we make a decomposition, i.e.

$$
\mathbf{X}'\mathbf{P}\mathbf{X} = \mathbf{X}'\mathbf{W}\mathbf{X} - \mathbf{X}'\mathbf{Z}_{11}\mathbf{X} - 2\mathbf{X}'\mathbf{Z}_{12}\mathbf{X} - \mathbf{X}'\mathbf{Z}_{22}\mathbf{X},
$$

where

$$
\begin{aligned}
\mathbf{X}'\mathbf{W}\mathbf{X} &= \underset{1\le d\le D}{\operatorname{col}'}(\mathbf{X}_d')\underset{1\le d\le D}{\operatorname{diag}}(\mathbf{W}_d)\underset{1\le d\le D}{\operatorname{col}}(\mathbf{X}_d) = \sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{X}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt},\\
\mathbf{X}'\mathbf{Z}_{11}\mathbf{X} &= \underset{1\le d\le D}{\operatorname{col}'}(\mathbf{X}_d')\underset{1\le d\le D}{\operatorname{diag}}(w_{dm_d.}^{-1}\mathbf{w}_{n_d}\mathbf{w}_{n_d}')\underset{1\le d\le D}{\operatorname{col}}(\mathbf{X}_d)\\
&= \sum_{d=1}^{D} w_{dm_d.}^{-1}\Big(\sum_{t=1}^{m_d}\mathbf{X}_{dt}'\mathbf{w}_{n_{dt}}\Big)\Big(\sum_{t=1}^{m_d}\mathbf{X}_{dt}'\mathbf{w}_{n_{dt}}\Big)',\\
\mathbf{X}'\mathbf{Z}_{12}\mathbf{X} &= -\underset{1\le d\le D}{\operatorname{col}'}(\mathbf{X}_d')\underset{1\le d\le D}{\operatorname{diag}}\Big(w_{dm_d.}^{-1}\operatorname{diag}\big[\underset{1\le t\le m_d-1}{\operatorname{col}}(\mathbf{w}_{n_{dt}})\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\underset{1\le d\le D}{\operatorname{col}}(\mathbf{X}_d)\\
&= -\sum_{d=1}^{D} w_{dm_d.}^{-1}\Big(\sum_{t=1}^{m_d-1}\mathbf{X}_{dt}'\mathbf{w}_{n_{dt}}\Big)\Big(\sum_{t=1}^{m_d-1}\mathbf{X}_{dt}'\mathbf{w}_{n_{dt}}\Big)',
\end{aligned}
$$

$$\begin{aligned}
\mathbf{X}'\mathbf{Z}_{22}\mathbf{X} &= \underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d')\underset{1\le d\le D}{\text{diag}}\Big(\text{diag}\big[\underset{1\le t\le m_d-1}{\text{diag}}(w_{dt.}^{-1}\mathbf{w}_{n_{dt}}\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)\\
&+ \underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d')\underset{1\le d\le D}{\text{diag}}\Big(w_{dm_d.}^{-1}\big[\underset{1\le t\le m_d-1}{\text{col}}(\mathbf{w}_{n_{dt}}),\mathbf{0}\big]'\big[\underset{1\le t\le m_d-1}{\text{col}'}(\mathbf{w}_{n_{dt}}'),\mathbf{0}\big]\Big)\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)\\
&= \sum_{d=1}^{D}\sum_{t=1}^{m_d-1} w_{dt.}^{-1}\mathbf{X}_{dt}'\mathbf{w}_{n_{dt}}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt}+\sum_{d=1}^{D} w_{dm_d.}^{-1}\Big(\sum_{t=1}^{m_d-1}\mathbf{X}_{dt}'\mathbf{w}_{n_{dt}}\Big)\Big(\sum_{t=1}^{m_d-1}\mathbf{X}_{dt}'\mathbf{w}_{n_{dt}}\Big)'.
\end{aligned}$$

To obtain a computationally efficient formula for $\mathbf{G}\mathbf{Z}'\mathbf{W}\mathbf{X}$, we do some previous calculations.

$$\mathbf{G}\mathbf{Z}'\mathbf{W}\mathbf{X}=\begin{pmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_1'\\ \widetilde{\mathbf{Z}}_2'\end{pmatrix}\mathbf{W}\mathbf{X}=\begin{pmatrix}\mathbf{R}_{11}+\mathbf{R}_{12}\\ \mathbf{R}_{21}+\mathbf{R}_{22}\end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{R}_{11} &= \mathbf{G}_{11}\mathbf{Z}_1'\mathbf{W}\mathbf{X}=\underset{1\le d\le D}{\text{diag}}(w_{dm_d.}^{-1})\underset{1\le d\le D}{\text{diag}}(\mathbf{1}_{n_d}')\underset{1\le d\le D}{\text{col}}(\mathbf{W}_d\mathbf{X}_d)=\underset{1\le d\le D}{\text{col}}\Big(w_{dm_d.}^{-1}\sum_{t=1}^{m_d}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt}\Big),\\
\mathbf{R}_{12} &= \mathbf{G}_{12}\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X}=-\underset{1\le d\le D}{\text{diag}}(w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}')\underset{1\le d\le D}{\text{diag}}\Big(\big[\underset{1\le t\le m_d-1}{\text{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}\big]\Big)\underset{1\le d\le D}{\text{col}}(\mathbf{W}_d\mathbf{X}_d)\\
&= -\underset{1\le d\le D}{\text{col}}\Big(w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}'\underset{1\le t\le m_d-1}{\text{col}}(\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt})\Big)=-\underset{1\le d\le D}{\text{col}}\Big(w_{dm_d.}^{-1}\sum_{t=1}^{m_d-1}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt}\Big),\\
\mathbf{R}_{21} &= \mathbf{G}_{21}\mathbf{Z}_1'\mathbf{W}\mathbf{X}=-\underset{1\le d\le D}{\text{diag}}(w_{dm_d.}^{-1}\mathbf{1}_{m_d-1})\underset{1\le d\le D}{\text{diag}}(\mathbf{1}_{n_d}')\underset{1\le d\le D}{\text{col}}(\mathbf{W}_d\mathbf{X}_d)\\
&= -\underset{1\le d\le D}{\text{col}}(w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}\mathbf{w}_{n_d}'\mathbf{X}_d)=-\underset{1\le d\le D}{\text{col}}\Big(\underset{1\le t\le m_d-1}{\text{col}}\Big(w_{dm_d.}^{-1}\sum_{t=1}^{m_d}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt}\Big)\Big),\\
\mathbf{R}_{22} &= \mathbf{G}_{22}\widetilde{\mathbf{Z}}_2'\mathbf{W}\mathbf{X}=\left[\underset{1\le d\le D}{\text{diag}}\Big(\underset{1\le t\le m_d-1}{\text{diag}}(w_{dt.}^{-1})\Big)+\underset{1\le d\le D}{\text{diag}}\Big(w_{dm_d.}^{-1}\mathbf{1}_{m_d-1}\mathbf{1}_{m_d-1}'\Big)\right]\\
&\cdot\ \underset{1\le d\le D}{\text{diag}}\Big(\big[\underset{1\le t\le m_d-1}{\text{diag}}(\mathbf{1}_{n_{dt}}'),\mathbf{0}\big]\Big)\underset{1\le d\le D}{\text{col}}(\mathbf{W}_d\mathbf{X}_d)\\
&= \underset{1\le d\le D}{\text{col}}\Big(\underset{1\le t\le m_d-1}{\text{col}}(w_{dt.}^{-1}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt})\Big)+\underset{1\le d\le D}{\text{col}}\Big(\underset{1\le t\le m_d-1}{\text{col}}\Big(w_{dm_d.}^{-1}\sum_{t=1}^{m_d-1}\mathbf{w}_{n_{dt}}'\mathbf{X}_{dt}\Big)\Big).
\end{aligned}$$

The calculation of matrix

$$\mathbf{A}=(\mathbf{X}^{(3)t}\mathbf{W}\mathbf{X}^{(3)})^{-1}=\begin{pmatrix}\mathbf{D}^{11} & \mathbf{D}^{12}\\ \mathbf{D}^{21} & \mathbf{D}^{22}\end{pmatrix}=\begin{pmatrix}\mathbf{D}^{11} & \mathbf{A}^{12} & \mathbf{A}^{13}\\ \mathbf{A}^{21} & \mathbf{A}^{22} & \mathbf{A}^{23}\\ \mathbf{A}^{31} & \mathbf{A}^{32} & \mathbf{A}^{33}\end{pmatrix}$$

is given below.

$$\mathbf{D}^{11} = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}=(\mathbf{X}'\mathbf{W}\mathbf{X}-\mathbf{X}'\mathbf{Z}_{11}\mathbf{X}-2\mathbf{X}'\mathbf{Z}_{12}\mathbf{X}-\mathbf{X}'\mathbf{Z}_{22}\mathbf{X})^{-1},$$

$$\begin{aligned}
\mathbf{D}^{21} &= -\mathbf{G}\mathbf{Z}'\mathbf{W}\mathbf{X}\mathbf{D}^{11}=[\mathbf{A}^{21},\mathbf{A}^{31}]',\quad \mathbf{D}^{12}=(\mathbf{D}^{21})'=[(\mathbf{A}^{21})',(\mathbf{A}^{13})']=[\mathbf{A}^{12},\mathbf{A}^{13}],\\
\mathbf{A}^{21} &= -[\mathbf{G}_{11}\mathbf{Z}_1'+\mathbf{G}_{12}\widetilde{\mathbf{Z}}_2']\mathbf{W}\mathbf{X}\mathbf{D}^{11}=-[\mathbf{R}_{11}+\mathbf{R}_{12}]\mathbf{D}^{11},\\
\mathbf{A}^{31} &= -[\mathbf{G}_{21}\mathbf{Z}_1'+\mathbf{G}_{22}\widetilde{\mathbf{Z}}_2']\mathbf{W}\mathbf{X}\mathbf{D}^{11}=-[\mathbf{R}_{21}+\mathbf{R}_{22}]\mathbf{D}^{11},
\end{aligned}$$

The matrix $\mathbf{D}^{22}$ admits the decomposition

$$\mathbf{D}^{22}=\mathbf{G}+\mathbf{G}\mathbf{Z}'\mathbf{W}\mathbf{X}\mathbf{D}^{11}\mathbf{X}'\mathbf{W}\mathbf{Z}\mathbf{G}=\begin{pmatrix}\mathbf{A}^{22} & \mathbf{A}^{23}\\ \mathbf{A}^{31} & \mathbf{A}^{33}\end{pmatrix}$$

where

$$\begin{aligned}\mathbf{A}^{22} &= \mathbf{G}_{11}+[\mathbf{R}_{11}+\mathbf{R}_{12}]\mathbf{D}^{11}[\mathbf{R}_{11}+\mathbf{R}_{12}]', \quad \mathbf{A}^{23}=\mathbf{G}_{12}+[\mathbf{R}_{11}+\mathbf{R}_{12}]\mathbf{D}^{11}[\mathbf{R}_{21}+\mathbf{R}_{22}]',\\ \mathbf{A}^{32} &= \mathbf{G}_{21}+[\mathbf{R}_{21}+\mathbf{R}_{22}]\mathbf{D}^{11}[\mathbf{R}_{11}+\mathbf{R}_{12}]', \quad \mathbf{A}^{33}=\mathbf{G}_{22}+[\mathbf{R}_{21}+\mathbf{R}_{22}]\mathbf{D}^{11}[\mathbf{R}_{21}+\mathbf{R}_{22}]'.\end{aligned}$$

Finally, we calculate the quadratic form $\mathbf{y}'\mathbf{M}_3\mathbf{y}$.

$$\begin{aligned}\mathbf{y}'\mathbf{M}_3\mathbf{y} &= \mathbf{y}'\mathbf{W}\mathbf{y}-\mathbf{y}'\mathbf{W}[\mathbf{X},\mathbf{Z}_1,\widetilde{\mathbf{Z}}_2]\mathbf{A}[\mathbf{X}',\mathbf{Z}_1',\widetilde{\mathbf{Z}}_2']'\mathbf{W}\mathbf{y}=\mathbf{y}'\mathbf{W}\mathbf{y}\\ &- \mathbf{y}'\mathbf{W}[\mathbf{X}\mathbf{D}^{11}\mathbf{X}'+\mathbf{Z}_1\mathbf{A}^{22}\mathbf{Z}_1'+\widetilde{\mathbf{Z}}_2\mathbf{A}^{33}\widetilde{\mathbf{Z}}_2'+2\mathbf{X}\mathbf{A}^{12}\mathbf{Z}_1'+2\mathbf{X}\mathbf{A}^{13}\widetilde{\mathbf{Z}}_2'+2\mathbf{Z}_1\mathbf{A}^{23}\widetilde{\mathbf{Z}}_2']\mathbf{W}\mathbf{y}\\ &= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{y}_{dt}-\Bigg\{\Big(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\Big)\mathbf{D}^{11}\Big(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\Big)'\\ &+ \underset{1\le d\le D}{\operatorname{col}'}\Big(\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}}\Big)\mathbf{A}^{22}\Big[\underset{1\le d\le D}{\operatorname{col}'}\Big(\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}}\Big)\Big]'\\ &+ \underset{1\le d\le D}{\operatorname{col}'}\Big(\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}})\Big)\mathbf{A}^{33}\Big[\underset{1\le d\le D}{\operatorname{col}'}\Big(\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}})\Big)\Big]'\\ &+ 2\Big(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\Big)\mathbf{A}^{12}\Big[\underset{1\le d\le D}{\operatorname{col}'}\Big(\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}}\Big)\Big]'\\ &+ 2\Big(\sum_{d=1}^{D}\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{W}_{dt}\mathbf{X}_{dt}\Big)\mathbf{A}^{13}\Big[\underset{1\le d\le D}{\operatorname{col}'}\Big(\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}})\Big)\Big]'\\ &+ 2\underset{1\le d\le D}{\operatorname{col}'}\Big(\sum_{t=1}^{m_d}\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}}\Big)\mathbf{A}^{23}\Big[\underset{1\le d\le D}{\operatorname{col}'}\Big(\underset{1\le t\le m_d-1}{\operatorname{col}'}(\mathbf{y}_{dt}'\mathbf{w}_{n_{dt}})\Big)\Big]'\Bigg\}.\end{aligned}$$

9.2.4 The EBLUP of the domain mean

The EBLUP of the linear parameter $\eta=\mathbf{a}'\mathbf{y}=\mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\mathbf{y}_r$ is

$$\widehat{\eta}=\mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta}+\widehat{\mathbf{V}}_{rs}\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s-\mathbf{X}_s\widehat{\beta})\right]$$

As $\mathbf{V}_{ers}=\mathbf{0}$, $\mathbf{V}_{rs}=\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'+\mathbf{V}_{ers}=\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'$ and $\widehat{\mathbf{u}}=\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_{ss}^{-1}(\mathbf{y}_s-\mathbf{X}_s\widehat{\beta})$, we have

$$\begin{aligned}\widehat{\eta} &= \mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta}+\mathbf{Z}_r\widehat{\mathbf{V}}_u\mathbf{Z}_s'\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s-\mathbf{X}_s\widehat{\beta})\right]=\mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta}+\mathbf{Z}_r\widehat{\mathbf{u}}\right]\\ &= \mathbf{a}'\left[\mathbf{X}\widehat{\beta}+\mathbf{Z}_1\widehat{\mathbf{u}}_1+\mathbf{Z}_2\widehat{\mathbf{u}}_2\right]+\mathbf{a}_s'\left[\mathbf{y}_s-\mathbf{X}_s\widehat{\beta}-\mathbf{Z}_{s1}\widehat{\mathbf{u}}_1-\mathbf{Z}_{s2}\widehat{\mathbf{u}}_2\right].\end{aligned}$$

Under model (9.7), $\overline{Y}_{dt}=\frac{1}{N_{dt}}\sum_{j=1}^{N_{dt}}y_{dtj}$ can be written as a linear parameter $\eta=\mathbf{a}'\mathbf{y}$, where

$$\begin{aligned}\mathbf{a}' &= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}',\dots,\mathbf{0}_{N_{d-1}}',\mathbf{0}_{N_{d1}}',\dots,\mathbf{0}_{N_{d(i-1)}}',\mathbf{1}_{N_{dt}}',\mathbf{0}_{N_{d(i+1)}}',\dots,\mathbf{0}_{N_{dm_d}}',\mathbf{0}_{N_{d+1}}',\dots,\mathbf{0}_{N_D}')\\ &= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}',\dots,\mathbf{0}_{N_{d-1}}',\underset{1\le k\le m_d}{\operatorname{col}'}\left[\delta_{tk}\mathbf{1}_{N_{dk}}'\right],\mathbf{0}_{N_{d+1}}',\dots,\mathbf{0}_{N_D}')=\frac{1}{N_{dt}}\underset{1\le \ell\le D}{\operatorname{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\operatorname{col}'}[\delta_{tk}\mathbf{1}_{N_{\ell k}}']\},\end{aligned}$$

with $\delta_{ab}=1$ if $a=b$ and $\delta_{ab}=0$ if $a\neq b$. It hold that $\mathbf{a}'\mathbf{X}=\overline{X}_{dt}$,

$$\begin{aligned}
\mathbf{a}'\mathbf{Z}_1 &= \frac{1}{N_{dt}}\underset{1\leq\ell\leq D}{\text{col}'}\{\delta_{d\ell}\underset{1\leq k\leq m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}}]\}\underset{1\leq\ell\leq D}{\text{diag}}(\mathbf{1}_{N_\ell})=\underset{1\leq\ell\leq D}{\text{col}'}\{\delta_{d\ell}\}=\overline{\mathbf{Z}}_{1,dt},\\
\mathbf{a}'\mathbf{Z}_2 &= \frac{1}{N_{dt}}\underset{1\leq\ell\leq D}{\text{col}'}\{\delta_{d\ell}\underset{1\leq k\leq m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}}]\}\underset{1\leq\ell\leq D}{\text{diag}}(\underset{1\leq k\leq m_\ell}{\text{diag}}(\mathbf{1}_{N_{\ell k}}))=\underset{1\leq\ell\leq D}{\text{col}'}\{\underset{1\leq k\leq m_\ell}{\text{col}'}\{\delta_{d\ell}\delta_{tk}\}\}=\overline{\mathbf{Z}}_{2,dt}.
\end{aligned}$$

If $n_{dt}>0$, the EBLUP of $\overline{Y}_{dt}$ is

$$\widehat{\overline{Y}}_{dt}^{eblup}=\overline{\mathbf{X}}_{dt}\widehat{\beta}+\overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1+\overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2+f_{dt}\left[\overline{\mathbf{y}}_{s,dt}-\overline{\mathbf{X}}_{s,dt}\widehat{\beta}-\overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1-\overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2\right],$$

where $\overline{\mathbf{y}}_{s,dt}=\frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}}y_{dtj}$, $\overline{\mathbf{X}}_{s,dt}=\frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}}\mathbf{x}_{dtj}$ and $f_{dt}=\frac{n_{dt}}{N_{dt}}$.

If $n_{dt}=0$, the EBLUP of $\overline{Y}_{dt}$ is

$$\widehat{\overline{Y}}_{dt}^{eblup}=\overline{\mathbf{X}}_{dt}\widehat{\beta}+\overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1+\overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2.$$

9.2.5 Mean squared error of the EBLUP

Let $\theta=(\sigma_0^2,\varphi_1,\varphi_2)$ be the vector of variance components. The mean squared error of the EBLUP of $\overline{Y}_{dt}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup})=g_1(\theta)+g_2(\theta)+g_3(\theta)+g_4(\theta),$$

where

$$\begin{aligned}
g_1(\theta) &= \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r,\\
g_2(\theta) &= [\mathbf{a}_r'\mathbf{X}_r-\mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r-\mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r],\\
g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}_s(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},\\
g_4(\theta) &= \mathbf{a}_r'\mathbf{V}_{er}\mathbf{a}_r.
\end{aligned}$$

Calculation of $g_1(\theta)$

The elements of formula $g_1(\theta)=\mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r$ are

$$\begin{aligned}
\mathbf{a}_r' &= \frac{1}{N_{dt}}\left(\mathbf{0}'_{N_1-n_1},\dots,\mathbf{0}'_{N_{d-1}-n_{d-1}},\underset{1\leq k\leq m_d}{\text{col}'}\left[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}\right],\mathbf{0}'_{N_{d+1}-n_{d+1}},\dots,\mathbf{0}'_{N_D-n_D}\right),\\
\mathbf{Z}_r &= [\mathbf{Z}_{1r}\,\mathbf{Z}_{2r}],\quad \mathbf{T}_s=\mathbf{V}_u-\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\Sigma_u=\begin{pmatrix}\mathbf{T}_{11s} & \mathbf{T}_{12s}\\ \mathbf{T}_{21s} & \mathbf{T}_{22s}\end{pmatrix},\\
\mathbf{V}_u &= \begin{pmatrix}\sigma_1^2\mathbf{I}_D & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}_M\end{pmatrix},\quad \mathbf{Z}_s=[\mathbf{Z}_{1s}\,\mathbf{Z}_{2s}],\quad \mathbf{V}_s^{-1}=\underset{1\leq d\leq D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}.
\end{aligned}$$

It holds that

$$\begin{aligned}
\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{V}_u &= \begin{pmatrix} \sigma_1^2\mathbf{Z}_{1s}' \\ \sigma_2^2\mathbf{Z}_{2s}' \end{pmatrix} \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\left[\sigma_1^2\mathbf{Z}_{1s}, \sigma_2^2\mathbf{Z}_{2s}\right] \\
&= \begin{pmatrix} \sigma_1^4\mathbf{Z}_{1s}' \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{1s} & \sigma_1^2\sigma_2^2\mathbf{Z}_{1s}' \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s} \\ \sigma_1^2\sigma_2^2\mathbf{Z}_{2s}' \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{1s} & \sigma_2^4\mathbf{Z}_{2s}' \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s} \end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{Z}_{1s}' \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{1s} &= \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{n_d}'\} \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\} \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{n_d}\} = \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{nd}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{nd}\}, \\
\mathbf{Z}_{1s}' \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s} &= \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{n_d}'\} \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\} \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\} \\
&= \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{nd}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\}, \\
\mathbf{Z}_{2s}' \underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\}\mathbf{Z}_{2s} &= \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\}.
\end{aligned}$$

The blocks of matrix $\mathbf{T}_s$ are

$$\begin{aligned}
\mathbf{T}_{11s} &= \sigma_1^2 \underset{1\le d\le D}{\text{diag}}\{1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}\}, \\
\mathbf{T}_{12s} &= -\sigma_1^2\sigma_2^2 \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\},\ \mathbf{T}_{21s} = (\mathbf{T}_{12s})', \\
\mathbf{T}_{22s} &= \sigma_2^2 \underset{1\le d\le D}{\text{diag}}\{\mathbf{I}_{m_d} - \sigma_2^2 \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\}.
\end{aligned}$$

The product $\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'$ is

$$\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r' = [\mathbf{Z}_{1r}\,\mathbf{Z}_{2r}]\,\mathbf{T}_s\left[\mathbf{Z}_{1r}'\,\mathbf{Z}_{2r}'\right]' = \mathbf{Z}_{1r}\mathbf{T}_{11s}\mathbf{Z}_{1r}' + \mathbf{Z}_{1r}\mathbf{T}_{12s}\mathbf{Z}_{2r}' + \mathbf{Z}_{2r}\mathbf{T}_{21s}\mathbf{Z}_{1r}' + \mathbf{Z}_{2r}\mathbf{T}_{22s}\mathbf{Z}_{2r}'.$$

It holds that

$$\begin{aligned}
\mathbf{M}_{11}^{rr} &= \mathbf{Z}_{1r}\mathbf{T}_{11s}\mathbf{Z}_{1r}' = \sigma_1^2 \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{N_d-n_d}\} \underset{1\le d\le D}{\text{diag}}\{1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}\} \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{N_d-n_d}'\} \\
&= \sigma_1^2 \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{N_d-n_d}[1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}]\mathbf{1}_{N_d-n_d}'\}, \\
\mathbf{M}_{12}^{rr} &= \mathbf{Z}_{1r}\mathbf{T}_{12s}\mathbf{Z}_{2r}' = -\sigma_1^2\sigma_2^2 \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{N_d-n_d}\} \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\} \\
&\cdot \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\} \\
&= -\sigma_1^2\sigma_2^2 \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{N_d-n_d}\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}) \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\}, \\
\mathbf{M}_{21}^{rr} &= (\mathbf{M}_{12}^{rr})', \\
\mathbf{M}_{22}^{rr} &= \mathbf{Z}_{2r}\mathbf{T}_{22s}\mathbf{Z}_{2r}' = \sigma_2^2 \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\} \\
&\cdot \underset{1\le d\le D}{\text{diag}}\{\mathbf{I}_{m_d}-\sigma_2^2 \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\} \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\} \\
&= \sigma_2^2 \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}) \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\} \\
&- \sigma_2^4 \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}) \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}) \\
&\cdot \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\}.
\end{aligned}$$

As

$$\mathbf{a}_r' = \frac{1}{N_{dt}} \underset{1\le \ell\le D}{\text{col}'}\left[\delta_{d\ell} \underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{\ell k}-n_{\ell k}}']\right] \quad \text{y} \quad f_{dt} = \frac{n_{dt}}{N_{dt}},$$

we obtain

$$\begin{aligned}
\mathbf{a}_r'\mathbf{M}_{11}^{rr}\mathbf{a}_r &= \sigma_1^2\mathbf{a}_r' \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{N_d-n_d}[1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}]\mathbf{1}_{N_d-n_d}'\}\mathbf{a}_r \\
&= \sigma_1^2(1-f_{dt})^2[1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}] = \sigma^2\varphi_1(1-f_{dt})^2[1-\varphi_1\mathbf{1}_{n_d}'\Sigma_{ds}^{-1}\mathbf{1}_{n_d}], \\
\mathbf{a}_r'\mathbf{M}_{12}^{rr}\mathbf{a}_r &= -\sigma_1^2\sigma_2^2\mathbf{a}_r' \underset{1\le d\le D}{\text{diag}}\{\mathbf{1}_{N_d-n_d}\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}) \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\}\mathbf{a}_r, \\
&= -\sigma^2\varphi_1\varphi_2(1-f_{dt})\mathbf{1}_{n_d}'\Sigma_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}) \underset{1\le k\le m_d}{\text{col}}[\delta_{tk}(1-f_{dk})] \\
\mathbf{a}_r'\mathbf{M}_{22}^{rr}\mathbf{a}_r &= \sigma_2^2\mathbf{a}_r' \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}) \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\}\mathbf{a}_r \\
&- \sigma_2^4\mathbf{a}_r' \underset{1\le d\le D}{\text{diag}}\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}) \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}) \\
&\cdot \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}}')\}\mathbf{a}_r \\
&= \sigma^2\varphi_2(1-f_{dt})^2 - \sigma^2\varphi_2^2 \underset{1\le k\le m_d}{\text{col}'}[(1-f_{dk})\delta_{tk}] \\
&\cdot \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\Sigma_{ds}^{-1} \underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}) \underset{1\le k\le m_d}{\text{col}}[(1-f_{dk})\delta_{tk}].
\end{aligned}$$

Finally,

$$g_1(\theta) = \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r = \mathbf{a}_r'\mathbf{M}_{11}^{rr}\mathbf{a}_r + 2\mathbf{a}_r'\mathbf{M}_{12}^{rr}\mathbf{a}_r + \mathbf{a}_r'\mathbf{M}_{22}^{rr}\mathbf{a}_r,$$

Calculation of $g_2(\theta)$

The formula for $g_2(\theta)$ is

$$g_2(\theta) = [\mathbf{a}_r'\mathbf{X}_r - \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r - \mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r] = [\mathbf{a}_{21}' - \mathbf{a}_{22}']\mathbf{Q}_s[\mathbf{a}_{21} - \mathbf{a}_{22}],$$

where $\mathbf{Q}_s = (\mathbf{X}_s'\mathbf{V}^{-1}\mathbf{X}_s)^{-1} = \sigma^2\left(\sum_{d=1}^D \mathbf{X}_{ds}'\Sigma_{ds}^{-1}\mathbf{X}_{ds}\right)^{-1}$ and $\mathbf{V}_{es}^{-1} = \sigma^{-2}\mathbf{W}_s$. On the one hand

$$\mathbf{a}_{21}' = \mathbf{a}_r'\mathbf{X}_r = \frac{1}{N_{dt}}\mathbf{1}_{N_{dt}-n_{dt}}'\mathbf{X}_{dt,r} = \frac{1}{N_{dt}}\sum_{j\in r}\mathbf{x}_{dtj} = (1-f_{dt})\overline{\mathbf{X}}_{dt}^*, \text{ where } \overline{\mathbf{X}}_{dt}^* = \frac{1}{N_{dt}-n_{dt}}\sum_{j\in r}\mathbf{x}_{dtj}.$$

On the other hand

$$\mathbf{a}_{22}' = \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s = \sigma^{-2}\mathbf{a}_r'(\mathbf{M}_{11}^{rs} + \mathbf{M}_{12}^{rs} + \mathbf{M}_{21}^{rs} + \mathbf{M}_{22}^{rs})\mathbf{W}_s\mathbf{X}_s = \mathbf{G}_{11} + \mathbf{G}_{12} + \mathbf{G}_{21} + \mathbf{G}_{22},$$

where

$$\mathbf{M}_{11}^{rs} = \mathbf{Z}_{1r}\mathbf{T}_{11s}\mathbf{Z}_{1s}',\ \mathbf{M}_{12}^{rs} = \mathbf{Z}_{1r}\mathbf{T}_{12s}\mathbf{Z}_{2s}'$$
$$\mathbf{M}_{21}^{rs} = \mathbf{Z}_{2r}\mathbf{T}_{21s}\mathbf{Z}_{1s}' = (\mathbf{M}_{12}^{sr})',\ \mathbf{M}_{22}^{rs} = \mathbf{Z}_{2r}\mathbf{T}_{22s}\mathbf{Z}_{2s}'.$$

Let us define $\mathbf{w}_{n_{dk}}' = (w_{dk1},\ldots,w_{dkn_{dk}})$. It holds that

$$\begin{aligned}
\mathbf{G}_{11} &= \sigma^{-2}\mathbf{a}_r'\mathbf{M}_{11}^{rs}\mathbf{W}_s\mathbf{X}_s = \frac{\sigma_1^2}{\sigma^2 N_{dt}}\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{dk}-n_{dk}}']\mathbf{1}_{N_d-n_d}[1-\sigma_1^2\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}]\mathbf{1}_{n_d}'\mathbf{W}_{ds}\mathbf{X}_{ds} \\
&= \varphi_1(1-f_{dt})[1-\varphi_1\mathbf{1}_{n_d}'\Sigma_{ds}^{-1}\mathbf{1}_{n_d}]\sum_{k=1}^{m_d}\mathbf{w}_{n_{dk}}'\mathbf{X}_{dk,s}, \\
\mathbf{G}_{12} &= \sigma^{-2}\mathbf{a}_r'\mathbf{M}_{12}^{rs}\mathbf{W}_s\mathbf{X}_s \\
&= -\frac{\sigma_1^2\sigma_2^2}{\sigma^2 N_{dt}}\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{dk}-n_{dk}}']\mathbf{1}_{N_d-n_d}\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{W}_{ds}\mathbf{X}_{ds} \\
&= -\varphi_1\varphi_2(1-f_{dt})\mathbf{1}_{n_d}'\Sigma_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\underset{1\le k\le m_d}{\text{col}}(\mathbf{w}_{n_{dj}}'\mathbf{X}_{dk,s}), \\
\mathbf{G}_{21} &= \sigma^{-2}\mathbf{a}_r'\mathbf{M}_{21}^{rs}\mathbf{W}_s\mathbf{X}_s \\
&= -\frac{\sigma_1^2\sigma_2^2}{\sigma^2 N_{dt}}\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{dk}-n_{dk}}']\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\mathbf{W}_{ds}\mathbf{X}_{ds} \\
&= -\varphi_1\varphi_2(1-f_{dt})\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}]\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\Sigma_{ds}^{-1}\mathbf{1}_{n_d}\sum_{k=1}^{m_d}\mathbf{w}_{n_{dk}}'\mathbf{X}_{dk,s}, \\
\mathbf{G}_{22} &= \sigma^{-2}\mathbf{a}_r'\mathbf{M}_{22}^{rs}\mathbf{W}_s\mathbf{X}_s \\
&= \frac{\sigma_2^2}{\sigma^2 N_{dt}}\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{dk}-n_{dk}}']\{\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}') \\
&- \sigma_2^2\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\}\mathbf{W}_{ds}\mathbf{X}_{ds} \\
&= \varphi_2(1-f_{dt})\underset{1\le k\le m_d}{\text{col}'}[\delta_{tk}]\left[\mathbf{I}_{m_d} - \varphi_2\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}}')\Sigma_{ds}^{-1}\underset{1\le k\le m_d}{\text{diag}}(\mathbf{1}_{n_{dk}})\right]\underset{1\le k\le m_d}{\text{col}}(\mathbf{w}_{n_{dk}}'\mathbf{X}_{dk,s}).
\end{aligned}$$

Calculation of $g_3(\theta)$

The formula for $g_3(\theta)$ is

$$g_3(\theta) \approx \operatorname{tr}\left\{(\nabla \mathbf{b}')\mathbf{V}_s(\nabla \mathbf{b}')' E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where

$$\begin{aligned}
\mathbf{b}' &= \mathbf{a}_r'\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1} = \mathbf{a}_r'[\mathbf{Z}_{1r}, \mathbf{Z}_{2r}]\operatorname{diag}\left\{\sigma_1^2\mathbf{I}_D, \sigma_2^2\mathbf{I}_M\right\}[\mathbf{Z}_{1s}', \mathbf{Z}_{2s}']'\mathbf{V}_s^{-1} \\
&= \mathbf{a}_r'\left[\sigma_1^2\mathbf{Z}_{1r}\mathbf{Z}_{1s}' + \sigma_2^2\mathbf{Z}_{2r}\mathbf{Z}_{2s}'\right]\mathbf{V}_s^{-1} = \sigma_1^2\mathbf{a}_r'\mathbf{Z}_{1r}\mathbf{Z}_{1s}'\mathbf{V}_s^{-1} + \sigma_2^2\mathbf{a}_r'\mathbf{Z}_{2r}\mathbf{Z}_{2s}'\mathbf{V}_s^{-1} \\
&= \mathbf{b}_1' + \mathbf{b}_2' = \underset{1\le\ell\le D}{\operatorname{col}'}[\delta_{d\ell}\mathbf{b}_{1\ell}'] + \underset{1\le\ell\le D}{\operatorname{col}'}[\delta_{d\ell}\mathbf{b}_{2\ell}'], \\
\mathbf{b}_{1d}' &= \frac{\sigma^2\varphi_1}{N_{dt}}\underset{1\le k\le m_d}{\operatorname{col}'}[\delta_{tk}\mathbf{1}_{N_{dk}-n_{dk}}']\mathbf{1}_{N_d-n_d}\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1} = \varphi_1(1-f_{dt})\mathbf{1}_{n_d}'\Sigma_{ds}^{-1}, \\
\mathbf{b}_{2d}' &= \frac{\sigma^2\varphi_2}{N_{dt}}\underset{1\le k\le m_d}{\operatorname{col}'}[\delta_{tk}\mathbf{1}_{N_{dk}-n_{dk}}']\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{N_{dk}-n_{dk}})\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{V}_{ds}^{-1}, \\
&= \varphi_2(1-f_{dt})\underset{1\le d\le D}{\operatorname{col}'}[\delta_{tk}]\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\Sigma_{ds}^{-1}.
\end{aligned}$$

Let us define $\mathbf{A}_{ds} = \mathbf{I}_{m_d} + \varphi_2 \underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{W}_{ds}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})$. Then

$$\begin{aligned}
\Sigma_{ds}^{-1} &= \mathbf{L}_{ds}^{-1} - \frac{\varphi_1}{1+\varphi_1\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1}, \\
\mathbf{L}_{ds}^{-1} &= \mathbf{W}_{ds} - \varphi_2\mathbf{W}_{ds}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{W}_{ds}.
\end{aligned}$$

By applying the formula $\frac{\partial \mathbf{A}^{-1}}{\partial\gamma} = -\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial\gamma}\mathbf{A}^{-1}$, we obtain the partial derivatives of $\mathbf{L}_{ds}^{-1}$.

$$\begin{aligned}
\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\sigma^2} &= \mathbf{0}_{n_d\times n_d}, \quad \frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_1} = \mathbf{0}_{n_d\times n_d}, \\
\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2} &= -\mathbf{W}_{ds}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{W}_{ds} + \varphi_2\mathbf{W}_{ds}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1} \\
&\cdot \underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{W}_{ds}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}})\mathbf{A}_{ds}^{-1}\underset{1\le k\le m_d}{\operatorname{diag}}(\mathbf{1}_{n_{dk}}')\mathbf{W}_{ds}.
\end{aligned}$$

The partial derivatives of Σ_{ds}^{-1} are

$$\begin{aligned}
\frac{\partial\Sigma_{ds}^{-1}}{\partial\sigma^2} &= \mathbf{0}_{n_d\times n_d}, \\
\frac{\partial\Sigma_{ds}^{-1}}{\partial\varphi_1} &= \frac{1}{[1+\varphi_1\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}]^2}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1}, \\
\frac{\partial\Sigma_{ds}^{-1}}{\partial\varphi_2} &= \frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2} + \frac{\varphi_1^2\mathbf{1}_{n_d}'\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2}\mathbf{1}_{n_d}}{[1+\varphi_1\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}]^2}\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1} \\
&- \frac{\varphi_1}{1+\varphi_1\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}}\left[\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\mathbf{L}_{ds}^{-1} + \mathbf{L}_{ds}^{-1}\mathbf{1}_{n_d}\mathbf{1}_{n_d}'\frac{\partial\mathbf{L}_{ds}^{-1}}{\partial\varphi_2}\right].
\end{aligned}$$

Let $\theta=(\theta_1,\theta_2,\theta_3)=(\sigma^2,\varphi_1,\varphi_2)$. The partial derivatives of $\mathbf{b}'_{1d}$ y $\mathbf{b}'_{2d}$ are

$$\begin{aligned}
\frac{\partial \mathbf{b}'_{1d}}{\partial \sigma^2} &= \mathbf{0}_{1\times n_d} \quad \frac{\partial \mathbf{b}'_{1d}}{\partial \varphi_1}=(1-f_{dt})\mathbf{1}'_{n_d}\left[\Sigma_{ds}^{-1}+\varphi_1\frac{\partial \Sigma_{ds}^{-1}}{\partial \varphi_1}\right],\\
\frac{\partial \mathbf{b}'_{1d}}{\partial \varphi_2} &= \varphi_1(1-f_{dt})\mathbf{1}'_{n_d}\frac{\partial \Sigma_{ds}^{-1}}{\partial \varphi_2},\\
\frac{\partial \mathbf{b}'_{2d}}{\partial \sigma^2} &= \mathbf{0}_{1\times n_d},\\
\frac{\partial \mathbf{b}'_{2d}}{\partial \varphi_1} &= \varphi_2(1-f_{dt})\underset{1\le k\le m_d}{\mathrm{col}'}[\delta_{tk}]\underset{1\le k\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dk}})\frac{\partial \Sigma_{ds}^{-1}}{\partial \varphi_1},\\
\frac{\partial \mathbf{b}'_{2d}}{\partial \varphi_2} &= (1-f_{dt})\underset{1\le k\le m_d}{\mathrm{col}'}[\delta_{tk}]\underset{1\le k\le m_d}{\mathrm{diag}}(\mathbf{1}'_{n_{dk}})\left[\Sigma_{ds}^{-1}+\varphi_2\frac{\partial \Sigma_{ds}^{-1}}{\partial \varphi_2}\right].
\end{aligned}$$

Let us define the matrix $\mathbf{Q}=(q_{ab})_{a,b=1,\dots,3}$, with elements

$$q_{ab}=\left(\frac{\partial \mathbf{b}'_{1d}}{\partial \theta_a}+\frac{\partial \mathbf{b}'_{2d}}{\partial \theta_a}\right)\sigma^2\Sigma_{ds}\left(\frac{\partial \mathbf{b}'_{1d}}{\partial \theta_b}+\frac{\partial \mathbf{b}'_{2d}}{\partial \theta_b}\right)', \quad a,b=1,2,3,$$

and the elements $F_{ab}=F_{\theta_a,\theta_b}$'s of the REML Fisher information matrix. Then

$$\begin{aligned}
g_3(\theta) &\approx \mathrm{tr}\left\{\mathbf{Q}E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}\\
&\approx \mathrm{tr}\left\{\begin{pmatrix} q_{11} & q_{12} & q_{13}\\ q_{21} & q_{22} & q_{23}\\ q_{31} & q_{32} & q_{33}\end{pmatrix}\begin{pmatrix} F_{\sigma^2\sigma^2} & F_{\sigma^2\varphi_1} & F_{\sigma^2\varphi_2}\\ F_{\varphi_1\sigma^2} & F_{\varphi_1\varphi_1} & F_{\varphi_1\varphi_2}\\ F_{\varphi_2\sigma^2} & F_{\varphi_2\varphi_1} & F_{\varphi_2\varphi_2}\end{pmatrix}^{-1}\right\}
\end{aligned}$$

Calculation of $g_4(\theta)$

We have that $g_4(\theta)=\mathbf{a}'_r\Sigma_{er}\mathbf{a}_r$, where

$$\begin{aligned}
\mathbf{a}'_r &= \frac{1}{N_{dt}}\underset{1\le\ell\le D}{\mathrm{col}'}\left[\delta_{d\ell}\underset{1\le k\le m_\ell}{\mathrm{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}-n_{\ell k}}]\right],\\
\mathbf{V}_{er}^{-1} &= \sigma^{-2}\mathbf{W}_r=\sigma^{-2}\underset{1\le d\le D}{\mathrm{diag}}\{\mathbf{W}_{dr}\}.
\end{aligned}$$

Therefore

$$\begin{aligned}
g_4(\theta) &= \frac{1}{N_{dt}}\underset{1\le\ell\le D}{\mathrm{col}'}\left[\delta_{d\ell}\underset{1\le k\le m_\ell}{\mathrm{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}-n_{\ell k}}]\right]\sigma^2\underset{1\le d\le D}{\mathrm{diag}}\{\mathbf{W}_{dr}^{-1}\}\frac{1}{N_{dt}}\underset{1\le\ell\le D}{\mathrm{col}}\left[\delta_{d\ell}\underset{1\le k\le m_\ell}{\mathrm{col}}[\delta_{tk}\mathbf{1}_{N_{\ell k}-n_{\ell k}}]\right]\\
&= \frac{\sigma^2}{N_{dt}^2}\mathbf{1}'_{N_{dt}-n_{dt}}\underset{j\in r}{\mathrm{diag}}\{w_{dtj}^{-1}\}\mathbf{1}_{N_{dt}-n_{dt}}=\frac{\sigma^2}{N_{dt}^2}\sum_{j\in r_{dt}}\frac{1}{w_{dtj}}.
\end{aligned}$$

9.2.6 Simulation experiment 1

The scope of simulation experiment 1 is to analyze the behavior of the REML and H3 estimates of model parameters. Samples are generated in the following way.

- Simulation of explanatory variable: For $d=1,\dots,D$, $t=1,\dots,m_d$, $j=1,\dots,n_{dt}$, generate
$$x_{dtj}=(b_{dt}-a_{dt})U_{dtj}+a_{dt}\ \text{ with } U_{dtj}=\frac{j}{n_{dt}+1},\quad j=1,\dots,n_{dt}.$$
Take $a_{dt}=1$, $b_{dt}=1+\frac{1}{m_d}(m_d(d-1)+t)$, $\quad d=1,\dots,D$, $t=1,\dots,m_d$.
- weights: For $d=1,\dots,D$, $t=1,\dots,m_d$, $j=1,\dots,n_{dt}$, do $w_{dtj}=1/x_{dtj}^{\ell}$, $\ell=0,1/2$, (2 possibilities).
- Simulation of random effects and errors: For $d=1,\dots,D$, $t=1,\dots,m_d$, $j=1,\dots,n_{dt}$, generate
$$u_{1,d}\sim N(0,\sigma_1^2),\quad u_{2,dt}\sim N(0,\sigma_2^2),\quad e_{dtj}\sim N(0,\sigma_0^2),\quad \text{con }\ \sigma_1^2=1,\ \sigma_2^2=1,\ \sigma_0^2=1.$$
- Simulation of target variable: For $d=1,\dots,D$, $t=1,\dots,m_d$, $j=1,\dots,n_{dt}$, generate
$$y_{dtj}=\beta x_{dtj}+u_{1,d}+u_{2,dt}+w_{dtj}^{-1/2}e_{dtj},\quad \text{with } \beta=1.$$

The steps of the simulation experiments are

1. Generate explanatory variables and weights.
2. Repeat $K=1000$ times ($k=1,\dots,K$)
   2.1. Generate a sample of size $n=\sum_{d=1}^{D}\sum_{t=1}^{m_d}n_{dt}$, with the corresponding values of the target vartiable, the random and fixed effects and the errors.
   2.2. Calculate $\widehat{\beta}_{(k)}$, $\widehat{\sigma}_{0,(k)}^2$, $\widehat{\sigma}_{1,(k)}^2$ and $\widehat{\sigma}_{2,(k)}^2$ by using the methods H3 and REML.
3. Output 1 is the empirical mean squared error of $\widehat{\beta}_{(k)}$, $\widehat{\sigma}_{0,(k)}^2$, $\widehat{\sigma}_{1,(k)}^2$ and $\widehat{\sigma}_{2,(k)}^2$, i.e.
$$EMSE(\widehat{\beta})=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\beta}_{(k)}-\beta)^2,\qquad EMSE(\widehat{\sigma}_0^2)=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\sigma}_{0,(k)}^2-\sigma_0^2)^2,$$
$$EMSE(\widehat{\sigma}_1^2)=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\sigma}_{1,(k)}^2-\sigma_1^2)^2,\qquad EMSE(\widehat{\sigma}_2^2)=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\sigma}_{2,(k)}^2-\sigma_2^2)^2.$$
4. Output 2 is the empirical bias of $\widehat{\beta}_{(k)}$, $\widehat{\sigma}_{0,(k)}^2$, $\widehat{\sigma}_{1,(k)}^2$ and $\widehat{\sigma}_{2,(k)}^2$:
$$B(\widehat{\beta})=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\beta}_{(k)}-\beta),\qquad B(\widehat{\sigma}_0^2)=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\sigma}_{0,(k)}^2-\sigma_0^2),$$
$$B(\widehat{\sigma}_1^2)=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\sigma}_{1,(k)}^2-\sigma_1^2),\qquad B(\widehat{\sigma}_2^2)=\frac{1}{K}\sum_{k=1}^{K}(\widehat{\sigma}_{2,(k)}^2-\sigma_2^2).$$

In the simulation experiment we take $D = 30$, $m_d = 5$, $d = 1,\dots,D$. We carry out 10 realizations of the simulation experiments with the sample sizes presented in Table 9.2.6.1.

g	1	2	3	4	5	6	7	8	9	10
$n_{dt}^{(g)}$	3	4	5	6	7	8	9	10	15	20
$n_d^{(g)}$	15	20	25	30	35	40	45	50	75	100
$n^{(g)}$	450	600	750	900	1050	1200	1350	1500	2250	3000

Table 9.2.6.1. Sample sizes in Experiment 1.

The obtained results are presented in Table 9.2.6.2.

	n	450	600	750	900	1050	1200	1350	1500
	n_d	15	20	25	30	35	40	45	50
	n_{dt}	3	4	5	6	7	8	9	10
EMSE	β	0.126	0.087	0.068	0.054	0.046	0.040	0.034	0.030
	σ_0^2	6.583	4.386	3.403	2.605	2.187	1.879	1.661	1.479
	σ_1^2	110.26	106.36	106.56	105.14	102.23	101.38	98.96	99.23
	σ_2^2	30.90	26.26	23.87	22.86	21.75	21.49	20.40	20.26
BIAS	β	0.069	-0.069	-0.036	-0.003	0.082	-0.083	0.065	0.009
	σ_0^2	-0.311	1.011	0.328	-0.179	-0.184	-0.239	-0.087	-0.385
	σ_1^2	-8.586	-0.432	0.562	5.623	-1.291	2.186	-4.623	1.709
	σ_2^2	-0.542	-2.486	1.926	0.680	-0.288	0.336	-1.626	0.947

Table 9.2.6.2. EMSE and BIAS (multiplied by 10^3) of $\widehat{\beta}$, $\widehat{\sigma}_0^2$, $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ for $\ell = 0$.

9.2.7 Simulation experiment 2

The scope of simulation experiment 1 is to analyze the behavior of the EBLUPs. Samples are generated in the following way.

1. Generation of deterministic elements

   - Simulation of explanatory variables: for $d = 1,\dots,D$, $t = 1,\dots,m_d$, $j = 1,\dots,N_{dt}$, generate
   $$x_{dtj} = (b_{dt} - a_{dt})U_{dtj} + a_{dt} \text{ with } U_{dtj} = \frac{j}{N_{dt}+1}, \quad j = 1,\dots,N_{dt}.$$
   Take $a_{dt} = 1$, $b_{dt} = 1 + \frac{1}{m_d}(m_d(d-1)+t)$, $\quad d = 1,\dots,D$, $t = 1,\dots,m_d$.
   - Weights: For $d = 1,\dots,D$, $t = 1,\dots,m_d$, $j = 1,\dots,N_{dt}$, hacer $w_{dtj} = 1/x_{dtj}^{\ell}$, $\ell = 0, 1/2$.

2. Repeat $K = 100000$ times ($k = 1,\dots,K$)

   (a) Generation of random elements

   - Simulation random effects and errors: For $d = 1,\dots,D$, $t = 1,\dots,m_d$, $j = 1,\dots,N_{dt}$, generate
   $$u_{1,d}^{(k)} \sim N(0,\sigma_1^2), \quad u_{2,dt}^{(k)} \sim N(0,\sigma_2^2), \quad e_{dtj}^{(k)} \sim N(0,\sigma_0^2), \quad \text{with } \sigma_0^2 = \sigma_1^2 = \sigma_2^2 = 1.$$

- Simulation of the target variable: For $d=1,\ldots,D,\ t=1,\ldots,m_d,\ j=1,\ldots,N_{dt}$, generate

$$y_{dtj}^{(k)}=\beta x_{dtj}+u_{1,d}^{(k)}+u_{2,dt}^{(k)}+w_{dtj}^{-1/2}e_{dtj}^{(k)},\quad \text{with } \beta=1.$$

(b) Extraction of samples. For $d=1,\ldots,D$, $t=1,\ldots,m_d$, select the n_{dt} units of the level dt in positions $\left[\frac{N_{dt}}{1+n_{dt}}\right]j$, $j=1,\ldots,n_{dt}$.

(c) Calculate $\widehat{\beta}_{(k)}$, $\widehat{\sigma}_{0,(k)}^2$, $\widehat{\sigma}_{1,(k)}^2$ and $\widehat{\sigma}_{2,(k)}^2$ by using the REML method.

(d) For $d=1,\ldots,D$, $t=1,\ldots,m_d$ calculate $\widehat{\overline{Y}}_{dt}^{eblup,(k)}$.

3. Output: For $d=1,\ldots,D$, $t=1,\ldots,m_d$ calculate

$$EMSE_{dt}=\frac{1}{K}\sum_{k=1}^{K}\left(\widehat{\overline{Y}}_{dt}^{eblup,(k)}-\overline{Y}_{dt}^{(k)}\right)^2,\quad BIAS_{dt}=\frac{1}{K}\sum_{k=1}^{K}\left(\widehat{\overline{Y}}_{dt}^{eblup,(k)}-\overline{Y}_{dt}^{(k)}\right).$$

$$EMSE=\frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d}EMSE_{dt},\quad BIAS=\frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d}BIAS_{dt}$$

where $\overline{Y}_{dt}^{(k)}=\frac{1}{N_{dt}}\sum_{j=1}^{N_{dt}}y_{dtj}^{(k)}$.

The simulation experiment is carried out with $D=30$, $m_d=5$, $d=1\ldots,D$. The obtained results appears in Table 9.2.7.1.

N	4500	6000	7500	9000	10500	12000	13500	15000
N_d	150	200	250	300	350	400	450	500
N_{dt}	30	40	50	60	70	80	90	100
n_{dt}	3	4	5	6	7	8	9	10
BIAS	0.00043	-0.00004	0.00122	-0.00010	0.00008	-0.00034	0.00001	0.00007
MSE	0.25191	0.19521	0.15969	0.13521	0.11699	0.10357	0.09280	0.08401

Table 9.2.7.1. EMSE and BIAS of $\widehat{\overline{Y}}_{dt}^{eblup}$ for $m_d=5$ and $\ell=0$.

9.2.8 Simulation experiment 3

The scope of simulation experiment 1 is to analyze the behavior of the mean squared error estimators of the EBLUPs. Samples are generated in the following way.

1. Generation of deterministic elements

- Simulation of explanatory variables: For $d=1,\ldots,D$, $t=1,\ldots,m_d,\ j=1,\ldots,N_{dt}$, generate

$$x_{dtj}=(b_{dt}-a_{dt})U_{dtj}+a_{dt}\ \text{ with } U_{dtj}=\frac{j}{N_{dt}+1},\quad j=1,\ldots,N_{dt}.$$

Take $a_{dt}=1$, $b_{dt}=1+\frac{1}{m_d}(m_d(d-1)+t)$, $\quad d=1,\ldots,D$, $t=1,\ldots,m_d$.

- Pesos: For $d=1,\ldots,D$, $t=1,\ldots,m_d,\ j=1,\ldots,N_{dt}$, do $w_{dtj}=1/x_{dtj}^{\ell}$, $\ell=0,\,1/2$.

2. Repeat $K = 100000$ times ($k = 1,\dots,K$)

   (a) Generation of random elements

   - Simulation of random factors and errors: For $d = 1,\dots,D, t = 1,\dots,m_d, j = 1,\dots,N_{dt}$, generate
$$u_{1,d}^{(k)} \sim N(0,\sigma_1^2), \quad u_{2,dt}^{(k)} \sim N(0,\sigma_2^2), \quad e_{dtj}^{(k)} \sim N(0,\sigma_0^2), \quad \text{with } \sigma_0^2 = \sigma_1^2 = \sigma_2^2 = 1.$$
   - Simulation of target variables: For $d = 1,\dots,D, t = 1,\dots,m_d, j = 1,\dots,N_{dt}$, generate
$$y_{dtj}^{(k)} = \beta x_{dtj} + u_{1,d}^{(k)} + u_{2,dt}^{(k)} + w_{dtj}^{-1/2} e_{dtj}^{(k)}, \quad \text{with } \beta = 1.$$
   - Extraction of samples. For $d = 1,\dots,D, t = 1,\dots,m_d$, select the n_{dt} units of the level dt in positions $\left[\frac{N_{dt}}{1+n_{dt}}\right] j$, $j = 1,\dots,n_{dt}$.
   - Calculate $\widehat{\beta}_{(k)}$, $\widehat{\sigma}_{0,(k)}^2$, $\widehat{\sigma}_{1,(k)}^2$ and $\widehat{\sigma}_{2,(k)}^2$ by using the REML.
   - For $d = 1,\dots,D, t = 1,\dots,m_d$ calculate $\widehat{\overline{Y}}_{dt}^{eblup,(k)}$ and $mse(\widehat{\overline{Y}}_{dt}^{eblup,(k)})$.

   (b) For $d = 1,\dots,D, t = 1,\dots,m_d$ read the values of $EMSE_{dt}$ calculated in Simulation 2.

3. Output: For $d = 1,\dots,D, t = 1,\dots,m_d$ calculate
$$E_{dt} = \frac{1}{K}\sum_{k=1}^{K}\left(mse(\widehat{\overline{Y}}_{dt}^{eblup,(k)}) - EMSE_{dt}\right)^2 B_{dt} = \frac{1}{K}\sum_{k=1}^{K}\left(mse(\widehat{\overline{Y}}_{dt}^{eblup,(k)}) - EMSE_{dt}\right),$$
$$E = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d} E_{dt}, \quad B = \frac{1}{M}\sum_{d=1}^{D}\sum_{t=1}^{m_d} B_{dt}$$

The simulation experiment is carried out with $D = 30$, $m_d = 5$, $d = 1\dots,D$. The obtained results appears in Table 9.2.8.1.

N	4500	6000	7500	9000	10500	12000	13500	15000
N_d	150	200	250	300	350	400	450	500
N_{dt}	30	40	50	60	70	80	90	100
n_{dt}	3	4	5	6	7	8	9	10
B	0.6933	0.6859	0.6760	0.6710	0.6681	0.6653	0.6668	0.6625
E	0.5096	0.4966	0.4813	0.4727	0.4689	0.4647	0.4673	0.4607

Table 9.2.8.1. E and B of $mse(\widehat{\overline{Y}}_{dt}^{eblup})$ for $m_d = 5$ and $\ell = 0$.

Chapter 10

M-quantile methods

In recent years there have been significant developments in model-based small area estimation. The most popular approach to small area estimation employs random effects models for estimating domain specific parameters (see Rao (2003)). An alternative approach to small area estimation that relaxes the parametric assumptions of random effects models by employing M-quantile models was recently proposed by Chambers and Tzavidis (2006) and Tzavidis et al. (2010). This model is presented in section 10.1 and estimation of small area means and quantiles under the M-quantile model is discussed. We further discuss the estimation of the Mean Squared Error (MSE) of the small area estimators and we present model-based simulation results for assesing the propoerties of point and MSE estimators. Having developed the methodology for estimating small area averages and quantiles, in Section 1.2 we focus on the estimation of poverty indicators which present a special case of estimating small area quantiles. We consider estimation for two popular poverty indicators namely, the Head Count Ratio (HCR) and the Poverty Gap. In addition, we also consider estimation for fuzzy set indicators that have more recently attracted interest in poverty studies. Two methods of poverty estimation are considered namely, the EBP approach (Molina and Rao, 2009), see Chapter 2, and the M-quantile approach that is based on the methodology proposed in Tzavidis et al. (2010). The two approaches are then contrasted in a model-based study and in a design-based simulation study.

When the functional form of the relationship between the response variable and the covariates is unknown or has a complicated functional form, an approach based on use of a nonparametric regression model using penalized splines can offer significant advantages compared with one based on a linear model. Pratesi et al. (2008) and Pratesi et al. (2009) have extended the p-spline regression model to the M-quantile method for the estimation of the small area parameters using a nonparametric specification of the conditional M-quantile of the response variable given the covariates. The model is discussed in section 10.3.

M-quantile models assume independence of the small area effects. In some applications, however, observations that are spatially close may be more related than observations that are further apart. This spatial correlation can be accounted for by assuming that the regression coefficients vary spatially across the geography of interest. In a recent paper Salvati et al. (2008) proposed an M-quantile Geographically Weighted Regression (GWR) small area model extending the traditional M-quantile regression model by allowing local rather than global parameters to be estimated. The model is shown in section 10.4.

10.1 Linear M-quantile regression models

A recently proposed approach to small area estimation is based on the use of M-quantile models (see Chambers and Tzavidis 2006)). M-quantile regression provides a "quantile-like" generalization of regression based on influence functions (see Breckling and Chambers (1988)). M-quantile models do not depend on strong distributional assumptions nor on a predefined hierarchical structure, and outlier robust inference is automatically performed when these models are fitted. The M-quantile of order q for the conditional density of y given $\mathbf{X}$ is defined as the solution $Q_q(x;\psi)$ of the estimating equation $\int \psi_q(y-Q)f(y|\mathbf{X})dy = 0$, where ψ denotes the influence function associated with the M-quantile. In a linear M-quantile regression model the q-th M-quantile $Q_q(x,\psi)$ of the conditional distribution of y given $\mathbf{X}$ is such that

$$Q_q(x;\psi) = \mathbf{X}\beta_\psi(q) \tag{10.1}$$

where $\psi_q(r_{iq\psi}) = 2\psi\{s^{-1}r_{iq\psi}\}\ \left\{qI(r_{jq\psi} > 0) + (1-q)I(r_{jq\psi} \leq 0)\right\}$ and s is a suitable robust estimate of scale, e.g. the MAD estimate $s = median\left|r_{jq\psi}\right|/0.6745$. A popular choice for the influence function is the Huber Proposal 2, $\psi(u) = uI(-c \leq u \leq c) + c\,\mathtt{sgn}(u)$. However, other influence functions are also possible. For specified q and continuous ψ, an estimate $\hat{\beta}_\psi(q)$ of $\beta_\psi(q)$ is obtained via iterative weighted least squares. Note that there is a different set of regression parameters for each q.

10.1.1 Estimation of small area means and quantiles

We now consider the problem of estimating the small area mean and the cumulative distribution function of a given variable of interest using M-quantile models under a unified estimation framework for estimating any small area target parameter that was defined by Tzavidis et al. (2010).

Let $\Omega_d = \{1,\ldots,N_d\}$ be the population of area d. Let $\mathbf{y}_d = (y_1,\ldots,y_{N_d})'$ denote the variable values for the N_d small area population elements. We consider a sample $s_d \subset \Omega_d$, of $n_d \leq N_d$ units, and we denote with $r_d = \Omega_d - s_d$ the set of non sampled units. For each population unit j, let $\mathbf{x}_j = (x_{1j},\ldots,x_{pj})$ denote a vector of p known auxiliary variables. The small area specific empirical distribution function of y for area d is

$$F_d = N_d^{-1}\Big[\sum_{j\in s_d} \mathrm{I}(y_j \leqslant t) + \sum_{j\in r_d} \mathrm{I}(y_j \leqslant t)\Big]. \tag{10.2}$$

The problem of estimating $F_d(t)$ given the sample data essentially reduces to predicting the values y_j for the non-sampled units in small area d. One straightforward way of achieving this is to simply replace the unknown non-sample values of y (10.2) by their predicted values $\hat{y}_j$ under an appropriate model, leading to a plug-in estimator of (10.2) of the form

$$\hat{F}_d = N_d^{-1}\Big[\sum_{j\in s_d} \mathrm{I}(y_j \leqslant t) + \sum_{j\in r_d} \mathrm{I}(\hat{y}_j \leqslant t)\Big]. \tag{10.3}$$

An estimator of the mean $\overline{Y}_d$ of y in area d is then defined by the value of the mean functional defined by (10.3). This leads to the usual plug-in estimator of the mean,

$$\hat{\overline{Y}}_d = \int_{-\infty}^{\infty} t d\hat{F}_d(t) = N_d^{-1}\left(\sum_{\mathbf{x}\in s_d} y_j + \sum_{j\in r_d} \hat{y}_j\right).$$

The predicted value of a non-sample unit j in area d corresponds to an estimate $\hat{\mu}_j$ of its expected value given that it is located in area d.

Following Chambers and Tzavidis (2006), an alternative to random effects for characterizing the variability across the population is to use the M-quantile coefficients of the population units. For unit j with values y_j and $\mathbf{x}_j$, this coefficient is the value θ_j such that $Q_{\theta_j}(\mathbf{x}_j;\psi) = y_j$. These authors observed that if a hierarchical structure does explain part of the variability in the population data, units within clusters (areas) defined by this hierarchy are expected to have similar M-quantile coefficients. When the conditional M-quantiles are assumed to follow a linear model, with $\beta_\psi(q)$ a sufficiently smooth function of q, this suggests an estimator of the distribution function:

$$\hat{F}_d^{MQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{j\in r_d} I(\mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d) \le t)\right\} \tag{10.4}$$

where $\mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d)$ is used to predict the unobserved value y_j for population unit $j \in r_d$. When there are no sampled observations in area d then $\hat{\theta}_d = 0.5$.

Using the empirical distribution function and the linear M-quantile small area models one can defined the estimator of the small area mean as:

$$\hat{\bar{Y}}_d^{MQ}(t) = \int_{-\infty}^{\infty} t d\hat{F}_d^{MQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} y_j + \sum_{j\in r_d} \mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d)\right\}. \tag{10.5}$$

We refer to the small area estimator that can be expressed as functionals of (10.3), with non-sample predictions derived as estimates of expected values.

Chambers and Tzavidis (2006) observed that the naive M-quantile mean estimator (10.5) can be biased. The distribution function estimator (10.3) underlying (10.4) is not consistent in general. Thus, when the non-sample predicted values in (10.3) are estimated expectations that converge in probability to the actual expected values, we see that

$$\sum_{j\in r_d} \mathrm{I}(\hat{y}_j \leqslant t) = \sum_{j\in r_d} \mathrm{I}(y_j - (y_j - \hat{y}_j) \leqslant t) = \sum_{j\in r_d} \mathrm{I}(y_j \leqslant t + \varepsilon_j) \neq \sum_{j\in r_d} \mathrm{I}(y_j \leqslant t),$$

where ε_j are the actual regression errors. If these errors are independently and identically distributed symmetrically about zero we expect that the summation on the left hand side above will closely approximate the summation on the right for values of t near the median of the non-sampled area d values of y but not anywhere else. More generally, for heteroskedastic and/or asymmetric errors this correspondence will typically occur elsewhere in the support of y, although one would expect that in most reasonable situations it will be "close" to the median of y. In other words, it is not advisable to use (10.3) to predict a quantile of the area d distribution of y other than the median.

By combining a smearing argument (Duan, 1983) with a model for the finite population distribution of y, Chambers and Dunstan (1986) (hereafter referred to as CD) developed a model-consistent estimator for a finite population distribution function. In the context of the small area distribution function (10.2), and assuming that the residuals are homoskedastic within the small area of interest, this is of the form

$$\hat{F}_d^{CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{k\in r_d} n_d^{-1} \sum_{j\in s_d} I(\hat{y}_k + (y_j - \hat{y}_j) \le t)\right\}. \tag{10.6}$$

It can be shown that under the CD estimator of the small area distribution function the mean functional defined by (10.6) takes the value

$$\hat{\bar{Y}}_d^{CD} = \int_{-\infty}^{\infty} t d\hat{F}_d^{CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} y_j + \sum_{j\in r_d}\hat{y}_j + (f_d^{-1}-1)\sum_{j\in s_d}(y_j-\hat{y}_j)\right\} \tag{10.7}$$

where $f_d = n_d N_d^{-1}$ is the sampling fraction in area d, $\hat{y}_j = \mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d)$ and $\hat{y}_j$ can be obtained under the linear M-quantile small area model. We refer to (10.7) as the bias adjusted M-quantile mean predictor. Due to the bias correction in (10.7), this predictor will have higher variability than (10.5) and so it should only be used when (10.4) is expected to have substantial bias, e.g. when there are large outlying data points. An alternative approach for dealing with this bias-variance trade off is to limit the variability of the bias correction term in (10.7) by using robust (huberized) residuals instead of raw residuals. In particular,

$$\hat{F}_d^{CDRob}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j\le t) + \sum_{k\in r_d} n_d^{-1}\sum_{j\in s_d} I\left(\hat{y}_k + \nu_j\psi\{y_j-\hat{y}_j\}\leqslant t\right)\right\} \tag{10.8}$$

where ν_j is a robust estimate of scale for area individual j in area d.

Wang and Dorfman (1996) pointed out that the CD estimator (10.6) is model-consistent but design-inconsistent. An alternative to this estimator that is both design-consistent and model-consistent has been proposed by Rao et al. (1990) (hereafter referred to as RKM). Under simple random sampling within the small areas the RKM estimator of the finite population distribution function is

$$\hat{F}_d^{RKM}(t) = n_d^{-1}\left\{\sum_{j\in s_d} I(y_j\le t) + N_d^{-1}\sum_{k\in r_d} n^{-1}\sum_{j\in s_d} I(y_j-\hat{y}_j\le t-\hat{y}_k)\right. \tag{10.9}$$

$$\left. -(n_d^{-1}-N_d^{-1})\sum_{k\in s_d} n_d^{-1}\sum_{j\in s_d} I(y_j-\hat{y}_j\le t-\hat{y}_k)\right\}.$$

Chambers et al. (1992) compared the large-sample mean squared errors of (10.6) and (10.9) and concluded that neither dominates the other. When the model is correctly specified we expect (10.6) to outperform (10.9). However RKM demonstrated that (10.6) can be substantially biased when model assumptions fail, while (10.9) is much less sensitive. Here we just note that the RKM estimator can be used to define an estimator of a small area characteristic that can be represented as a functional of the small area distribution function in exactly the same way as the CD-type estimator (10.7) can be used for this purpose. In general, the resulting estimators will not be the same. An exception is the RKM-based estimator of the area d mean, which is the same as the CD-based estimator of this mean under simple random sampling.

Turning now to the small area quantiles we note that an estimator of the qth quantile of the distribution of y in area d is straightforwardly defined as the solution to the estimating equation

$$\int_{-\infty}^{\hat{\mu}_{qd}} d\hat{F}_d(t) = q, \tag{10.10}$$

where $\hat{F}_d(t)$ is suitable estimator of the area d distribution of y such as the CD or the RKM estimators and $\hat{\mu}_{qd}$ is the estimated qth quantile in small area d. As the preceding discussion makes clear, we anticipate that a better approach for quantiles other than the median is to use either the CD-type specifications or the RKM specification for $\hat{F}_d(t)$, with $\hat{y}_j$ defined by an M-quantile linear small area model.

10.1.2 Mean Squared Error (MSE) estimation for estimators of small area means and quantiles

A robust mean squared error estimation method for the naive M-quantile estimator (10.5) was described in Chambers and Tzavidis (2006). Here we extend this argument to define an estimator that is a first order approximation to the mean squared error of the estimator (10.6) of the small area mean when this is based on an M-quantile regression fit. A more detailed discussion of this approach to mean squared error estimation is set out in Chambers et al. (2008). To start, we note that since an iteratively reweighted least squares algorithm is used to calculate the M-quantile regression fit at $\hat{\theta}_d$, we have

$$\hat{\beta}_\psi(\hat{\theta}_d) = (\mathbf{X}_s' \mathbf{W}_{s_d} \mathbf{X}_s)^{-1} \mathbf{X}_s' \mathbf{W}_{s_d} \mathbf{y}_s$$

where $\mathbf{X}_s$ and $\mathbf{y}_s$ denote the matrix of sample x values and the vector of sample y values respectively, and $\mathbf{W}_{s_d}$ denotes the diagonal weight matrix of order n that defines the estimator of the M-quantile regression coefficient with $q = \hat{\theta}_d$. It immediately follows that (10.5) can be written

$$\hat{\bar{Y}}_d^{MQ} = \mathbf{w}_{s_d}' \mathbf{y}_s, \tag{10.11}$$

where $\mathbf{w}_{s_d} = (w_{jd}) = n_d^{-1} \Delta_{s_d} + (1 - N_d^{-1} n_d) \mathbf{W}_d \mathbf{X}_s (\mathbf{X}_s' \mathbf{W}_d \mathbf{X}_s)^{-1} \{\bar{\mathbf{x}}_{r_d} - \bar{\mathbf{x}}_{s_d}\}$ with Δ_{sd} denoting the n-vector that "picks out" the sample units from area d. Here $\bar{\mathbf{x}}_{s_d}$ and $\bar{\mathbf{x}}_{r_d}$ denote the sample and non-sample means of x in area d. Also, these weights are 'locally calibrated' on x since

$$\sum_{j \in s} w_{jd} \mathbf{x}_j = \bar{\mathbf{x}}_{s_d} + (1 - f_d)(\bar{\mathbf{x}}_{r_d} - \bar{\mathbf{x}}_{s_d}) = \bar{\mathbf{x}}_d.$$

A first order approximation to the mean squared error of (10.11) then treats the weights as fixed and applies standard methods of robust mean squared error estimation for linear estimators of population quantities (see Royall and Cumberland (1978)). With this approach, the prediction variance of $\hat{\bar{Y}}_d^{CD}$ is estimated by

$$\widehat{Var}(\hat{\bar{Y}}_d^{CD}) = \sum_{g=1}^{d} \sum_{j \in s_g} \lambda_{jdg} \left(y_j - \mathbf{x}_j \hat{\beta}_\psi(\hat{\theta}_g) \right)^2, \tag{10.12}$$

where $\lambda_{jdg} = \{(w_{jd} - 1)^2 + (n_d - 1)^{-1}(N_d - n_d)\} \mathrm{I}(g = d) + w_{jg}^2 \mathrm{I}(g \neq d)$. This prediction variance estimator implicitly assumes a model where the regression of y on x varies between areas, and that this variation is consistently estimated by the fit of the M-quantile regression model in each area. Furthermore, since the weights defining $\hat{\bar{Y}}_d^{CD}$ are locally calibrated on x, it immediately follows that (10.6) is unbiased under the same model and hence no correction for its bias is necessary when estimating its mean squared error. This can be compared with the estimator of the mean squared error of the naive M-quantile estimator $\hat{\bar{Y}}_d^{MQ}$ described in Chambers and Tzavidis (2006), which includes a squared bias term.

The linearization-based prediction variance estimator (10.12) is defined only when the estimator of interest can be written as a weighted sum of sample values. Consequently, it cannot be used with quantile estimators defined by solving (10.10). In this section we describe a nonparametric bootstrap approach to MSE estimation of small area quantiles that was described in Tzavidis et al.(2010) and is based on the approach of Lombardía et al. (2003).

We define two bootstrap schemes that resample residuals from an M-quantile model fit. The first scheme draws samples from the empirical distribution of suitably recentered residuals. The second scheme draws samples from a smoothed version of this empirical distribution. Using these two schemes, we generate a bootstrap population, from which we then draw bootstrap small area samples. In order to define the bootstrap population, we first calculate the M-quantile small area model residuals $\varepsilon_{jd} = y_{jd} - \hat{\beta}_{\psi}(\hat{\theta}_d)$.

A bootstrap finite population $U^* = (y^*_{jd}, \mathbf{x}_{jd}), j \in U, d = 1, \cdots, D$ with

$$y^*_{jd} = \mathbf{x}_{jd}\hat{\beta}_{\psi}(\hat{\theta}_d) + \varepsilon^*_{jd}$$

is then generated, where the bootstrap residuals ε^*_{jd} are obtained by sampling from an estimator of the distribution function $\hat{G}(u)$ of the model residuals ε_{jd}. In order to define $\hat{G}(u)$ we consider two approaches: (i) sampling from the empirical distribution function of the model residuals and (ii) sampling from a smoothed distribution function of these residuals. In each case sampling of the residuals can be done in two ways, (i) by sampling from the distribution of all residuals without conditioning on the small area - we refer to this as the unconditional approach; (ii) by sampling from the conditional distribution of residuals within small area d - we refer to this as the conditional approach. The empirical unconditional distribution of the residuals is

$$\hat{G}(u) = n^{-1}\sum_{d=1}^{D}\sum_{j\in s_d} \mathrm{I}(\varepsilon_{jd} - \bar{\varepsilon}_s \le u)$$

where $\bar{\varepsilon}_s$ is the sample mean of the ε_{jd}. Similarly, the empirical conditional distribution of these residuals in area d is

$$\hat{G}_d(u) = n_d^{-1}\sum_{j\in s_d} \mathrm{I}(\varepsilon_j - \bar{\varepsilon}_{sd} \le u)$$

where $\bar{\varepsilon}_{sd}$ is the sample mean of the ε_{jd} in area d. A smoothed estimator of the unconditional distribution is

$$\hat{G}(u) = n^{-1}\sum_{d=1}^{D}\sum_{j\in s_d} K\Big(\frac{u - (\varepsilon_{jd} - \bar{\varepsilon}_s)}{h}\Big)$$

where $h > 0$ is a smoothing parameter and K is the distribution function corresponding to a bounded symmetric kernel density k,

$$K(u) = \int_{-\infty}^{u} k(z)\,dz.$$

Similarly a smoothed estimator of the conditional distribution in area d is

$$\hat{G}_d(u) = n_d^{-1}\sum_{j\in s_d} K\Big(\frac{u - (\varepsilon_{jd} - \bar{\varepsilon}_s)}{h_d}\Big)$$

, where $h_d > 0$ and K are the same as above. K is defined by the Epanechnikov kernel,

$$k(u) = \frac{3}{4}(1-u^2)\mathrm{I}(|u| < 1),$$

while the smoothing parameters h and h_d are chosen so that they minimize the cross-validation criterion suggested by Bowman et al. (1998). That is, in the unconditional case h is chosen in order to minimize

$$CV(h) = n^{-1}\sum_{d=1}^{D}\sum_{j\in s_d}\int \Big(\mathrm{I}\Big((\varepsilon_{jd} - \bar{\varepsilon}_s) \le u\Big) - \hat{G}_{-j}(u)\Big)^2 du,$$

where $\hat{G}_{-j}(u)$ is the version of $G(u)$ that omits sample unit j with the extension to the conditional case being obvious. It can be shown (see Section 1.5 in Li and Racine (2007)) that choosing h and h_d in this way is asymptotically equivalent to using the MSE optimal values of these parameters. In the simulation studies reported in the next section, we compute both the conditional and unconditional smoothed distribution functions of residuals using the np package of Hayfield and Racine (2008) in the R software environment (R Development Core Team (2008)) that implements the above approach. In either case, bootstrap samples s_d^* are then drawn using simple random sampling within the small areas and without replacement. In what follows we denote by $F_{N,d}(t)$ the unknown true distribution function of the finite population values in area d, by $\hat{F}_d^{CD}(t)$ the CD estimator of $F_{N,d}(t)$ based on sample s_d, by $F_{N,d}^*(t)$ the known true distribution function of the bootstrap population U_d^* in area d, and by $\hat{F}_d^{*,CD}(t)$ the CD estimator of $F_{N,d}^*(t)$ based on bootstrap sample s_d^*. We then estimate the mean squared error of the CD estimator (10.6) as follows. Starting from sample s, selected from a finite population U without replacement, we generate B bootstrap populations, U^{*b}, using one of the four above mentioned methods for estimating the distribution of the residuals. From each bootstrap population, U^{*b}, we select L samples using simple random sampling within the small areas and without replacement in a way such that $n_d^* = n_d$. Finally, bootstrap estimators of the bias and variance of the CD estimator of the distribution function in area j are defined respectively by

$$\widehat{Bias}_d = B^{-1}L^{-1}\sum_{b=1}^{B}\sum_{l=1}^{L}\left(\hat{F}_d^{bl,CD}(t) - F_{N,d}^{*b}(t)\right)$$

and

$$\widehat{Var}_d = B^{-1}L^{-1}\sum_{b=1}^{B}\sum_{l=1}^{L}\left(\hat{F}_d^{*bl,CD}(t) - \bar{\hat{F}}_d^{*bl,CD}(t)\right)^2,$$

, where

$$\bar{\hat{F}}_d^{*bl,CD}(t) = L^{-1}\sum \hat{F}_d^{*bl,CD}(t)$$

is the distribution function of the bth bootstrap population and $\hat{F}_d^{*bl,CD}(t)$ is the CD estimator of $F_{N,d}^{*,b}(t)$ computed from the lth sample of the bth bootstrap population, $(b = 1,\cdots,B, l = 1,\cdots,L)$. The bootstrap estimator of the mean squared error of the CD-based small area estimate is finally calculated as

$$\widehat{MSE}_d\left(\hat{F}_d^{CD}(t)\right) = \widehat{Var}_d + \widehat{Bias}_d^2. \qquad (10.13)$$

Note that the above bootstrap procedure can also be used to construct confidence intervals for the value of $F_{N,d}(t)$ by "reading off" appropriate quantiles of the bootstrap distribution of $F_d^{CD}(t)$. Clearly, the procedure can be used with any small area estimator, and so can be used to compute bootstrap estimates of the mean squared errors of the M-quantile estimates of the small area means as well as associated confidence intervals, which can be contrasted with the estimates derived using the analytic mean squared error estimator.

10.1.3 Model-based simulations for the estimators of small area means and quantiles

In this section we present results from a simulation study used to compare the performance of the robust M-quantile small area estimators. In particular, we considered a model-based simulation in which small area population and sample data were simulated based on a two-level linear mixed model with different parametric assumptions for the area and unit level random effects.

Two methods were used to simulate bivariate population values (y,x) in $d = 30$ small areas. In both, $N = 232,500$ with $N_d = 500$ in area d. For each area d, we selected a simple random sample (without replacement) of size $n_d = 30$, leading to an overall sample size of $n = 900$. The sample values of y and the population values of x were then used to estimate the small area target parameters, which were taken to be the small area means and selected quantiles of y. This process was repeated 1000 times.

The first simulation experiment (scenario 1) generated population values of y using $y_{jd} = 5 + x_{jd} + \gamma_j + \varepsilon_{jd}$, with the x_{jd} generated as independently and identically distributed realisations from $N(\xi_d, \xi_d^2/36)$. The small area x-means ξ_d were themselves drawn at random from the uniform distribution on the interval (40, 120), and held fixed over the simulations. Similarly, the random effects γ_d and ε_{jd} were independently and identically generated as $N(0,1)$ and $N(0,64)$ realisations respectively. The second simulation experiment (scenario 2) generated values of the target variable using the same linear model as in scenario 1, but in this case values of x_{jd} were generated as independently and identically distributed realisations from $\chi^2(Z_d)$, with the Z_d drawn at random from the integers 1 to 200, and held fixed over the simulations. Also, the random effects γ_d and ε_{jd} were independently and identically generated as mean-corrected $\chi^2(1)$ and $\chi^2(3)$ realisations respectively. The purpose of scenario 2 was to examine the effect of misspecification of the Gaussian assumptions of a mixed model. Two different types of small area linear models were fitted to the sample data obtained in these Monte Carlo simulations. These were (a) a linear mixed model, and (b) a linear M-quantile regression specification. The random intercepts model used in (a) was fitted using the default settings of the lme function (see Section 10.3 in Venables and Ripley (2002)) in the R software package. The M-quantile linear regression fit underpinning (b) was obtained using a modified version of the rlm function (see Section 8.3 in Venables and Ripley (2002)) in R. Estimated model coefficients obtained from these fits were then used to compute a range of EBLUP and M-quantile-based estimators of means and quantiles in the different areas.

Biases and mean squared errors over these simulations, averaged over the 30 areas, are set out in Table 10.1 (scenario 1) and in Table 10.2 (scenario 2). Under scenario 1 all estimators performed reasonably well. The differences between the estimators were much more pronounced under scenario 2 (area effects distributed as chi-squared). Here we see that the use of naive estimators led to substantial biases as far as quantiles were concerned. In contrast, the estimators (both EBLUP and M-quantile) based on (10.6) and (10.9) were essentially unbiased, even for extreme quantiles, with the CD-based estimators somewhat more efficient. On the basis of these results it would appear that estimators that are defined as functionals of the CDF estimators (10.6) or (10.9) are preferable if there is concern about misspecification of the distribution of area effects.

Method	Target Parameters					
	10th	25th	50th	Mean	75th	90th
	Relative Bias (%)					
EBLUP/Naive	0.088	0.041	-0.002	-0.002	-0.036	-0.062
EBLUP/CD	0.096	0.046	0.051	-0.002	0.072	0.160
EBLUP/RKM	0.005	0.015	-0.024	-0.002	0.015	0.105
M-quantile/Naive	0.090	0.044	0.003	0.003	-0.030	-0.055
M-quantile/CD	0.058	0.003	-0.003	-0.002	0.008	0.064
M-quantile/RKM	-0.011	0.002	0.008	-0.002	0.009	0.014
	Relative RMSE (%)					
EBLUP/Naive	0.29	0.23	0.20	0.23	0.19	0.19
EBLUP/CD	0.34	0.25	0.22	0.24	0.21	0.26
EBLUP/RKM	0.31	0.25	0.21	0.24	0.20	0.20
M-quantile/Naive	0.46	0.38	0.33	0.32	0.31	0.30
M-quantile/CD	0.34	0.25	0.21	0.24	0.21	0.24
M-quantile/RKM	0.32	0.25	0.22	0.24	0.21	0.22

Table 10.1: Model-based simulation results for Scenario 1 (Gaussian area effects) averaged over 30 small areas. The target parameters are the small area means and selected percentiles of the small area distributions.

Method	Target Parameters					
	10th	25th	50th	Mean	75th	90th
	Relative Bias (%)					
EBLUP/Naive	22.48	9.731	0.420	0.024	-4.708	-6.969
EBLUP/CD	0.373	0.205	0.079	-0.018	-0.073	-0.186
EBLUP/RKM	0.216	0.599	0.125	-0.018	-0.348	0.001
M-quantile/Naive	17.24	5.653	-2.641	-1.794	-7.021	-8.787
M-quantile/CD	0.373	0.176	0.028	-0.018	-0.086	-0.188
M-quantile/RKM	0.211	0.596	0.124	-0.018	-0.348	0.003
	Relative RMSE (%)					
EBLUP/Naive	22.56	9.99	2.86	1.97	4.93	7.03
EBLUP/CD	3.23	3.08	3.01	2.01	3.32	3.90
EBLUP/RKM	4.10	3.56	3.30	2.01	3.46	4.12
M-quantile/Naive	17.60	6.70	3.30	2.49	7.04	8.80
M-quantile/CD	3.23	3.09	3.11	2.01	3.48	3.89
M-quantile/RKM	4.11	3.56	3.36	2.01	3.46	4.12

Table 10.2: Model-based simulation results for Scenario 2 (Chi-squared area effects) averaged over 30 small areas. The target parameters are the small area means and selected percentiles of the small area distributions.

In order to evaluate the performance of the linearization-based MSE estimator (10.12) and the bootstrap MSE estimator (10.13), we carried out a further model-based simulation study. In this study we focussed on MSE estimation for the 25th, 50th and 75th percentiles using the bootstrap estimator (10.13), and for the mean using either the linearization-based estimator (10.12) or the bootstrap estimator (10.13). A total of 200 Monte Carlo simulations were carried out for each percentile and 100 Monte Carlo simulations for the mean, with the bootstrap MSE estimation implemented by generating a single bootstrap population at each Monte Carlo simulation and taking L = 500 bootstrap samples from this population. The bootstrap population was generated unconditionally, with bootstrap population values obtained by sampling from the smoothed residual distribution generated by the sample data obtained in each Monte Carlo simulation. Although it would have been theoretically preferable to have generated multiple bootstrap populations from each Monte Carlo sample, computing limitations restricted our investigation to B = 1. Since the estimates generated by the bootstrap procedure were then averaged over the 200 Monte Carlo simulations in our evaluation, this limitation is not as severe as it might appear to be, since the Monte Carlo simulations themselves serve as proxies for multiple bootstrap populations. Simulation results evaluating the resulting MSE estimators are set out in Tables 3 and 4 and in Figure 10.1. Focusing first on Table 10.3, we note that under both simulation scenarios, the linearization-based and the bootstrap MSE estimators tracked the true MSEs of the small area mean estimators very well, and provided coverage rates that were close to the nominal 95%.

Method	MSE estimator for the small area mean					
	Min	25th	50th	Mean	75th	Max
	Gaussian area effects					
True	0.271	0.331	0.411	0.419	0.481	0.783
Linearization	0.289	0.317	0.400	0.416	0.500	0.680
Bootstrap	0.282	0.319	0.401	0.418	0.504	0.715
Coverage Linearization	0.88	0.93	0.95	0.94	0.97	0.99
Coverage Bootstrap	0.88	0.94	0.96	0.96	0.97	0.99
	Chi-squared area effects					
True	0.344	0.453	0.549	0.589	0.736	1.051
Linearization	0.411	0.453	0.552	0.592	0.689	0.980
Bootstrap	0.398	0.444	0.559	0.589	0.706	1.003
Coverage Linearization	0.87	0.89	0.92	0.93	0.96	0.98
Coverage Bootstrap	0.92	0.95	0.96	0.96	0.97	1.00

Table 10.3: Across areas distribution of true (i.e. Monte Carlo) mean squared error and average over Monte Carlo simulations of estimated mean squared error and coverage rates of nominal 95% confidence intervals for the M-quantile/CD estimator 10.11. Estimated mean squared errors based on (10.13) using the smoothed unconditional approach (Bootstrap) or (10.12) (Linearization). Intervals were defined as the M-quantile/CD estimator (10.11) plus or minus twice its estimated standard error, calculated as the square root of (10.12) or (10.13).

Focusing next on Table 10.4 and Figure 10.1 we see that the bootstrap MSE estimator also performed well in terms of approximating the true MSEs of the small area quantile estimators. Again, coverage rates generated by 95% prediction intervals based on these MSE estimates were close to their nominal level.

MSE		Percentiles of across areas distribution					
		Min	25th	50th	Mean	75th	Max
		Gaussian area effects					
0.25 quantile	True	0.354	0.391	0.491	0.514	0.595	0.887
	Estimated	0.345	0.383	0.475	0.500	0.598	0.857
0.50 quantile	True	0.311	0.353	0.444	0.469	0.547	0.761
	Estimated	0.314	0.348	0.433	0.455	0.543	0.774
0.75 quantile	True	0.339	0.386	0.495	0.516	0.611	0.909
	Estimated	0.338	0.375	0.471	0.495	0.592	0.867
		Chi-squared area effects					
0.25 quantile	True	0.289	0.357	0.454	0.471	0.569	0.919
	Estimated	0.314	0.346	0.437	0.458	0.554	0.795
0.50 quantile	True	0.376	0.454	0.575	0.594	0.735	1.087
	Estimated	0.395	0.439	0.554	0.578	0.696	1.001
0.75 quantile	True	0.594	0.678	0.848	0.893	1.035	1.727
	Estimated	0.592	0.666	0.843	0.877	1.058	1.579

Table 10.4: Across areas distribution of the true (i.e. Monte Carlo) mean squared error and average over Monte Carlo simulations of estimated mean squared error for the CD estimates of 0.25, 0.50 and 0.75 quantiles from (10.10). Estimated mean squared error for quantiles is based on (10.13) using smoothed unconditional approach.

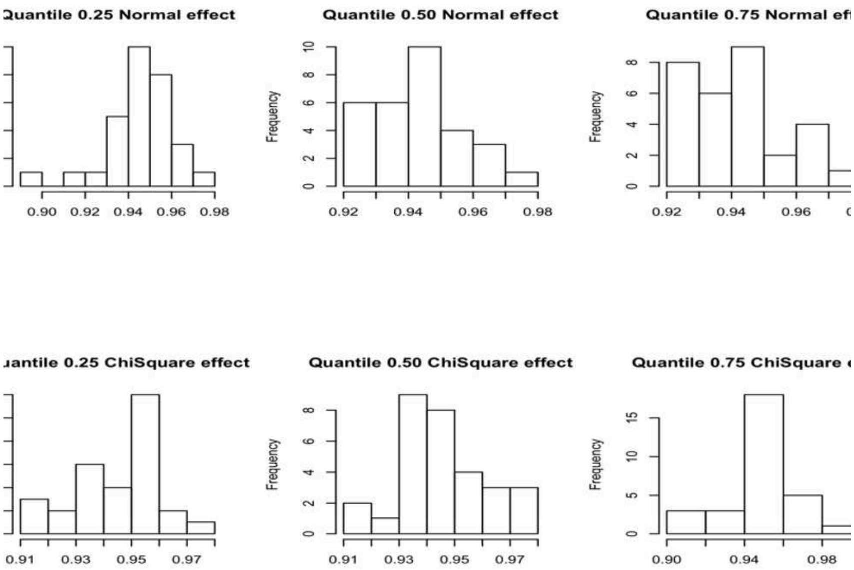


Figure 10.1: Distribution of area-specific coverage rates of nominal 95% confidence intervals for small area quantiles in the model-based simulations. Intervals were defined as the M-quantile/ CD estimator (10.10) plus or minus twice its estimated standard error, calculated as the square root of (10.13).

10.2 Small Area Models for Poverty Estimation

In this report we have already discussed small area estimation of averages and quantiles using the unit-level nested error regression model and the M-quantile small area model. Both models can be also utilized for estimating more complex statistics such as small area poverty indicators (Foster et al., 1984). Recently, Molina and Rao (2009) proposed the Empirical Best Prediction (EBP) approach to poverty estimation under the nested error regression model. Under this model, the EBP approach provides the best estimator of the target parameter. Nevertheless, with real data the assumptions of statistical models may hold only approximately and in fact, on many occasions there are significant departures from the model assumptions. An alternative approach to poverty estimation is based on the M-quantile small area model. The M-quantile method does not impose strong distributional assumptions and is outlier robust. Hence, the use of the M-quantile model for poverty estimation may protect us against departures from assumptions of the unit-level nested error regression model. In this section we contrast these two small area methodologies for poverty estimation both when the assumptions of the unit-level nested error regression model hold and when these assumptions are violated.

10.2.1 Definitions of poverty indicators

Although small area averages are widely used in small area applications, relying only on averages may not provide an very informative picture about the distribution of wealth in a small area. In economic applications for example, estimates of average income may not provide an accurate picture of the area wealth due to the high within area inequality. Here we focus exclusively on the estimation of two poverty indicators i.e. the incidence of poverty or *Head Count Ratio* F_0 and the *Poverty Gap* F_1 (see Foster et al. (1984)). Denoting by t the poverty line, the FGT poverty measures for a small area d are defined as

$$F_{\alpha d} = (\frac{t - y_{jd}}{t})^{\alpha} \mathrm{I}(y_{jd} \leqslant t). \tag{10.14}$$

Setting $\alpha = 0$ defines the *Head Count Ratio* whereas setting $\alpha = 1$ defines the *Poverty Gap*.

10.2.2 The M-quantile approach for poverty estimation

In this section we discuss estimation of the poverty indicators of interest under the M-quantile model. To start with, the target is to estimate F_{ad} using the M-quantile small area model

$$F_{\alpha d} = N_d^{-1}\Big[\sum_{j\in s_d} F_{\alpha d} + \sum_{j \in r_d} F_{\alpha d}\Big], \tag{10.15}$$

The question again is how to estimate the out of sample component in the expression above. This can be achieved using the ideas we described in Section 10.1.1 for estimating the small area distribution function under the M-quantile small area model. As we mentioned in Section 10.1.1, using the empirical distribution of the small area population distribution function may provide biased results especially when the aim is to estimate small area quantiles. Poverty estimation is a special case of quantile estimation since we are interested in estimating the number of individuals/households below a threshold. As a result one approach to estimating $F_{\alpha d}$ is by using a smearing-type estimator of the distribution function such as the Chambers-Dunstan estimator. In this case, an estimator $\hat{F}_{\alpha d}^{MQ}$ of $F_{\alpha d}^{MQ}$ is

$$\hat{F}_{ad} = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{k\in r_d} n_d^{-1} \sum_{j \in s_d} I(\hat{y}_k + (y_j - \hat{y}_j) \le t)\right\}$$

The above can be evaluated using the following procedure.

1 Fit the M-quantile small area model (1.1) using the raw $\mathbf{y}_s$ sample values and obtain estimates of β and q_d;

2 draw an out of sample vector using

$$y^*_{jdr} = \mathbf{x}_{jdr}\hat{\beta}(\hat{\theta}_d) + e^*_{jdr},$$

where e^*_{jdr} is a vector of size $N_d - n_d$ drawn from the Empirical Distribution Function (EDF) of the estimated M-quantile regression residuals or from a smooth version of this distribution and $\hat{\beta}$, $\hat{\theta}_d$ are obtained from the previous step;

3 repeat the process H times. Each time combine the sample data and out of sample data for estimating the target using

$$\hat{F}_{ad}^{MQ} = N_d^{-1}\Big[\sum_{j\in s_d} \mathrm{I}(y_j \leqslant t) + \sum_{j\in r_d} \mathrm{I}(y^*_j \leqslant t)\Big];$$

4 average the results over H simulations.

At this point, we should elaborate on some aspects of the approach used for estimating the poverty indicators under the M-quantile small area model. To start with, we must point out that one can use different approaches for drawing e^{*}_{jdr}. One can draw conditional (upon the small area) or unconditional residuals from the EDF or from a smoothed version of the EDF. The outlined approach for estimating the poverty indicators, although less parametric, is similar in spirit to the EBP approach proposed by Molina and Rao (2009). Note for example that y^{*}_{jdr} is generated using $\mathbf{x}_{jdr}\hat{\beta}(\hat{\theta}_d)$ i.e. from the conditional M-quantile model plus a draw from the empirical distribution of residuals. Under this approach $\hat{\theta}_d$ play the role of the area random effects in the M-quantile modelling framework. These area-specific M-quantile coefficients are not fixed i.e. they are estimated under the M-quantile model using fewer assumptions than the ones utilized by the unit-level nested error regression model for estimating the area random effects. Of course, when the assumptions of the random area effects model hold, the EBP approach of Molina and Rao (2009) offers the best predictor. However, when the assumptions of the random area effects model are not met, the M-quantile approach for estimating the incidence of poverty may offer a competitive alternative. A mean squared error of the M-quantile estimates of the incidence of poverty can be obtained using the non-parametric bootstrap approach described in Tzavidis et al. (2010). In the following section, we use two model-based simulation scenarios for contrasting the EBP and M-quantile approaches to poverty estimation. Under the first scenario the assumptions of unit-level nested error regression model are perfectly met and we assess the efficiency gains from using the EBP approach. Under the second scenario we generate population data under alternative parametric assumptions and we aim at assessing whether the less parametric M-quantile approach offers any efficiency gains in this case.

10.2.3 A Model-based Simulation

The model-based simulation scenario we consider is exactly the same as the one employed by Molina and Rao (2009). In particular, we simulated populations of size $N = 20000$ for 80 small areas with $N_d = 250$. The response variable for the population units y_{jd} was generated from a linear mixed model taking as auxiliary variables two dummies plus an intercept term. The values of these two dummies for the population units were generated from Bernoulli distributions with success probabilities increasing with the area index for X1 and constant for X2 and more specifically,

$$p_{1d} = 0.3 + \frac{0.5d}{80}, p_{2d} = 0.2,$$

and welfare variables are exponential functions of the responses y_{jd}. A set of sample indices s_d with $nd = 20$ was drawn independently in each area d using simple random sampling without replacement. The values of the auxiliary variables for the population units and the sample indices were kept fixed over the $H = 5000$ Monte Carlo simulations. The intercept and the regression coefficients associated with the two auxiliary variables used to generate populations were $\beta = (3, 0.03, .0.04)$ and the poverty line t was fixed as 0.6 times the median of the y. We considered two scenarios for generating the area-level and unit-level residuals.

- Scenario 1: Scenario 1: $u_d \sim N(0, 0.15^2)$ and $e_{jd} \sim N(0, 0.5^2)$
- Scenario 2: Mean-centered X^2 with $u_d \sim X^2(2)$ and $e_{jd} \sim X^2(4)$.

For each Monte-Carlo population we computed the true poverty indicators F_{ad}^{TR} and we also used three small area estimators (a) the direct estimator $\hat{F}_{ad}^{Dir}$, (b) the EBP $\hat{F}_{ad}^{EBP}$, and (c) the M-quantile $\hat{F}_{ad}^{MQ}$. The performance of these three estimators is evaluated using the following two measures

- Bias: $RB_d = \frac{1}{H}\sum_{h=1}^{H}\frac{(\hat{F}_{ad}-F_{ad}^{TR})}{F_{ad}^{TR}}$

- Root Mean Square Error $RRMSE_d = [\frac{1}{H}\sum_{h=1}^{H}(\hat{F}_{ad}-F_{ad}^{TR})^2]^{\frac{1}{2}}$

The results of the model-based simulations are summarised in Figures 1.2-1.5. As expected when the assumptions of the unit-level nested error regression model hold, the EBP approach offers estimates with the smallest MSE. The M-quantile estimates have a bit larger MSE and the direct estimates are the most inefficient. On the other hand, when the Gaussian assumptions of the unit-level nested error regression model do not hold, we notice that the M-quantile approach offers small area estimates of poverty that are more efficient while the EBP estimates are more efficient than the direct ones. These results indicate that the less parametric M-quantile approach can protect us against model mispecification. However, if we are certain that the model assumptions are met, the EBP approach will always offer the most efficient method to poverty estimation.

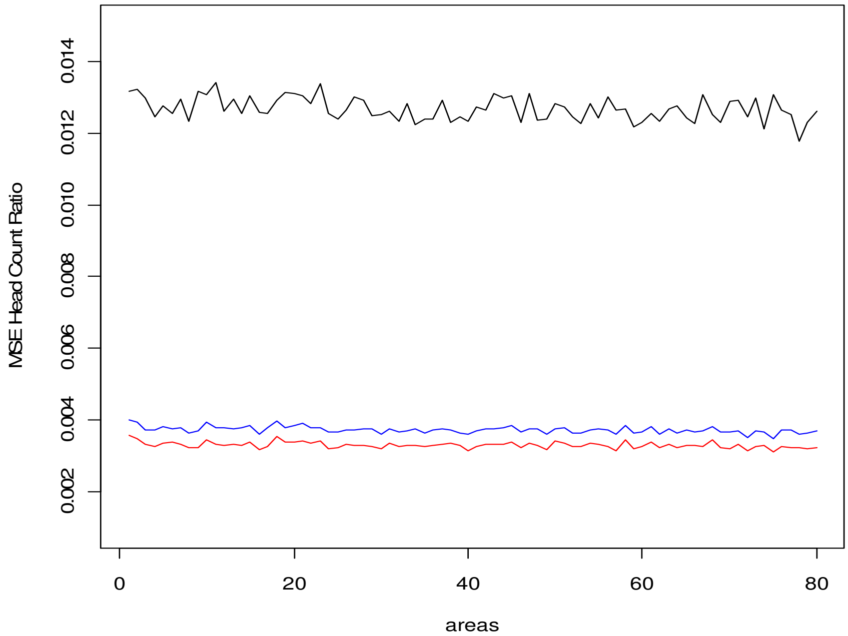


Figure 10.2: Model based simulations: MSE of the EBP (red), direct (black), and MQ (blue) estimators of HCR, for each area when the normality assumptions hold.

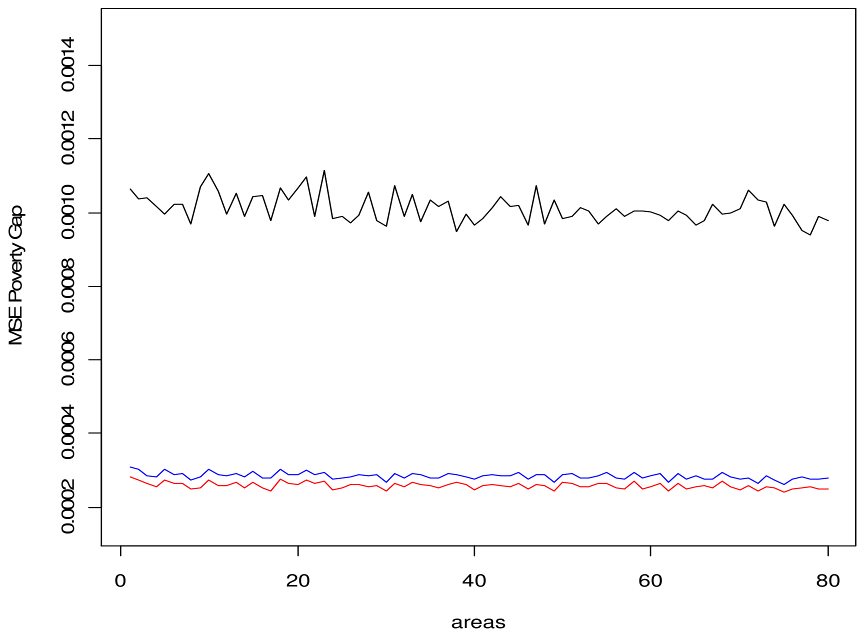


Figure 10.3: Model based simulations: MSE of the EBP (red), direct (black), and MQ (blue) estimators of PG, for each area when the normality assumptions hold.

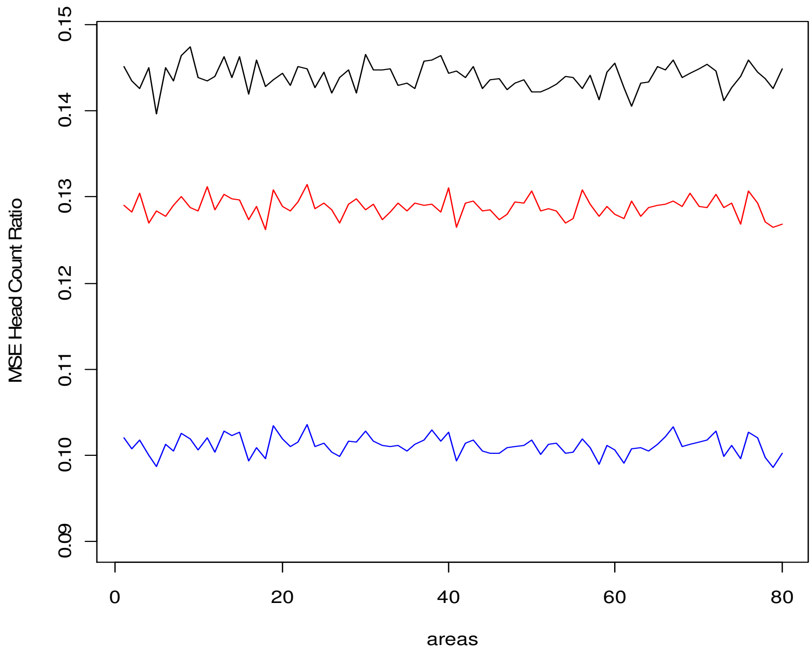


Figure 10.4: Model based simulations: MSE of the EBP (red), direct (black), and MQ (blue) estimators of HCR, for each area when the chi-square assumptions hold.

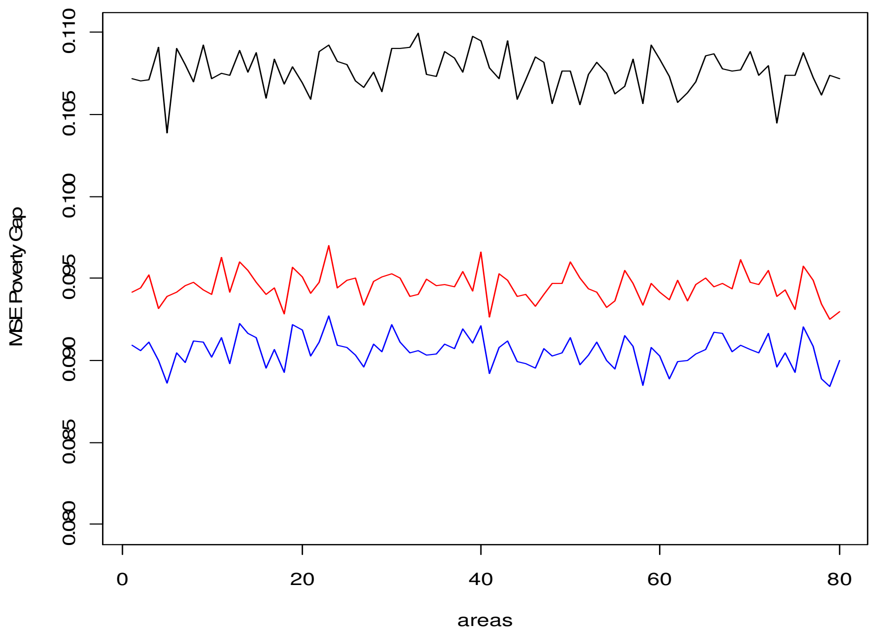


Figure 10.5: Model based simulations: MSE of the EBP (red), direct (black), and MQ (blue) estimators of PG, for each area when the chi-square assumptions hold.

10.2.4 A Design-based Simulation

The aim of this section is to empirically contrast the two methodologies using a design-based simulation that utilizes real data from the 2007 European Survey on Income and Living Conditions (EU-SILC) in Italy. Our target is the estimation of the incidence of poverty for 29 Italian Provinces in three Regions: Lombardia (Northern Italy), Toscana (Central Italy) and Campania (Southern Italy). Data on the household equivalised income, on household characteristics (size of the household in square meters) and on individual characteristics of the head of the household (gender, education, marital status and employment) are available.

The synthetic population data on which this simulation is based is generated by nonparametrically bootstrapping (within each of the 29 target small areas) the EU-SILC original sample dataset. This synthetic population was kept fixed and five hundred within province simple random samples of size equal to the sample size of each province in the original dataset were selected independently. Estimated values of the incidence of poverty at province level were obtained using the MQ (on the raw equivalised income) and the EBP (on the logarithm of equivalised income) estimators described in the previouos sections. The simulation results are set out in Figures 10.6, 10.7, 10.8 which show the relative bias, the square root of the variance, and the root mean squared error for the estimation of the HCR. This assists us in understanding how the different components i.e. bias and variance contribute to the mean squared error. To start with, we note that the M-quantile-based poverty estimation method has smaller RMSE in most provinces (Figure 10.8). To explain this result we first focus on the variance results. It is clear that the EBP method has lower variance in most provinces (Figure 10.7), which we may expect given that the EBP is based on a random effects model. This result indicates that the worse performance of the EBP in terms of RMSE must be due to bias. Indeed, the bias of the EBP is higher than the bias of the M-quantile estimates for most provinces (Figure 10.6).

At this point it is important to remind that the synthetic population of this design-based simulation is not generated under a model but by non-parametrically bootstrapping the original sample data. One may argue that the nonparametric bootstrap for creating the synthetic population creates an over-representation of influential points in this population that affects the EBP approach and favors the M-quantile approach. For this reason, we replicated the design-based simulation this time generating the synthetic population by non-parametrically bootstrapping the original sample using also the survey weights. The results from this second design-based simulation are not reported here but the conclusions about the performance of the two approaches to poverty estimation remain the same as above. One explanation about higher bias in the EBP estimates is that this may be due to the effect of the simple exponential backtransformation. However, we are not convinced that the backtransformation is the source of the problem. Looking at the fit of the random effects model to the sample data, we noted that the Gaussian assumptions utilized by the EBP method are not met even when the log-transformed income is used. Hence generating area random effects and individual errors under these assumptionsmay not be realistic in this case. On the other hand, utilizing the empirical distribution of the residuals, as in the case of the M-quantile approach, may protect us against the misspecification of the parametric assumptions.

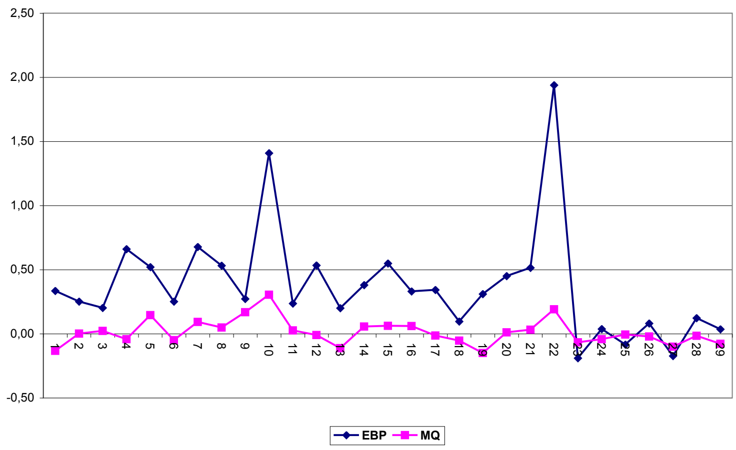


Figure 10.6: Design based simulations: relative bias of the EBP and MQ estimators of the HCR, for each area.

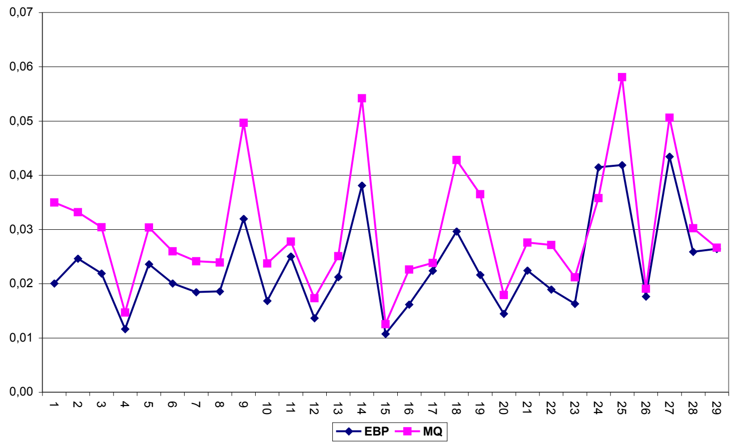


Figure 10.7: Design based simulations: square root of the variance of the EBP and MQ estimators of the HCR, for each area.

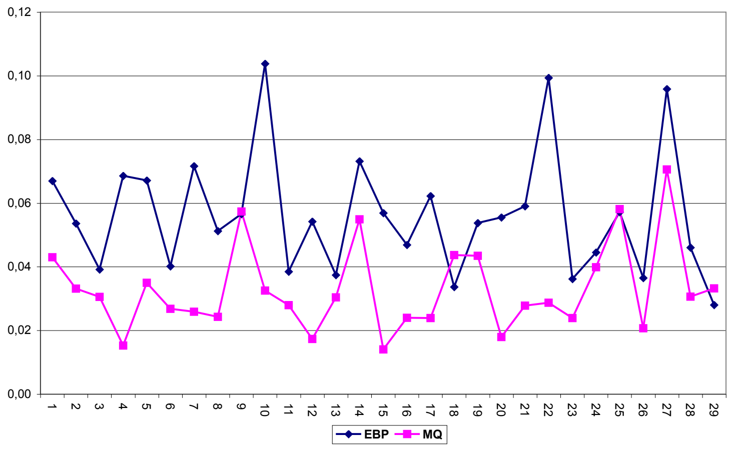


Figure 10.8: Design based simulations: root MSE of the EBP and MQ estimators of the HCR, for each area.

10.2.5 Alternative measures for poverty: fuzzy indicators at a small area level with M-quantile models

In this section we propose estimators for the fuzzy monetary and fuzzy supplementary indicators at a small area level based on the M-quantile model. Fuzzy approach considers poverty as a matter of degree rather than an attribute that is simply present or absent for individuals in the population of Betti et al.(2009).

Fuzzy indicators have been made such that they vary from 0 to 1, where 0 indicates the richest person in the population while 1 indicates the poorest.

The fuzzy monetary indicator is based on the equivalised income, E. For small area d is defined as:

$$FM_d = N_d^{-1} \sum_{j=1}^{N_d} FM_j$$

where FM_j is the fuzzy monetary index for the jth unit in the population:

$$FM_j = \left[(N_d - 1)^{-1} \sum_{k=1}^{N_d} I(E_k > E_j) \right]^{\alpha - 1} \left[\frac{\sum_{k=1}^{N_d} E_k I(E_k > E_j)}{\sum_{k=1}^{N_d} E_k} \right].$$

The parameter α is arbitrary, but Cheli and Betti (1999) have chosen α so that the mean of FM_j for the whole population is equal to the head count ratio computed for the official poverty line.

An estimator of FM_d under the M-quantile model is given by

$$\widehat{FM}_d^{MQ} = N_d^{-1}\left[\sum_{j\in s_d} FM_j + \sum_{j\in r_d} \widehat{FM}_j^{MQ}\right]. \tag{10.16}$$

Under the M-quantile model an empirical approach for estimating equation (10.16) is implemented using the same Monte-Carlo approximation described in section 10.2.2.

1. Fit the M-quantile model (10.1) using the raw E sample values and obtain estimates of $\beta(\theta_d)$;

2. draw an out of sample vector using

$$E_j^* = \mathbf{x}_j\hat{\beta}(\hat{\theta}_d) + e_j^* \quad j \in r_d,$$

   where $e_j^*, j \in r_d$ is drawn from the EDF of the estimated M-quantile regression residuals and $\hat{\beta}(\hat{\theta}_d)$ is obtained from the previous step;

3. repeat the process H times. Each time combine the sample data and the out of sample data for estimating the target using

$$\widehat{FM}_d^{MQ} = N_d^{-1}\left[\sum_{j\in s_d} FM_j + \sum_{j\in r_d} \widehat{FM}_j^{MQ}\right],$$

   where $\widehat{FM}_j^{MQ}$ is estimated using the observed and the predicted equivalised incomes $E_d = \{E_j, j \in s_d \cup E_j^*, j \in r_d\}$.

4. average the results over L simulations.

The fuzzy supplementary indicator put together diverse indicators of deprivation, such as housing conditions, possession of durable goods, perception of hardship, expectations, norms and values.

To quantify and put together diverse indicators several steps are necessary (see Betti et al. (2009)):

1. Identification of items;

2. transformation of the items into the $[0,1]$ interval;

3. exploratory and confirmatory factor analysis;

4. calculation of weights within each dimension;

5. calculation of scores for each dimension;

6. calculation of an overall score and the parameter α;

7. construction of the fuzzy deprivation measure in each dimension (and overall).

The steps 1 to 3 are used to identify different dimensions of the poverty where each dimension is composed of a given number of items.

Let K be the number of dimensions and k_h the number of items ($i = 1, \dots, k_h$) within the hth dimension ($h = 1, \dots, K$). The weight, w, for a given dimension h formed by k_h items and a given item z_i is

$$w_{hi} \propto \left[\frac{\sigma_{h,z_i}}{1-\bar{z}_i}\right]\left[\left(1+\sum_{i=1}^{k_h} r_{h,z_i;h,z_i'} I(r_{h,z_i;h,z_i'} < r^*_{h,z_i;h,z_i'})\right)^{-1}\left(1+\sum_{i=1}^{k_h} r_{h,z_i;h,z_i'} I(r_{h,z_i;h,z_i'} > r^*_{h,z_i;h,z_i'})\right)^{-1}\right],$$

where $\bar{z}_i$ and σ_{h,z_i} are the mean and the standard deviation of the ith item respectively. $r_{h,z_i;h,z_i'}$ is correlation coefficient between deprivation indicators corresponding to item z_i and the other items present in the hth dimension, $r^*_{h,z_i;h,z_i'}$ is the critical value of the correlation coefficient (see Betti et al. (2009)). Once weights are calculated for each items in a given dimension they are scaled in such a way that they sum to 1 within the dimension.

The score for the hth dimension for the jth individual is then computed as

$$s_{h,j} = \sum_{i=1}^{k_h} w_{hi}\frac{z_{hi,j}}{w_{hi}},$$

where $z_{hi,j}$ is the value of the ith item in the hth dimension for the individual j.

An overall score for the jth individual is calculated as the following unweighted mean:

$$s_j = K^{-1}\sum_{h=1}^{K} s_{h,j}.$$

Finally the fuzzy supplementary for the jth individual over all dimensions is computed as:

$$FS_j = \left[\frac{\sum_{g=j+1}^{N} I(s_g > s_j)}{\sum_{g=2}^{N} I(s_g > s_j)}\right]^{\alpha-1}\left[\frac{\sum_{g=j+1}^{N} s_g I(s_g > s_j)}{\sum_{g=2}^{N} s_g I(s_g > s_j)}\right],$$

where α is computed as for the fuzzy monetary indicator. Using the score of a given dimension we can compute fuzzy supplementary indicators for each dimension.

At a small area level we propose to compute scores for the sampled units as mentioned before and then use the M-quantile model (or other proper models) to predict scores for the non sampled units in the population. Once we have the scores for each individual, where s are observed and r are predicted[1], we can estimate the fuzzy supplementary indicator for each small areas using the same Monte-Carlo approximation described for the fuzzy monetary index.

We remand to the final WP1 deliverable for a deepened discussion about fuzzy sets. Here, we want just to provide a first approach to estimate fuzzy indicators at a small area level. The used approach is similar in spirit to the EBP approach proposed by Molina and Rao (2009).

[1]Note that in this statement s indicates the set of the sampled units and not the score variable.

Bias$*10^5$	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
EBP	−1.00	1.69	2.57	2.59	3.70	5.40
MQ	−17.45	−14.93	−13.39	−12.84	−11.14	−5.07
MSE$*10^4$	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
EBP	5.23	5.44	5.50	5.52	5.61	5.89
MQ	5.21	5.43	5.52	5.52	5.61	5.85

Table 10.5: Distribution over areas of the bias $(*10^5)$ and mean squared error $(*10^4)$ of the EBP and MQ estimators of the small area fuzzy monetary index

10.2.6 Model based simulation for fuzzy monetary indicator at a small area level

The model-based simulation has been carried out using the same scenario as the one employed by Molina and Rao (2009). This scenario has been already described in section 10.2.3 as scenario 1.

We evaluate the performance of the fuzzy monetary indicator using the following two measures:

- Bias:

$$B_d = L^{-1}\sum_{l=1}^{L}\widehat{FM}_d - FM_d^{TR};$$

- Mean Squared Error:

$$MSE_d = L^{-1}\sum_{l=1}^{L}(\widehat{FM}_d - FM_d^{TR})^2;$$

where FM_d^{TR} is the true value of the fuzzy monetary indicator in area d, $\widehat{FM}_d$ is an estimator of the fuzzy monetary chosen between the EBP estimator and the M-quantile estimator. The EBP estimator is based on the EBP approach described in chapter 2.

Results for EBP and MQ estimators are summarised in Table 1.5. In Figures 10.9 and 10.10 we show the performance in each area of the EBP and MQ estimators of the fuzzy indicator.

The EBP shows the best performance in terms of bias while the mean squared error is similar for EBP and MQ estimator. The MQ estimator shows a bigger bias than EBP and this is probably due to the fact that we used a scenario that fit perfectly with the parametric assumptions of the EBP so that MQ estimator can't reach the same level of precision.

Fuzzy supplementary indicators are thought to resume a large set of variables that are proxies for poverty and social exclusion. Typically one have to identify a given number of dimensions with factor analysis, where each dimension is build by a set of items. Given that, since in our scenario we considered only two dummy variables, we decided to not include the fuzzy supplementary indicator in the simulation study: we believe that it does not have sense to compute this indicator with a so poor set of variables.

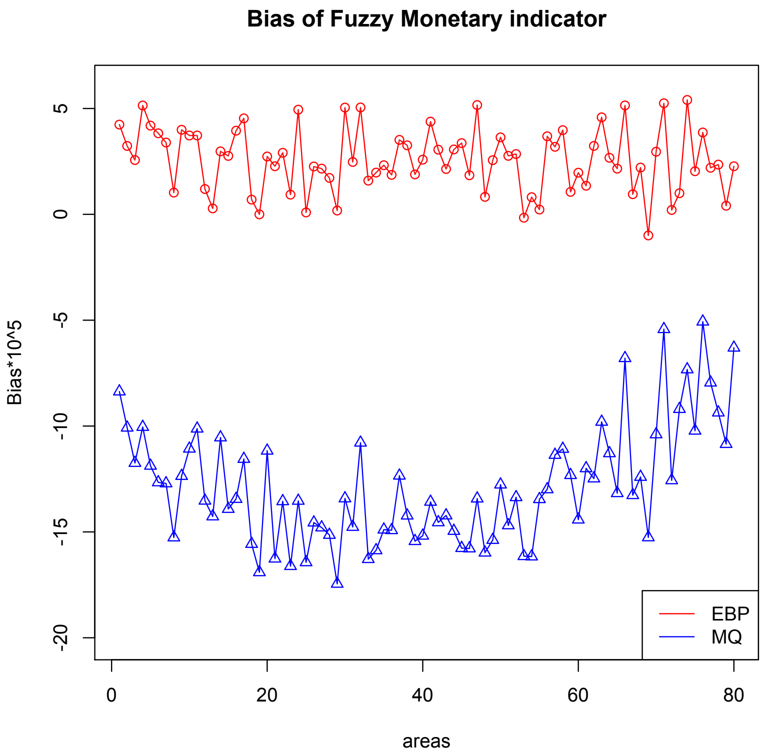


Figure 10.9: Model based simulations: bias of the EBP and MQ estimators of the fuzzy monetary indicator, for each area.

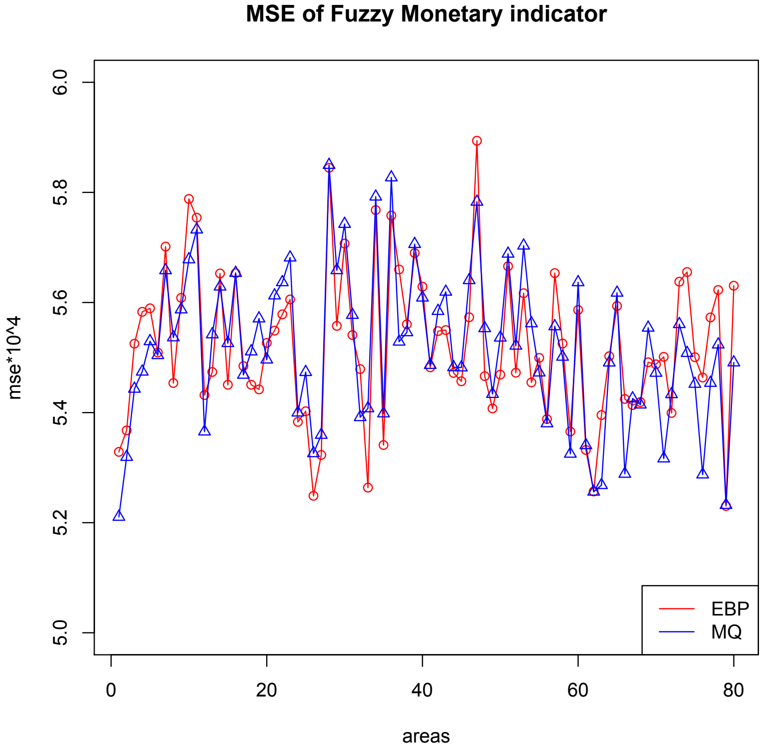


Figure 10.10: Model based simulations: mean squared error of the EBP and MQ estimators of the fuzzy monetary indicator, for each area.

10.3 Nonparametric M-quantile regression models in small area estimation

M-quantile models do not depend on strong distributional assumptions, but they assume that the quantiles of the distribution are some known parametric function of the covariates. When the functional form of the relationship between the q-th M-quantile and the covariates deviates from the assumed one, the traditional M-quantile regression can lead to biased estimates of the β coefficients. Pratesi et al. (2008) and Salvati et al. (2010b) have extended this approach to the M-quantile method for the estimation of the small area parameters using a nonparametric specification of the conditional M-quantile of the response variable given the covariates. When the functional form of the relationship between the q-th M-quantile and the covariates deviates from the assumed one, the traditional M-quantile regression can lead to biased estimators of the small area parameters. Using p-splines for M-quantile regression, beyond having the properties of M-quantile models, allows for dealing with an undefined functional relationship that can be estimated from the data. When the relationship between the q-th M-quantile and the covariates is not linear, a p-splines M-quantile regression model may have significant advantages compared to the linear M-quantile model.

Let us consider only smoothing with one covariate x_1, a nonparametric model for the qth quantile can be written as $Q_q(x_1,\psi) = \tilde{m}_{\psi,q}(x_1)$, where the function $\tilde{m}_{\psi,q}(\cdot)$ is unknown and, in the smoothing context, usually assumed to be continuous and differentiable. Here, we will assume that it can be approximated sufficiently well by the following function

$$m_{\psi,q}[x_1;\beta_\psi(q),\gamma_\psi(q)] = \beta_{0\psi}(q) + \beta_{1\psi}(q)x_1 + \ldots + \beta_{p\psi}(q)x_1^p + \sum_{k=1}^{K}\gamma_{k\psi}(q)(x_1-\kappa_k)_+^p, \qquad (10.17)$$

where p is the degree of the spline, $(t)_+^p = t^p$ if $t > 0$ and 0 otherwise, κ_k for $k = 1,\ldots,K$ is a set of fixed knots, $\beta_\psi(q) = (\beta_{0\psi}(q),\beta_{1\psi}(q),\ldots,\beta_{p\psi}(q))^t$ is the coefficient vector of the parametric portion of the model and $\beta\gamma_\psi(q) = (\gamma_{1\psi}(q),\ldots,\gamma_{K\psi}(q))^t$ is the coefficient vector for the spline one. The latter portion of the model allows for handling nonlinearities in the structure of the relationship. The spline model (10.17) uses a truncated polynomial spline basis to approximate the function $\tilde{m}_{\psi,q}(\cdot)$. Other bases can be used; in particular radial basis functions can be used to handle bivariate smoothing. More details on bases and knots choice can be found in Ruppert et al. (2003).

The influence of the knots is limited by putting a constraint on the size of the spline coefficients: typically $\sum_{k=1}^{K}\gamma_{k\psi}^2(q)$ is bounded by some constant, while the parametric coefficients $\beta_\psi(q)$ are left unconstrained. Therefore, by dropping the area subscript d for ease of notation, estimation can be accommodated by mimicking penalization of an objective function and solving the following set of estimating equations

$$\sum_{j=1}^{n}\psi_q(y_j - \mathbf{x}_j\beta_\psi(q) - \mathbf{z}_j\gamma_\psi(q))(\mathbf{x}_j,\mathbf{z}_j)\text{trace} + \lambda\begin{bmatrix}\mathbf{0}_{(1+p)}\\ \gamma_\psi(q)\end{bmatrix} = \mathbf{0}_{(1+p+K)}, \qquad (10.18)$$

assuming that

$$\psi_q(r_{jq\psi}) = 2\psi\{s^{-1}r_{jq\psi}\}\{(1-q)I(r_{jq\psi}\le 0) + qI(r_{jq\psi} > 0)\}$$

where $r_{jq\psi} = y_j - \mathbf{x}_j\beta_\psi(q) - \mathbf{z}_j\gamma_\psi(q)$, s is a robust estimate of scale, e.g. the MAD estimate $s = \text{median}|r_{jq\psi}|/0.6745$, $\mathbf{x}_j$ here is the j-th row of the $n\times(1+p)$ matrix

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}^p \end{bmatrix},$$

while $\mathbf{z}_j$ is the j-th row of the $n\times K$ matrix

$$\mathbf{Z} = \begin{bmatrix} (x_{11}-\kappa_1)_+^p & \cdots & (x_{11}-\kappa_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_{1n}-\kappa_1)_+^p & \cdots & (x_{1n}-\kappa_K)_+^p \end{bmatrix},$$

and λ is a Lagrange multiplier that controls the level of smoothness of the resulting fit. An iterative solution is needed here to obtain estimates $\hat{\beta}_\psi(q)$ and $\hat{\gamma}_\psi(q)$. Consider the Huber proposal 2 influence function (see Huber(1981)) an algorithm based on iteratively reweighted penalized least squares is proposed in Pratesi et al. (2009) to effectively compute the parameter estimates.

Once parameter estimates are obtained, $\hat{m}_{\psi,q}[x_1] = m_{\psi,q}[x_1;\hat{\beta}_\psi(q),\hat{\gamma}_\psi(q)]$ can be computed as an estimate for $Q_q(x_1,\psi)$. The approximation ability of this final estimate will heavily depend on the value of the smoothing parameter λ. Generalized Cross Validation (GCV) has been usefully applied in the context of smoothing splines (see Craven and Wahba (1979)) and is used also in Pratesi et al. (2009).

Extension to bivariate smoothing can be handled by assuming $Q_q(x_1,x_2,\psi) = \tilde{m}_{\psi,q}(x_1,x_2)$. This is of central interest in a number of application areas when referenced responses need to be converted to maps, as in environment and poverty mapping. In particular, the following model is assumed at quantile q for unit i:

$$m_{\psi,q}[x_{1j},x_{2j};\beta_\psi(q),\gamma_\psi(q)] = \beta_{0\psi}(q)+\beta_{1\psi}(q)x_{1j}+\beta_{2\psi}(q)x_{2j}+\mathbf{z}_i\gamma_\psi(q). \tag{10.19}$$

Here $\mathbf{z}_j$ is the j-th row of the following $n\times K$ matrix

$$\mathbf{Z} = [C(\tilde{\mathbf{x}}_j-\mathbf{k}_k)]_{\substack{1\leqslant j\leqslant n\\1\leqslant k\leqslant K}}[C(\mathbf{k}_k-\mathbf{k}_{k'})]^{-1/2}_{1\leqslant k,k'\leqslant K}, \tag{10.20}$$

where $C(\beta t) = ||\beta t||^2\log||\beta t||$, $\tilde{\mathbf{x}}_j = (x_{1j},x_{2j})$ and $\mathbf{k}_k$, $k=1,\ldots,K$ are knots. See Pratesi et al. (2009) for details on this. Here, it is enough to note that the estimation procedure can again be pursued with (10.18) where $\mathbf{x}_j = (1,\tilde{\mathbf{x}}_j)$.

The choice of knots in two dimensions is more challenging than in one. Two solutions suggested in literature that provide a subset of observations nicely scattered to cover the domain are *space filling designs* (see Nychka and Saltzman (1998)) and the *clara* algorithm. The first one is based on the maximal separation principle of K points among the unique $\tilde{\mathbf{x}}_j$ and is implemented in the `fields` package of the `R` language. The second one is based on clustering and selects K representative objects out of n; it is implemented in the package `cluster` of `R`.

It should be noted, then, that the estimating equations in (10.18) can be used to handle univariate smoothing and bivariate smoothing by suitably changing the parametric and the spline part of the model, i.e. once the **X** and the **Z** matrices are set up. Finally, other continuous or categorical variables can be easily inserted parametrically in the model by adding columns to the **X** matrix. This allows for semiparametric modeling, as intended in Ruppert et al. (2003), to be inherited and applied to M-quantile regression.

10.3.1 Small area estimator of the mean and of the quantiles

Salvati et al. (2010b) have applied the P-splines M-quantile regression to the estimation of a small area mean as follows. The first step is to estimate the M-quantile coefficients q_{jd} as illustrated in paragraph 10.1 for the linear case treated in Chambers and Tzavidis (2006). Recall that the M-quantile coefficient q_{jd} of unit j in area d is the value q_{jd} such that $Q_{q_{jd}}(x_{1jd},\psi) = y_{jd}$. The unit level coefficients are estimated by defining a fine grid of values on the interval $(0,1)$ and using the sample data to fit the p-splines M-quantile regression functions at each value q on this grid. If a data point lies exactly on the q-th fitted curve, then the coefficient of the corresponding sample unit is equal to q. Otherwise, to obtain q_{jd}, a linear interpolation over the grid is used. An estimate of the mean quantile for area d is obtained by taking the corresponding average value of the sample M-quantile coefficient of each unit in area d. The small area estimator of the mean may be taken as:

$$\hat{\bar{Y}}_d = \frac{1}{N_d}\Big\{\sum_{j\in s_d} y_{jd}+\sum_{j\in r_d}\hat{y}_{jd}\Big\}, \tag{10.21}$$

where the unobserved value for population unit $j\in r_d$ is predicted using

$$\hat{y}_{jd} = \mathbf{x}_{jd}\hat{\beta}_\psi(\hat{\theta}_d)+\mathbf{z}_{jd}\hat{\gamma}_\psi(\hat{\theta}_d),$$

where $\hat{\beta}_\psi(\hat{\theta}_d)$ and $\hat{\gamma}_\psi(\hat{\theta}_d)$ are the coefficient vectors of the parametric and spline portion, respectively, of the fitted p-splines M-quantile regression function at $\hat{\theta}_d$.

However, the estimator of the small area mean can be biased for small areas containing outliers. This has already been noted in Tzavidis and Chambers (2006) for the estimator under the a linear M-quantile regression model. They propose an adjustment for bias based on the Chambers and Dunstan (1986) estimator of the small area distribution function. This adjustment can be used also in case of p-splines M-quantile regression models. The bias-adjusted estimator for the mean is given by

$$\hat{\bar{Y}}_d^{NPMQ/CD} = \frac{1}{N_d}\Big\{\sum_{j\in U_d}\hat{y}_{jd} + \frac{N_d}{n_d}\sum_{j\in s_d}(y_{jd}-\hat{y}_{jd})\Big\}, \qquad (10.22)$$

where $\hat{y}_{jd}$ denotes the predicted values for the population units in s_d and in U_d. This estimator will be here denoted with NPMQ when using the p-splines M-quantile regression model, and with MQ when using the linear one.

Due to the bias correction in (10.22), this predictor will have higher variability and so should only be used when the estimator (10.21) is expected to have substantial bias, e.g. when there are large outlying data points. An alternative approach to dealing with the bias-variance trade off in (10.22) in such a situation is to limit the variability of the bias correction term in (10.22) by using robust (huberized) residuals instead of raw residuals. In particular,

$$\hat{\bar{Y}}_d^{NPMQ/Rob} = \frac{1}{N_d}\Big\{\sum_{j\in s_d} y_{jd} + \sum_{j\in r_d}\hat{y}_{jd} + \frac{N_d-n_d}{n_d}\sum_{j\in s_d}\nu_d\psi\Big(\frac{y_{jd}-\hat{y}_{jd}}{\nu_d}\Big)\Big\} \qquad (10.23)$$

where ν_d is a robust estimate of scale for area d (see Tzavidis and Chambers (2007)).

Using the nonparametric M-quantile predictor for the non sampled units we can define a model unbiased estimator of the small area distribution function (10.6):

$$\hat{F}_d^{NPQM/CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j\le t) + \sum_{k\in r_d} n_d^{-1}\sum_{j\in s_d} I(\mathbf{x}_{kd}\hat{\beta}_\psi(\hat{\theta}_d) - \mathbf{z}_{jd}\hat{\gamma}_\psi(\hat{\theta}_d) + (y_j - \mathbf{x}_{jd}\hat{\beta}_\psi(\hat{\theta}_d) + \mathbf{z}_{jd}\hat{\gamma}_\psi(\hat{\theta}_d)) \le t)\right\}. \qquad (10.24)$$

Similarly to M-quantile small area models, the qth quantile $\hat{\mu}_{qd}$ of the distribution of y in area d is straightforwardly estimated by the solution to

$$\int_{-\infty}^{\hat{\mu}_{qd}} d\hat{F}_d^{NPQM}(t) = q. \qquad (10.25)$$

10.3.2 Mean squared error estimation

Salvati et al. (2010b) also propose an estimator of the MSE of the small area mean. For fixed q and λ, the $\hat{\bar{Y}}_j$ in (10.23) can be written as the following linear combination of the observed y_{jd} plus an additional part due to the huberized residuals. In particular,

$$\hat{\bar{Y}}_d^{NPMQ/Rob} = \frac{1}{N_d}\sum_{j\in s} w_{jd} y_{jd}, \qquad (10.26)$$

where the weights $\mathbf{w}_d = (w_{1d}, \ldots, w_{nd})^T$ are given by

$$\mathbf{w}_d = \left\{1 + \frac{N_d - n_d}{n_d} b_{jd}\right\} \mathbf{1}_{s_d} +$$

$$+\mathbf{W}(\hat{\theta}_d)[\mathbf{X}, \mathbf{Z}]\Big([\mathbf{X}, \mathbf{Z}]\text{trace}\mathbf{W}(\hat{\theta}_d)[\mathbf{X}, \mathbf{Z}] + \lambda \mathbf{G}\Big)^{-1}\left(\mathbf{T}_{r_d} - \frac{N_d - n_d}{n_d}\mathbf{T}_{s_d}\right) \quad (10.27)$$

with $b_{jd} = \psi\Big(\frac{y_{jd}-\hat{y}_{jd}}{\nu_d}\Big)/\Big(\frac{y_{jd}-\hat{y}_{jd}}{\nu_d}\Big)$, $\mathbf{1}_{s_d}$ the n-vector with j^{th} component equal to one whenever the corresponding sample unit is in area j and to zero otherwise, $\mathbf{W}(\hat{\theta}_d)$ a diagonal $n \times n$ matrix that contains the final set of weights produced by the iteratively reweighted penalized least squares algorithm used to estimate the regression coefficients, $\mathbf{G} = \text{diag}\{\mathbf{0}_{1+p}, \mathbf{1}_K\}$ with $1+p$ the number of columns of $\mathbf{X}$ and K the number of columns of $\mathbf{Z}$, and with $\mathbf{T}_{r_d}$ and $\mathbf{T}_{s_d}$ the totals of the covariates for the non-sampled and the sampled units in area d, respectively. Note that $\mathbf{T}_{s_d} = \sum_{j \in s_d} [\mathbf{x}_{jd}\ \mathbf{z}_{jd}]^T b_{jd}$.

The weights derived from (10.27) are treated as fixed and a "plug in" estimator of the mean squared error of estimator (10.26) can be proposed by using standard methods for robust estimation of the variance of unbiased weighted linear estimators (see Royall and Cumberland (1978)) and by following the results due to Chambers and Tzavidis (2006). The prediction variance of (10.26) can be approximated by

$$\text{var}(\hat{\bar{Y}}_d^{NPMQ/Rob} - \bar{Y}_d) \approx \frac{1}{N_d^2}\Big[\sum_{j \in s_d}\Big\{d_{jd}^2 + \frac{N_d - n_d}{n_d - 1}\Big\}\text{var}(y_{jd}) + \sum_{j \in s \setminus s_d} b_{jd}^2 \text{var}(y_{jd})\Big] \quad (10.28)$$

with $b_{jd} = w_{jd} - 1$ if $j \in s_d$ and $b_{jd} = w_{jd}$ otherwise, and $s \setminus s_d$ the set of sampled units outside area d. Following the area level residual approach of Tzavidis and Chambers (2006), we can interpret $\text{var}(y_{jd})$ conditionally to the specific area d from which y_d is drawn and hence replace $\text{var}(y_{jd})$ in (10.28) by $(y_{jd} - \hat{y}_{jd})^2$. Salvati et al. (2010b) develop a robust estimator of the mean squared error of (10.26) that is given by

$$\widehat{\text{var}}(\hat{\bar{Y}}_d^{NPMQ/Rob}) = \frac{1}{N_d^2}\left[\sum_{j \in s_d}\left\{b_{jd}^2 + \frac{N_d - n_d}{n_d - 1}\right\}(y_{jd} - \hat{y}_{jd})^2 + \sum_{j \in s \setminus s_d} b_{jd}^2 (y_{jd} - \hat{y}_{jd})^2\right]. \quad (10.29)$$

Since the bias-adjusted nonparametric M-quantile estimator is an approximately unbiased estimator of the small area mean, the squared bias will not impact significantly the mean squared error estimator. The main limitation of the MSE estimator is that it does not account for the variability introduced in estimating the area specific q's and λ. We note also that we can obtain an estimate only for areas where there are at least two sampled units. Details on the property of the MSE estimator can be found in Tzavidis et al. (2010) and Salvati et al. (2010a).

10.3.3 Simulations for nonparametric M-quantile models

In this section we use simulation studies to illustrate the finite-sample performance of the small area mean estimator based on p-splines M-quantile regression – NPMQ. It is compared with the estimator computed by standard linear M-quantile regression – MQ – and with the Empirical Best Linear Unbiased Prediction estimators based on Battese et al. (1988) model – EBLUP – and on nonparametric regression model by Opsomer et al. (2008) – NPEBLUP. We carried out one simulation study where the properties of the estimators of the small area mean and of its MSE have been assessed by Monte Carlo experiments using models with a single covariate. These results are reported in Salvati et al. (2010b) and will be published in the *Journal of Statistical Computation and Simulation*. We also consider and compare the estimators of the small area quantiles, namely of quantiles 0.25, 0.5 and 0.75. In this case

Given the number of small areas $d = 30$, three synthetic populations of size $N = 10,550$ are generated under the random intercepts model

$$y_{jd} = m(x_{jd}) + \gamma_d + \varepsilon_{jd}$$

with x drawn from a Uniform distribution $[0,1]$ and area effects γ_d were independently drawn from $N(0,0.04)$. The true underlying relationship between the covariate x and the expected value of the response variable y $E(y|x) = m(x)$ were generated by the following models:

Linear. $m(x) = 3 + 2(x - 0.5)$: it represents a situation in which MQ and EBLUP are based on a good representation of the true model and NPMQ and NPEBLUP may be too complex;

Cycle. $m(x) = 2\sin(2\pi x)$: it defines an increasingly more complicated structure of the relationship between y and x;

Jump. $m(x) = 1 + 2(x - 0.5)I(x \leqslant 0.5) + 0.5I(x > 0.5)$: it is a discontinuous function for which all estimators are based on a misspecified model.

Two different settings are considered for the individual effects ε_{jd}:

- Gaussian errors with mean 0 and standard deviation 0.4 for the units belonging to twenty-four small areas and
- 15% contaminated Gaussian errors for the units belonging to the other six small areas where 85% percent of errors are generated from a Normal distribution with mean 0 and standard deviation 0.4 and the remaining 15% percent of errors are generated from a Normal distribution with mean 0 and standard deviation 2.

The setting for the first 24 small areas is considered as a situation of ÔregularlyÕ noisy data with $\rho = \frac{\sigma_\gamma^2}{\sigma_\gamma^2+\sigma_\varepsilon^2} = 0.2$. The setting for the second group of six small areas, on the contrary, defines a situation of more noisy data with the likely presence of outlying observations. We will denote by contaminated data the latter setting, while by uncontaminated data the former.

A sample of size $n = 600$ was selected from the simulated population, by simple random sampling. Each population was kept fixed for all simulation runs. A total of $T = 500$ simulations were carried out. For each sample MQ, NPMQ, EBLUP and NPEBLUP have been used to estimate the small area means. First, second and third quantile have been computed with NPMQ and MQ models under linear, cycle and jump signal. For MQ and NPMQ, the Huber Proposal 2 influence function is used with $c = 1.345$. This value gives reasonably high efficiency in the normal case it produces 95% efficiency when the errors are normal and still offers protection against outliers (see Huber(1981)). Moreover, in the correction term of the MQ and NPMQ estimators, robust (huberized) residuals instead of raw residuals are used. For each estimator and for each small area we computed the Monte Carlo estimate of the percentage relative bias

$$\text{RB}\%_d = \frac{\text{B}_{dMC}}{Y_d} 100; \tag{10.30}$$

the Root Mean Squared Error

$$\text{RMSE}_{dMC} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{\bar{Y}_{dt}} - \bar{Y}_d)^2}, \tag{10.31}$$

and the corresponding percentage Relative Root Mean Squared Error

$$\text{RRMSE}\%_d = \frac{\text{RMSE}_{dMC}}{\bar{Y}_d}100. \tag{10.32}$$

Figures 10.11, 10.12 and 10.13 report the RB$\%_d$, RRMSE$\%_d$ values obtained for this study under the linear, cycle and jump signal, respectively. Tables 1.6 , 1.7 and 1.8 show the behavior of quantiles estimates under the same types of signal. MSE estimation was monitored comparing MSE estimates and Monte Carlo MSEs, and by checking 95% confidence intervals coverage rates CR%. For MSE estimation of the NPMQ estimator we used expression (10.29), whereas the MSE estimation of MQ predictor was carried out following the method suggested in Chambers and Tzavidis (2006). MSE estimation of the EBLUP and NPEBLUP comes from analytical expressions introduced in Prasad and Rao (1990) and Opsomer et al. (2008), respectively. Intervals are defined by the small area mean estimate plus or minus twice their corresponding estimated root mean squared error. Areas are arranged in order of increasing population size and divided between uncontaminated and contaminated.

From Figure 10.11 – linear signal – we can see that in the uncontaminated areas 1-24 M-quantile type estimators (NPMQ and MQ) have a much better performance in terms of bias than the Mixed Model (MM) type estimators (EBLUP and NPEBLUP), while in terms of efficiency they all perform almost the same. In the contaminated areas 25-30, on the other hand, things change. Bias becomes an issue also for MQ estimators, although it seems that most of it comes from their poor performance in area 27. In addition, MM estimators are less efficient than MQ estimators in these areas. As of MSE estimation investigated through coverage rates, it should be noted that MQ estimators have a much better performance than MM estimators for both contaminated and uncontaminated areas. In the former, MM estimators have a way too low coverage, while in the latter a way too high. Finally, note that the fact that estimators based on nonparametric models – NPMQ and NPEBLUP – have the same performance of those based on linear models – MQ and EBLUP – shows that they do not loose efficiency under a linear model even using a too complex model. Focusing on the estimation of the quantiles from Table 1.6 it stems out that NPMQ and MQ have the same performance in terms of relative bias and relative mean squared error. The result is expected as the signal is linear: both in contaminated and uncontaminated areas the two methods give very similar results in average and median values of bias and variability.

From Figure 10.12 we can see that MQ type estimators have again smaller bias than MM estimators for uncontaminated areas. Efficiency, in this case, heavily depends on the model used, with the nonparametric estimators having smaller variability than the linear ones. NPMQ and NPEBLUP have similar performance, the former seems to have a better performance on average, while the latter in median. This is due to the poor performance of NPEBLUP in area 16. In the contaminated areas, NPMQ shows the best performance in terms of bias and efficiency. Coverage rates are comparable to those observed in the Linear signal simulation. Results from Figure 10.13 are very similar in substance to those from the Linear signal. This is due to the fact that this signal is linear for most of its domain. MQ estimators are again better in terms of bias in the first 24 areas. Efficiency is comparable for all estimators with NPMQ having the best performance on average. MQ estimators have a better performance in terms of bias and efficiency in the contaminated areas. Under the cycle setting Table 1.7 shows the results obtained in the estimation of the quantiles. The main finding here is that when relaxing the assumption of

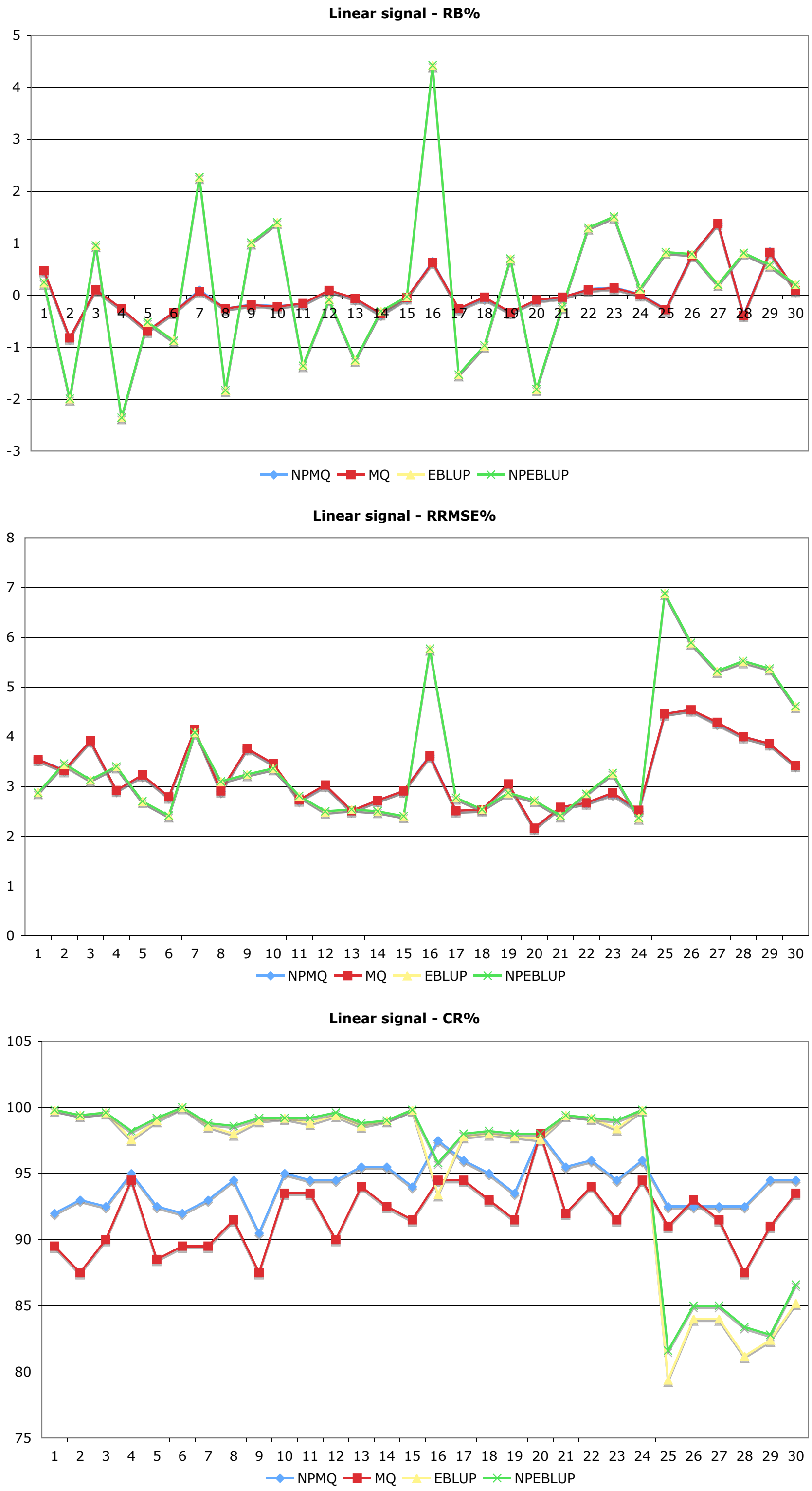


Figure 10.11: Relative Bias (RB%), Relative Root Mean Squared Errors (RRMSE%) and 95% Coverage Rate (CR%) in case of Linear signal. Areas are arranged in order of increasing population size.

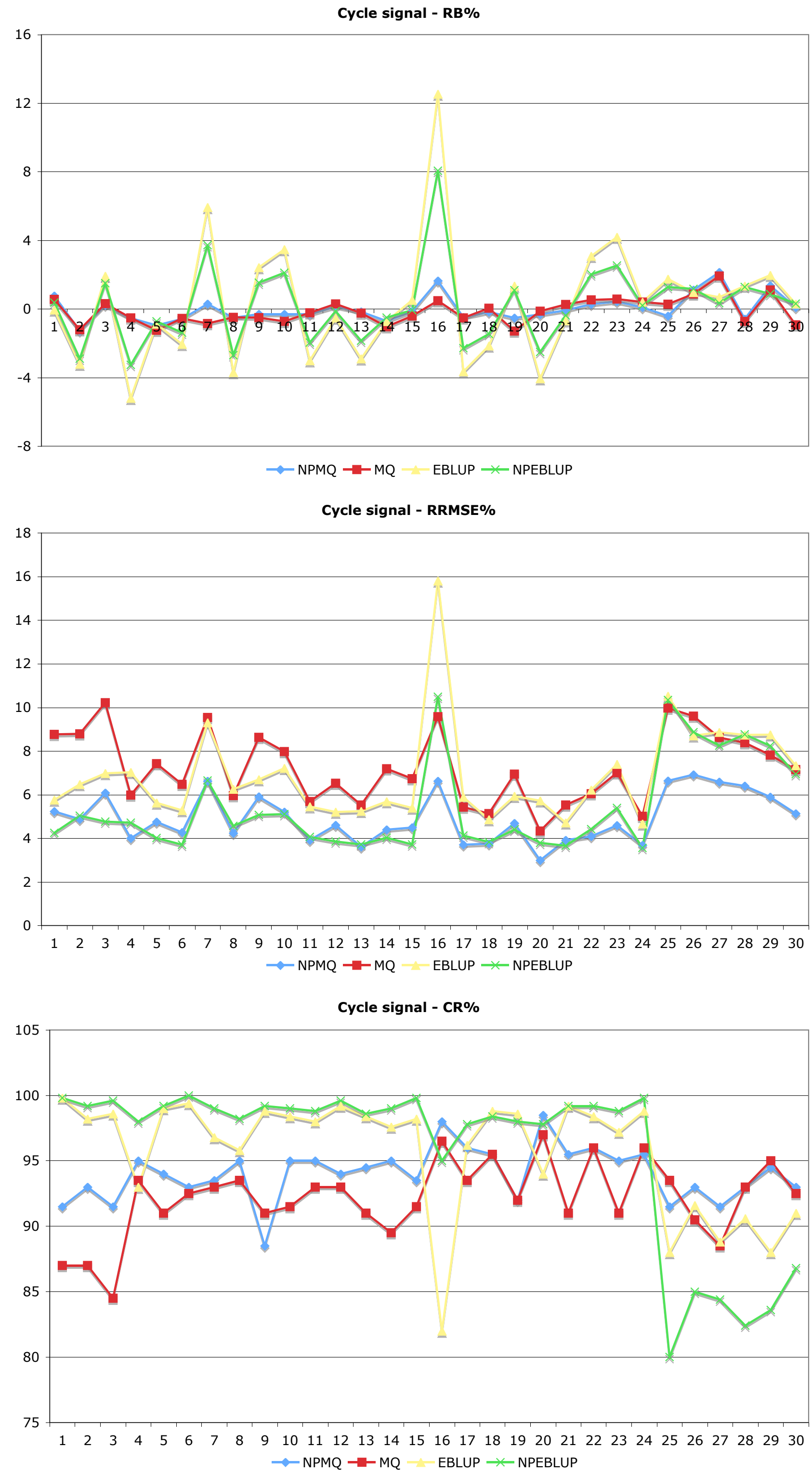


Figure 10.12: Relative Bias (RB%), Relative Root Mean Squared Errors (RRMSE%) and 95% Coverage Rate (CR%) in case of Cycle signal. Areas are arranged in order of increasing population size.

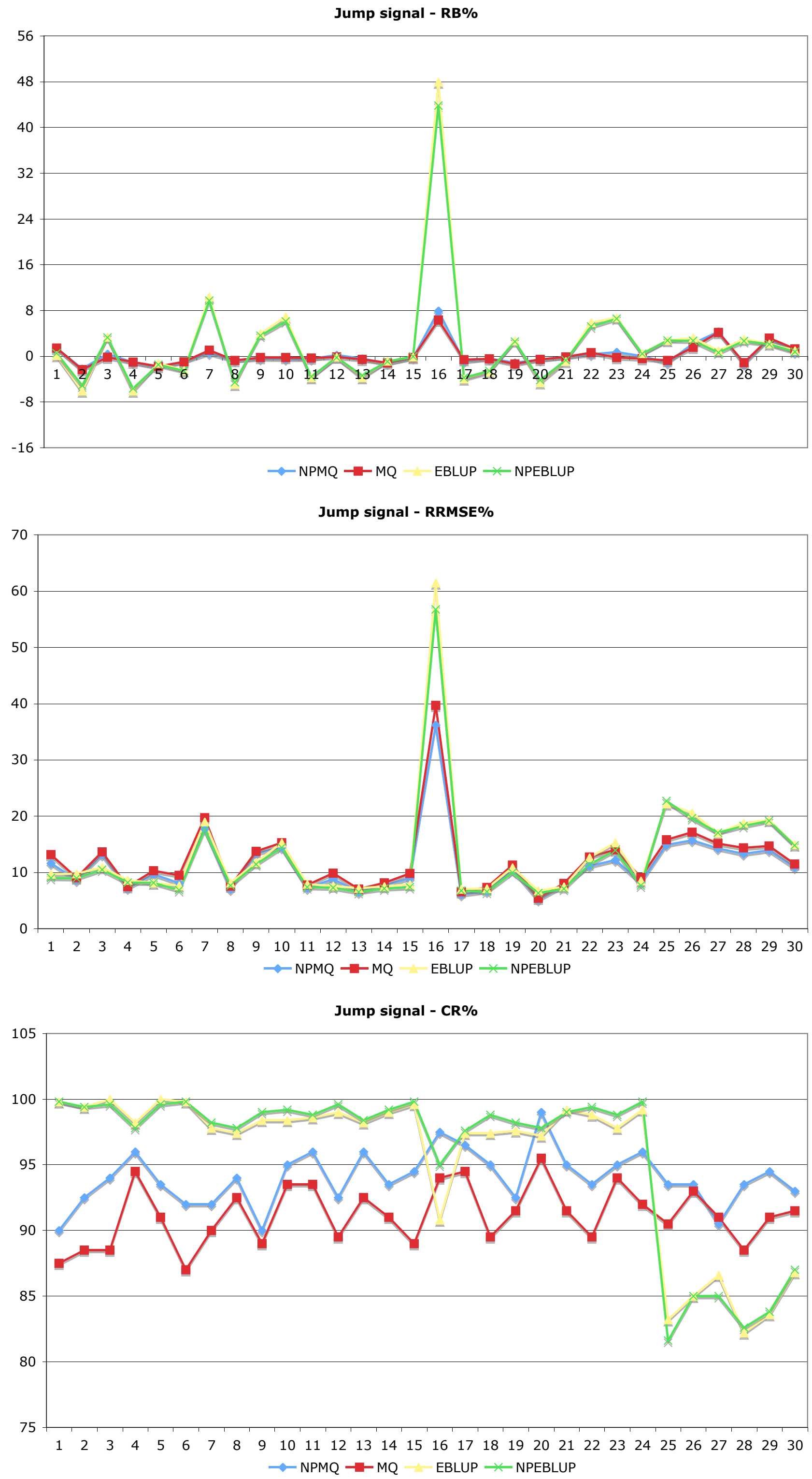


Figure 10.13: Relative Bias (RB%), Relative Root Mean Squared Errors (RRMSE%) and 95% Coverage Rate (CR%) in case of Jump signal. Areas are arranged in order of increasing population size.

linearity NPMQ definitely has a better performance than the traditional MQ estimator. Relative bias and Relative Root Mean Squared Error are always lower in the case of non parametric model with relevant gains especially in the last six contaminated areas. The result is confirmed under the jump setting of Table 1.8. Here the mean and median level of bias are generally higher than in the linear and cycle setting and also the RRMSE states a much higher variation of the estimates both in contaminated and uncontaminated areas. However the performance of NPMQ estimator is still appreciable in comparison with the traditional MQ estimators which rely on the linearity assumption.

Figures 10.14, 10.15 and 10.16 show how different root mean squared estimators track the true root mean squared error of the different estimator under linear, cycle and jump signal. Each figure has the same structure. Top left is the NPMQ predictor (10.22) with RMSE estimated using (10.29). Top right is the MQ predictor with RMSE estimator suggested by Tzavidis and Chambers (2007). Bottom left is EBLUP predictor with RMSE estimator suggested by Prasad and Rao (1990) and bottom right is the NPEBLUP estimator with RMSE analytical estimator suggested by Opsomer et al. (2008).

Figure 10.14 shows the area-specific values of RMSE and average estimated RMSE in case of linear signal. Estimator (10.29) performs well, showing only a small amount of undercoverage both for NPMQ and MQ estimators. Given that all its underlying assumptions are met, the Prasad and Rao (1990) and Opsomer et al. (2008) estimators of RMSE works very well in terms of empirical coverage. However, we note that they have a smoothing effect on the estimated variability of the small areas. This is due the fact that RMSE estimates are based on Prasad and Rao (1990) type-estimator, which targets the unconditional RMSE, whereas in the simulation experiments each population is kept fixed, then the empirical MSE is conditioned on the small area effects. We have also run simulations in which we randomized over the small area distribution and the results show still a good performance of the conditional RMSE estimators for the M-quantile type-estimators. As of variance estimators for mixed model-type estimators, their performance is better than in the conditional case considered here, even if it still suffers from the smoothing effect. Figure 10.17 shows such results in case of linear signal. The other detailed results are available to the interested reader from the authors. However, we believe that the setting considered here, by effectively fixing the differences between the small areas, constitutes a more practical and appropriate representation of the small area estimation problem in a finite population perspective and the conditioned RMSE is likely closer to the RMSE of interest to people using small area methods.

In case of cycle signal (Figure 10.15) the NPMQ and the MQ MSE estimators have the best performance in tracking the true variability. EBLUP and NPEBLUP MSE estimators smooth the behavior across the areas. Under the jump signal (Figure 10.16) estimator (10.29) for NPMQ and MQ estimators tracks the true behavior of RMSE. Both Prasad and Rao (1990) and Opsomer et al. (2008) estimators confirm their smoothing effect.

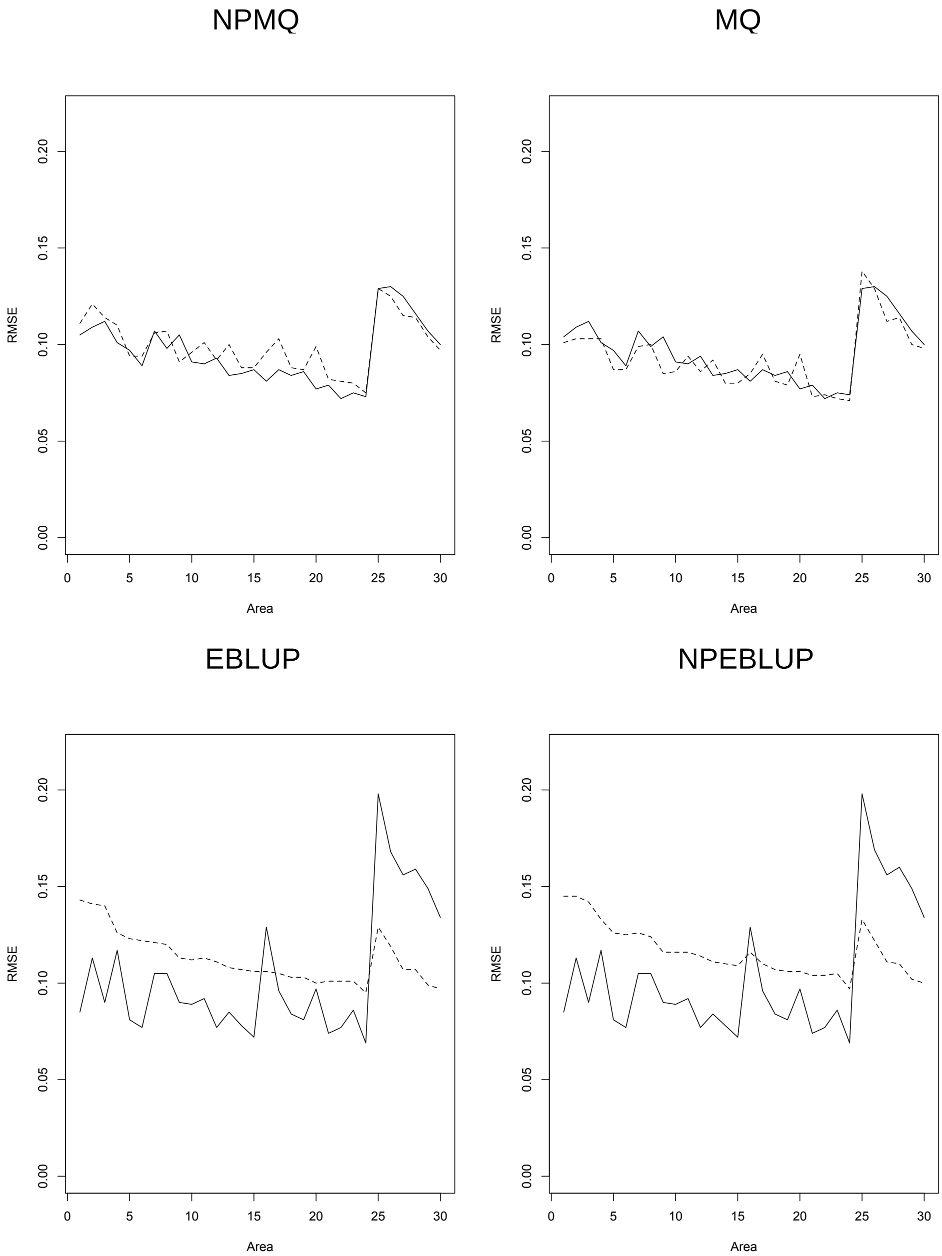


Figure 10.14: Area-specific values of RMSE (solid line) and average estimated RMSE (dashed line) in case of Linear signal. Areas are arranged in order of increasing population size (last six areas are the contaminated areas).

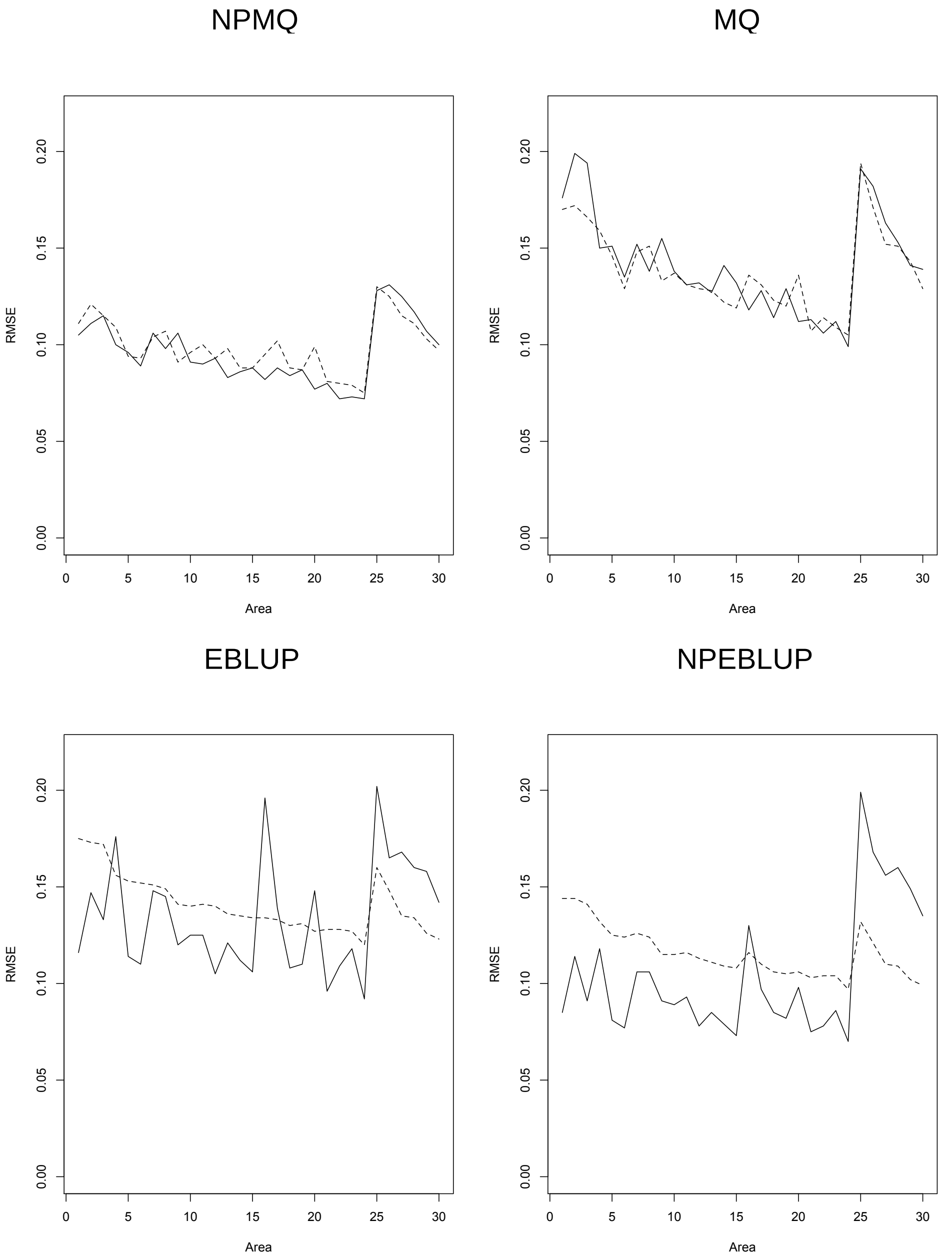


Figure 10.15: Area-specific values of RMSE (solid line) and average estimated RMSE (dashed line) in case of Cycle signal. Areas are arranged in order of increasing population size (last six areas are the contaminated areas).

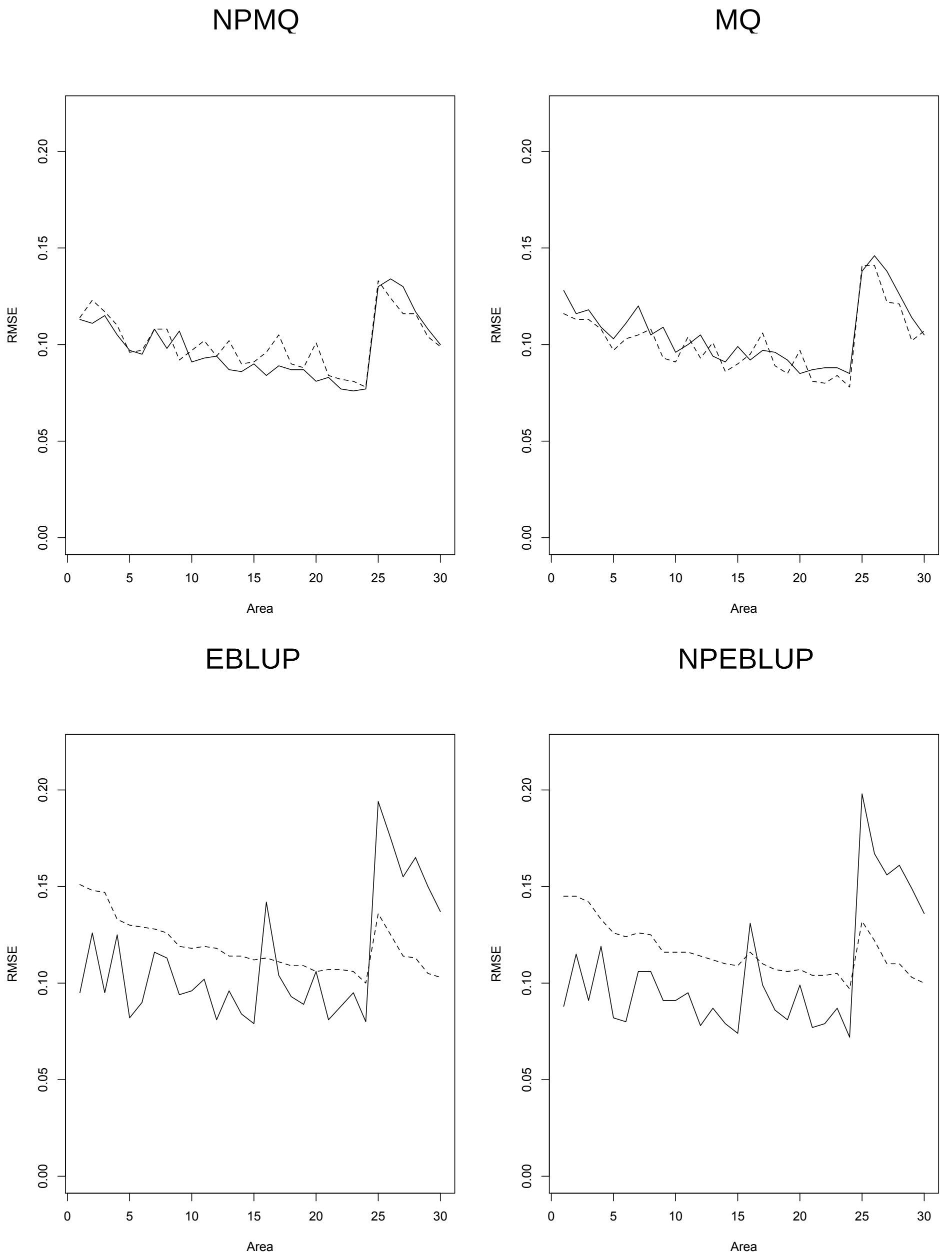


Figure 10.16: Area-specific values of RMSE (solid line) and average estimated RMSE (dashed line) in case of Jump signal. Areas are arranged in order of increasing population size (last six areas are the contaminated areas).

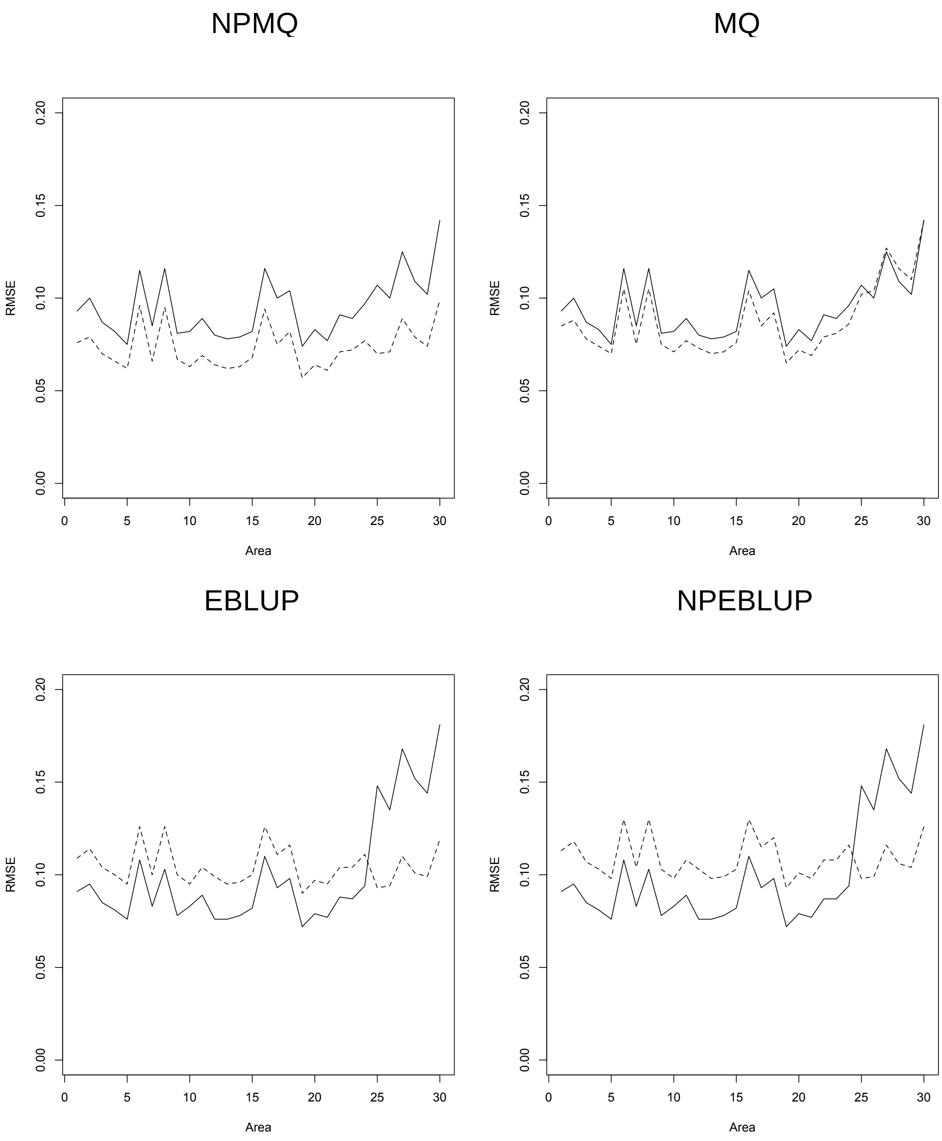


Figure 10.17: Area-specific values of unconditional RMSE (solid line) and average estimated RMSE (dashed line) in case of Linear signal. Areas are arranged in order of increasing population size (last six areas are the contaminated areas).

	Quantile 0.25				Quantile 0.5				Quantile 0.75			
Area	NPMQ		MQ		NPMQ		MQ		NPMQ		MQ	
	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%
Uncontaminated areas												
1	-0.64	5.20	-0.71	5.15	-0.06	3.97	-0.23	3.96	-0.28	3.38	-0.44	3.39
2	-0.74	5.51	-0.58	5.52	-0.65	4.48	-0.60	4.48	-0.26	3.72	-0.24	3.74
3	-0.74	4.84	-0.60	4.68	-0.41	3.86	-0.59	3.92	-0.30	3.49	-0.38	3.52
4	-0.71	4.63	-0.49	4.41	-0.44	3.78	-0.54	3.63	-0.44	3.31	-0.30	3.12
5	-0.53	4.25	-0.36	4.28	-0.32	3.58	-0.45	3.54	-0.28	3.06	-0.31	3.15
6	-0.06	5.95	-0.05	5.92	-0.16	4.65	-0.01	4.63	-0.02	4.14	-0.11	4.21
7	-0.74	4.61	-0.71	4.81	-0.07	3.67	-0.40	3.76	-0.18	3.25	-0.22	3.26
8	-0.68	6.30	-0.81	6.32	-0.69	4.97	-0.50	5.00	-0.64	4.26	-0.74	4.43
9	-0.22	4.53	-0.36	4.65	-0.44	3.84	-0.45	3.89	-0.24	3.31	-0.42	3.21
10	-0.71	4.68	-0.54	4.51	-0.47	3.49	-0.47	3.37	-0.31	3.10	-0.35	2.97
11	-0.53	4.75	-0.61	4.83	-0.27	3.75	-0.33	3.80	-0.27	3.46	-0.23	3.41
12	-0.68	4.69	-0.61	4.71	-0.31	3.74	-0.26	3.58	-0.37	3.14	-0.35	3.21
13	-0.61	4.46	-0.69	4.63	-0.41	3.32	-0.49	3.34	-0.22	2.93	-0.41	2.92
14	-0.54	4.73	-0.70	4.71	-0.58	3.74	-0.60	3.71	-0.41	3.33	-0.50	3.35
15	-0.37	4.72	-0.32	4.86	-0.37	3.86	-0.35	3.69	-0.23	3.36	-0.23	3.26
16	-0.55	6.15	-0.47	6.21	-0.31	5.06	-0.44	5.19	-0.39	4.42	-0.41	4.42
17	-1.48	5.54	-1.17	5.40	-0.77	4.35	-0.84	4.22	-0.57	3.74	-0.63	3.77
18	-0.11	5.65	-0.14	5.59	-0.15	4.42	-0.26	4.48	-0.22	3.67	-0.17	3.95
19	-0.34	4.15	-0.52	4.08	-0.40	3.41	-0.42	3.43	-0.27	2.94	-0.42	3.00
20	-0.63	4.53	-0.60	4.54	-0.53	3.68	-0.73	3.75	-0.27	3.08	-0.39	3.11
21	-0.60	4.27	-0.88	4.52	-0.60	3.64	-0.74	3.59	-0.32	3.03	-0.45	2.99
22	-1.00	5.04	-0.82	4.88	-0.43	4.11	-0.51	3.77	-0.35	3.41	-0.28	3.36
23	-0.61	5.29	-0.51	5.17	-0.40	4.22	-0.48	4.24	-0.48	3.52	-0.36	3.56
24	-0.35	5.24	-0.39	5.06	-0.24	4.14	-0.19	4.01	-0.01	3.50	-0.23	3.62
Mean (abs. values)	0.59	4.99	0.57	4.98	0.39	3.99	0.45	3.96	0.31	3.44	0.36	3.46
Median (abs. values)	0.61	4.74	0.59	4.82	0.40	3.85	0.46	3.79	0.28	3.37	0.35	3.36
Contaminated areas												
25	0.08	5.73	0.00	5.57	-0.33	4.29	-0.39	4.13	-0.16	4.20	-0.10	4.04
26	-0.84	5.80	-0.85	6.11	-0.19	4.11	-0.24	4.16	0.29	3.78	0.25	3.77
27	-0.75	6.82	-1.24	6.98	-1.26	5.47	-1.37	5.27	-0.70	5.01	-0.95	5.21
28	-0.18	6.80	-0.42	6.72	-0.32	4.54	-0.22	4.83	-0.26	4.14	-0.14	4.23
29	-0.99	5.99	-0.93	6.03	-0.50	4.29	-0.40	4.20	-0.28	4.18	-0.31	4.37
30	-1.25	8.90	-1.51	9.61	-0.77	6.26	-0.75	6.34	-0.19	6.13	-0.28	6.32
Mean (abs. values)	0.68	6.67	0.83	6.84	0.56	4.83	0.56	4.82	0.31	4.57	0.34	4.66
Median (abs. values)	0.80	6.39	0.89	6.42	0.41	4.42	0.39	4.51	0.27	4.19	0.26	4.30

Table 10.6: Relative Bias (RB%) and Relative Root Mean Squared Errors (RRMSE%) in case of Linear signal. Areas are arranged in order of increasing population size.

	Quantile 0.25				Quantile 0.5				Quantile 0.75			
Area	NPMQ		MQ		NPMQ		MQ		NPMQ		MQ	
	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%
Uncontaminated areas												
1	-0.96	9.92	3.67	14.20	-0.29	6.81	1.19	8.61	-0.27	4.91	-2.78	7.32
2	-1.61	10.14	6.42	15.01	-0.65	7.11	-1.37	8.97	-0.34	5.27	-3.80	8.14
3	-0.98	8.32	3.91	12.76	-0.43	5.46	-1.80	7.74	-0.46	4.28	-2.54	7.05
4	-0.97	8.54	0.77	11.35	-0.36	5.79	-0.03	7.44	-0.56	4.45	-1.19	6.30
5	-0.77	8.03	7.02	13.49	-0.51	5.95	-0.16	7.34	-0.13	4.08	-4.08	6.89
6	-0.27	11.16	4.40	15.05	0.19	7.25	0.39	9.75	-0.18	5.64	-2.09	7.72
7	-0.98	9.13	3.82	12.82	-0.23	6.13	1.15	8.31	-0.25	4.84	-2.82	6.81
8	-1.37	10.51	5.14	16.55	-0.84	7.55	-0.77	11.13	-0.87	5.87	-4.45	9.52
9	-1.00	9.08	8.61	15.51	-0.51	6.12	-0.09	7.67	-0.21	4.53	-4.07	7.46
10	-1.36	8.45	7.47	13.83	-0.59	5.51	-0.92	7.25	-0.59	4.09	-3.92	6.86
11	-1.07	9.01	4.36	13.47	-0.30	5.77	0.08	7.94	-0.17	4.47	-1.74	6.37
12	-1.00	8.91	4.36	12.74	-0.54	5.73	1.76	7.55	-0.50	4.41	-3.65	6.90
13	-0.70	7.60	5.13	11.84	-0.51	5.06	-1.63	7.20	-0.45	4.08	-2.51	6.13
14	-1.14	8.32	1.30	10.36	-0.54	5.78	-0.81	6.80	-0.76	4.68	-1.36	5.77
15	-0.65	8.86	8.47	15.64	-0.45	5.95	-0.32	7.98	-0.39	4.71	-4.04	7.25
16	-1.96	12.36	10.24	18.77	-0.35	8.04	-0.59	10.26	-0.58	6.15	-6.40	10.58
17	-1.85	9.80	6.75	15.39	-1.28	7.10	-0.55	9.08	-0.70	5.07	-4.73	8.51
18	-0.60	10.04	5.53	14.70	-0.52	6.80	-1.29	8.97	-0.13	5.28	-2.75	7.54
19	-0.62	7.70	7.56	12.96	-0.13	5.18	-0.61	6.91	-0.34	3.97	-4.00	6.43
20	-1.41	8.38	5.79	13.58	-0.83	5.80	0.55	7.56	-0.32	4.45	-4.21	7.38
21	-1.24	8.19	4.98	11.82	-0.98	5.61	-0.71	7.29	-0.64	4.25	-2.71	6.32
22	-1.73	8.96	5.33	14.18	-0.66	6.50	-0.90	8.13	-0.38	5.01	-3.92	7.41
23	-0.98	9.79	6.79	14.33	-0.82	6.18	-0.13	7.98	-0.39	4.96	-4.08	7.33
24	-0.69	10.11	0.13	13.94	-0.07	6.96	0.98	9.07	0.22	5.25	-1.74	7.31
Mean (abs. values)	1.08	9.22	5.33	13.93	0.52	6.26	0.78	8.21	0.41	4.78	3.32	7.30
Median (abs. values)	0.99	8.98	5.23	13.88	0.51	6.03	0.74	7.96	0.38	4.69	3.72	7.28
Contaminated areas												
25	-0.82	11.42	6.81	15.87	-0.11	7.54	-0.31	7.93	0.28	5.22	-3.01	7.51
26	-0.70	11.19	6.95	16.50	0.20	7.32	-1.12	8.12	0.19	5.08	-2.97	7.27
27	1.24	12.78	7.73	18.78	1.72	9.18	1.80	10.44	0.27	6.41	-2.32	8.51
28	-0.94	12.17	2.99	16.23	-0.64	7.89	-0.40	8.92	-0.55	6.10	-1.46	7.82
29	-1.34	11.35	5.27	16.36	-0.55	7.60	-0.34	9.02	-0.36	5.67	-3.69	8.31
30	-3.31	16.47	1.02	21.18	-0.01	9.89	-1.23	11.87	0.45	8.20	-1.69	10.75
Mean (abs. values)	1.39	12.56	5.13	17.49	0.54	8.24	0.87	9.38	0.35	6.11	2.52	8.36
Median (abs. values)	1.09	11.80	6.04	16.43	0.38	7.75	0.76	8.97	0.32	5.89	2.64	8.07

Table 10.7: Relative Bias (RB%) and Relative Root Mean Squared Errors (RRMSE%) in case of Cycle signal. Areas are arranged in order of increasing population size.

	Quantile 0.25				Quantile 0.5				Quantile 0.75			
Area	NPMQ		MQ		NPMQ		MQ		NPMQ		MQ	
	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%	RB%	RRMSE%
Uncontaminated areas												
1	-2,92	26,43	3,39	27,10	1,39	12,22	-6,91	13,27	-0,72	8,42	-3,03	9,00
2	-4,23	29,43	2,61	31,45	-1,65	13,39	-6,73	15,40	-1,07	9,12	-3,54	10,38
3	-4,27	26,69	3,70	28,97	-1,08	11,44	-6,36	14,01	-0,96	8,27	-3,86	9,68
4	-3,09	24,65	5,06	26,02	-0,04	11,01	-7,32	13,36	-0,88	7,50	-3,23	8,46
5	-2,52	20,02	0,19	22,37	0,03	11,21	-5,97	12,08	-0,35	7,05	-2,63	8,09
6	0,05	30,59	6,75	33,72	1,75	14,18	-5,73	15,35	-0,12	10,05	-3,30	11,29
7	-3,08	21,29	-2,71	22,65	0,89	11,09	-6,43	12,65	-0,76	7,74	-1,63	8,35
8	-4,39	31,04	2,49	34,46	-1,13	14,31	-7,25	17,00	-1,27	10,22	-5,42	12,33
9	-1,67	22,04	4,27	23,68	0,08	11,46	-6,70	13,11	-0,86	7,70	-4,20	9,08
10	-3,42	22,89	2,05	23,36	-0,39	9,69	-6,40	11,92	-0,65	7,29	-3,76	8,49
11	-2,21	21,86	-2,26	24,45	0,09	12,10	-6,83	13,18	-0,49	8,17	-1,90	9,09
12	-3,06	21,67	0,46	22,48	0,37	11,02	-6,62	12,15	-0,62	7,21	-2,36	8,19
13	-3,36	24,16	2,05	23,30	-0,90	9,84	-5,73	11,66	-0,62	7,34	-2,73	8,09
14	-2,48	23,18	2,25	24,57	-0,35	11,29	-7,07	13,93	-1,04	7,63	-3,59	9,07
15	-2,15	22,71	3,48	24,39	0,32	11,25	-5,82	12,18	-0,64	7,77	-2,85	8,75
16	-3,69	31,87	6,58	34,82	-0,07	14,88	-5,43	16,37	-1,02	10,83	-4,11	11,92
17	-6,71	28,69	2,66	29,61	-1,29	13,01	-7,77	15,16	-1,24	8,93	-4,94	10,57
18	-1,54	28,34	3,05	29,57	0,17	12,81	-4,74	14,46	-0,46	9,26	-2,26	9,86
19	-1,42	19,89	4,77	22,13	-0,20	10,45	-6,27	11,77	-0,66	7,08	-4,00	8,42
20	-3,60	21,31	0,41	21,79	-1,06	11,44	-7,33	13,03	-0,64	7,18	-2,36	8,24
21	-2,66	20,83	-0,09	22,63	-0,64	11,50	-7,38	12,93	-0,76	7,39	-3,41	8,28
22	-4,92	24,14	2,53	24,94	-1,43	12,33	-6,58	13,41	-0,64	8,20	-3,50	9,41
23	-3,77	27,58	4,41	27,64	-0,10	13,01	-6,85	14,50	-0,91	8,45	-3,59	9,84
24	-0,27	24,51	-1,11	26,77	0,46	12,85	-7,05	14,07	0,36	8,40	-0,32	9,13
Mean (abs. values)	2,98	24,82	2,89	26,37	0,66	11,99	6,55	13,62	0,74	8,22	3,19	9,33
Median (abs. values)	3,07	24,15	2,64	24,75	0,42	11,48	6,66	13,32	0,69	7,97	3,36	9,07
Contaminated areas												
25	1,20	33,06	6,28	30,91	-1,27	13,07	-5,93	13,47	-0,47	9,23	-1,89	9,91
26	-4,09	34,34	6,21	34,32	-0,41	12,50	-5,10	13,33	0,42	8,75	-1,75	9,46
27	-5,39	35,60	-2,62	37,68	-2,96	15,58	-8,02	17,10	-0,79	11,23	-1,65	12,29
28	0,06	40,83	4,69	40,84	-0,41	14,34	-6,18	15,07	-0,39	9,57	-1,39	10,58
29	-3,78	29,62	-0,41	29,58	-0,22	13,31	-6,22	13,63	-0,47	9,09	-0,54	10,01
30	-4,69	50,00	-0,04	50,10	-1,12	18,40	-6,88	20,52	-1,00	13,73	-2,82	15,14
Mean (abs. values)	3,20	37,24	3,38	37,24	1,07	14,53	6,39	15,52	0,59	10,26	1,67	11,23
Median (abs. values)	3,93	34,97	3,65	36,00	0,77	13,83	6,20	14,35	0,47	9,40	1,70	10,30

Table 10.8: Relative Bias (RB%) and Relative Root Mean Squared Errors (RRMSE%) in case of Jump signal. Areas are arranged in order of increasing population size.

10.4 M-quantile GWR models

Typically, random effects models assume independence of the random area effects. This independence assumption is also implicit in M-quantile small area models. In economic applications, however, observations that are spatially close may be more related than observations that are further apart. This spatial correlation can be accounted for by extending the random effects model to allow for spatially correlated area effects using, for example, a Simultaneous Autoregressive (SAR) model (see Petrucci and Salvati (2006), Pratesi and Salvati (2008) and Pratesi and Salvati (2009)). An alternative approach to incorporate the spatial information in the regression model is by assuming that the regression coefficients vary spatially across the geography of interest. Geographically Weighted Regression (GWR) (see Brundson et al. (1996)) extends the traditional regression model by allowing local rather than global parameters to be estimated. In a recent paper Salvati et al. (2008) proposed an M-quantile GWR small area model. The authors proposed an extension to the GWR model, the M-quantile GWR model, i.e. a locally robust model for the M-quantiles of the conditional distribution of the outcome variable given the covariates. Here we report a brief description of the M-quantile GWR model.

10.4.1 M-quantile geographically weighted regression

In this Section we define a spatial extension to linear M-quantile regression based on GWR. Since M-quantile models do not depend on how areas are specified, we also drop the area subscript d from our notation in this Section.

Given n observations at a set of L locations $\{u_l; l=1,\ldots,L; L\leqslant n\}$ with n_l data values $\{(y_{jl},\mathbf{x}_{jl}); i=1,\ldots,n_l\}$ observed at location u_l, a linear GWR model is a special case of a locally linear approximation to a spatially non-linear regression model and is defined as follows

$$y_{jl}=\mathbf{x}_{jl}^T\beta(u_l)+\varepsilon_{jl}, \tag{10.33}$$

where $\beta(u_l)$ is a vector of p regression parameters that are specific to the location u_l and the ε_{il} are independently and identically distributed random errors with zero expected value and finite variance. The value of the regression parameter 'function' $\beta(u)$ at an arbitrary location u is estimated using weighted least squares

$$\hat{\beta}(u)=\Big\{\sum_{l=1}^{L}w(u_l,u)\sum_{i=1}^{n_l}\mathbf{x}_{jl}\mathbf{x}_{jl}^T\Big\}^{-1}\Big\{\sum_{l=1}^{L}w(u_l,u)\sum_{i=1}^{n_l}\mathbf{x}_{jl}y_{jl}\Big\},$$

where $w(u_l,u)$ is a spatial weighting function whose value depends on the distance from sample location u_l to u in the sense that sample observations with locations close to u receive more weight than those further away. In this paper we use a Gaussian specification for this weighting function

$$w(u_l,u)=\exp\Big\{-d_{u_l,u}^2/2b^2\Big\}, \tag{10.34}$$

where $d_{u_l,u}$ denotes the Euclidean distance between u_l and u and $b>0$ is the bandwidth. As the distance between u_l and u increases the spatial weight decreases exponentially. For example, if $w(u_l,u)=0.5$ and $w(u_m,u)=0.25$ then observations at location u_l have twice the weight in determining the fit at location u compared with observations at location u_m. Alternative weighting functions, corresponding to density functions other than the Gaussian, can be used. See Fotheringham et al. (2002) for a discussion of other weighting functions.

The bandwidth b is a measure of how quickly the weighting function decays with increasing distance, and so determines the 'roughness' of the fitted GWR function. A spatial weighting function with a small bandwidth will typically result in a rougher fitted surface than the same function with a large bandwidth. In this paper we use a single bandwidth for our extension of GWR to M-quantile regression. This global bandwidth is defined by minimising the cross-validation criterion proposed by Fotheringham et al. (2002):

$$CV = \sum_{l=1}^{L}\sum_{j=1}^{n_l}[y_{jl} - \hat{y}_{(j)l}(b)]^2,$$

where $\hat{y}_{(j)l}(b)$ is the predicted value of y_{jl}, using bandwidth b, with the observation y_{jl} omitted from the model fitting process. The value of b that minimises CV is then selected. An alternative approach is to use optimal local bandwidths (see Farber and Páez (2007)). However, this significantly increases the computational intensity of the model fitting process.

The GWR model (10.33) is a linear model for the conditional expectation of y given $\mathbf{X}$ at location u. That is, this model characterises the local behaviour of the conditional expectation of y given $\mathbf{X}$ as a linear function of $\mathbf{X}$. However, a more complete picture of the relationship between y and $\mathbf{X}$ at location u can be constructed by specifying a model for the conditional distribution of y given $\mathbf{X}$ at this location. Since the M-quantiles serve to characterise this conditional distribution, such a model can be defined by extending

$$Q_q(\mathbf{x}_j;\psi) = \mathbf{x}_j^T\beta_\psi(q). \tag{10.35}$$

to specify a linear model for the M-quantile of order q of the conditional distribution of y given $\mathbf{X}$ at location u, writing

$$Q_q(\mathbf{x}_{jl};\psi,u) = \mathbf{x}_{jl}^T\beta_\psi(u;q), \tag{10.36}$$

where now $\beta_\psi(u;q)$ varies with u as well as with q. Like (10.33), we can interpret (10.36) as a local linear approximation, in this case to the (typically) non-linear order q M-quantile regression function of y on $\mathbf{X}$, thus allowing the entire conditional distribution (not just the mean) of y given $\mathbf{X}$ to vary from location to location. The parameter $\beta_\psi(u;q)$ in (10.36) at an arbitrary location u can be estimated by solving

$$\sum_{l=1}^{L} w(u_l,u)\sum_{j=1}^{n_l}\psi_q\{y_{jl} - \mathbf{x}_{jl}^T\beta_\psi(u;q)\}\mathbf{x}_{jl} = \mathbf{0}, \tag{10.37}$$

where $\psi_q(\varepsilon) = 2\psi(s^{-1}\varepsilon)\{qI(\varepsilon > 0) + (1-q)I(\varepsilon \leqslant 0)\}$, s is a suitable robust estimate of the scale of the residuals $y_{jl} - \mathbf{x}_{jl}^T\beta_\psi(u;q)$, e.g. $s = median|y_{jl} - \mathbf{x}_{jl}^T\beta_\psi(u;q)|/0.6745$, and we will typically assume a Huber Proposal 2 influence function, $\psi(\varepsilon) = \varepsilon I(-c \leqslant \varepsilon \leqslant c) + sgn(\varepsilon)I(|\varepsilon| > c)$. Provided c is bounded away from zero, we can solve (10.37) by combining the iteratively re-weighted least squares algorithm used to fit the 'spatially stationary' M-quantile model (10.35) and the weighted least squares algorithm used to fit a GWR model. Put $w_\psi(\varepsilon) = \psi_q(\varepsilon)/\varepsilon$ and $w_{\psi jl} = w_\psi(\varepsilon_{jl})$. Then (10.37) can be written as

$$\sum_{l=1}^{L} w(u_l,u)\sum_{j=1}^{n_l} w_{\psi jl}\{y_{jl} - \mathbf{x}_{jl}^T\beta_\psi(u;q)\}\mathbf{x}_{jl} = \mathbf{0}.$$

Note that the spatial weights $w(u_l,u)$ in (10.37) do not depend on q. That is, the degree of spatial smoothing is the same at every value of q. Spatial weights that vary with q are straightforward to define by allowing the bandwidth underpinning these weights to vary with q. Such a q-specific optimal bandwidth b can be obtained by minimising the following function with respect to b

$$\sum_{l=1}^{L}\sum_{j=1}^{n_l}[y_{jl}-\hat{y}_{(j)l}(q;b)]^2,$$

where $\hat{y}_{(j)l}(q;b)$ is the estimated value of the right hand side of (10.36) at quantile q and location u_{jl}, using bandwidth b when the observation y_{jl} is omitted from the model fitting process. However, using this q-specific cross-validation criterion can significantly increase the computational time. In this work we therefore use the optimal bandwidth at $q = 0.5$ for all other values of q. We note that this choice could potentially lead to over-smoothing for small or large values of q and hence bias. Nevertheless, it is a reasonable first approximation to the q-specific optimal bandwidth that can be computed reasonably quickly.

An R function that implements an iterative re-weighted least squares algorithm for for fitting (10.36) is available from Salvati and Tzavidis (2010). The steps of this algorithm are as follows:

1. For specified q and for each location u of interest, define initial estimates $\beta_\psi^{(0)}(u;q)$.

2. At each iteration t, calculate residuals $\varepsilon_{jl}^{(t-1)} = y_{jl} - \mathbf{x}_{jl}^T\beta_\psi^{(t-1)}(u;q)$ and associated weights $w_{\psi jl}^{(t-1)}$ from the previous iteration.

3. Compute the new weighted least squares estimates from

$$\hat{\beta}_\psi^{(t)}(u;q) = \left\{\mathbf{X}^T W^{*(t-1)}(u;q)\mathbf{X}\right\}^{-1}\mathbf{X}^T W^{*(t-1)}(u;q)\mathbf{y}, \tag{10.38}$$

   where $\mathbf{y}$ is the vector of n sample values and $\mathbf{X}$ is the corresponding matrix of order $n \times p$ of sample x values. The matrix $W^{*(t-1)}(u;q)$ is a diagonal matrix of order n with entry, corresponding to a particular sample observation, set equal to the product of this observation's spatial weight, which depends on its distance from location u, and the weight that this observation has when the sample data are used to calculate the 'spatially stationary' M-quantile estimate $\hat{\beta}_\psi(q)$.

4. Repeat steps 1-3 until convergence. Convergence is achieved when the difference between the estimated model parameters obtained from two successive iterations is less than a very small value.

The fitted regression surface $\hat{Q}_q(\mathbf{x}_{jl};\psi,u) = \mathbf{x}_{jl}^T\hat{\beta}_\psi(u;q)$ then defines the fit of the M-quantile GWR model for the regression M-quantile of order q of y given $\mathbf{X}$ at location u.

One may argue that (10.36) is over-parameterised as it allows for both local intercepts and local slopes. An alternative spatial extension of the M-quantile regression model (10.35) that has a smaller number of parameters is one that combines local intercepts with global slopes and is defined as

$$Q_q(\mathbf{x}_{jl};\psi,u) = \mathbf{x}_{jl}^T\beta_\psi(q) + \delta_\psi(u;q), \tag{10.39}$$

where $\delta_\psi(u;q)$ is a real valued spatial process with zero mean function over the space defined by locations of interest. Model (10.39) is fitted in two steps. At the first step we ignore the spatial structure in the data and estimate $\beta_\psi(q)$ directly via the iterative re-weighted least squares algorithm used to fit the standard linear M-quantile regression model (10.35). Denote this estimate by $\hat{\beta}_\psi(q)$. At the second step we use geographic weighting to estimate $\delta_\psi(u;q)$ via

$$\hat{\delta}_\psi(u;q) = n^{-1}\sum_{l=1}^{L} w(u_l,u)\sum_{j=1}^{n_l}\psi_q\{y_{jl} - \mathbf{x}_{jl}^T\hat{\beta}_\psi(q)\}. \tag{10.40}$$

Choosing between (10.36) and (10.39) will depend on the particular situation and whether it is reasonable to believe that the slope coefficients in the M-quantile regression model vary significantly between locations. However, it is clear that since (10.39) is a special case of (10.36), the solution to (10.37) will have less bias and more variance than the solution to (10.40). Hereafter we refer to (10.36) and (10.39) as the MQGWR and MQGWR-LI (Local Intercepts) models respectively.

Note that estimates of the local (GWR) M-quantile regression parameters are derived by solving the estimating equation (10.37) using iterative re-weighted least squares, without any assumption about the underlying conditional distribution of y_{jl} given $\mathbf{x}_{jl}$ at each location u_l. That is, the approach is distribution-free. For details see Salvati et al. (2008).

10.4.2 Using M-quantile GWR models in small area estimation

SAR models allow for spatial correlation in the model error structure to be used to improve SAE. Alternatively, this spatial information can be incorporated directly into the model regression structure via an M-quantile GWR model for the same purpose. In this Section we describe how this can be achieved. We now assume that we have only one population value per location, allowing us to drop the index l. We also assume that the geographical coordinates of every unit in the population are known, which is the case with geo-coded data. The aim is to use these data to predict the area d mean of y using the M-quantile GWR models (10.36) and (10.39).

Following Chambers and Tzavidis (2006), we first estimate the M-quantile GWR coefficients $q_j; j \in s$ of the sampled population units without reference to the small areas of interest. A grid-based interpolation procedure for doing this under (10.35) is described by Chambers and Tzavidis (2006) and can be used directly with (10.39). We adapt this approach to the GWR M-quantile model (10.36) by first defining a fine grid of q values in the interval $(0,1)$. Chambers and Tzavidis (2006) use a grid that ranges between 0.01 and 0.99 with step 0.01. We employ the same grid definition and then use the sample data to fit (10.36) for each distinct value of q on this grid and at each sample location. The M-quantile GWR coefficient for unit j with values y_j and $\mathbf{x}_j$ at location u_j is finally calculated by using linear interpolation over this grid to find the unique value q_j such that $\hat{Q}_{q_j}(\mathbf{x}_j;\psi,u_j) \approx y_j$.

Provided there are sample observations in area d, an area d specific M-quantile GWR coefficient, $\hat{\theta}_d$ can be defined as the average value of the sample M-quantile GWR coefficients in area d, otherwise we set $\hat{\theta}_d = 0.5$. Following Tzavidis et al. (2010), the bias-adjusted M-quantile GWR predictor of the mean $\bar{Y}_d$ in small area d is then

$$\hat{\bar{Y}}_d^{MQGWR/CD} = N_d^{-1}\Big[\sum_{j\in U_d}\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi,u_j) + \frac{N_d}{n_d}\sum_{j\in s_d}\{y_j - \hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi,u_j)\}\Big], \tag{10.41}$$

where $\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi,u_j)$ is defined either via the MQGWR model (10.36) or via the MQGWR-LI model (10.39).

Variants of the M-quantile GWR model (10.36) can be defined by changing the value of the tuning constant c in the Huber Proposal 2 influence function. For example, an expectile version of the M-quantile GWR model can be fitted by substituting a large positive value for the tuning constant c in this influence function. Empirical comparisons of the 'large c' (i.e. expectile) and the more robust 'small c' Huber-type M-quantile small area models are reported in Chambers and Tzavidis (2006).

There are situations where we are interested in estimating small area characteristics for domains (areas) with no sample observations. The conventional approach to estimating a small area characteristic, say the mean, in this case is synthetic estimation. Under the linear mixed model the synthetic mean predictor for out of sample area d is $\widehat{\bar{Y}}_d^{LM/SYNTH} = N_d^{-1}\sum_{j\in U_d}\mathbf{x}_j^T\hat{\beta}$. The SAR model-based version of this predictor, $\widehat{\bar{Y}}_d^{SAR/SYNTH}$, has the same form, but substitutes the estimator $\hat{\beta}$.Under the M-quantile GWR model (10.36) the synthetic mean predictor for out of sample area d is $\widehat{\bar{Y}}_d^{MQGWR/SYNTH} = N_d^{-1}\sum_{j\in U_d}\hat{Q}_{0.5}(\mathbf{x}_j;\psi,u_j)$. We note that with MQGWR-based synthetic estimation all variation in the area-specific predictions comes from the area-specific auxiliary information, including the locations of the population units in the area. We expect that when a truly spatially non-stationary process underlies the data, use of $\widehat{\bar{Y}}_d^{MQGWR/SYNTH}$ will lead to improved efficiency relative to more conventional synthetic mean predictors.

10.4.3 Mean squared error estimation

A "pseudo-linearization" MSE estimator for M-quantile small area estimators was recommended by Chambers and Tzavidis (2006) and it has now been used successfully in empirical studies reported in a number of published papers on SAE, including the recent publications by Tzavidis et al. (2010) and Salvati et al. (2010a). Below we extend the argument of these papers to defining an estimator of a first order approximation to the mean squared error of (10.41). This extension is based on (i) a model where the regression of y on $\mathbf{X}$ for a particular population unit depends on its location, with this regression specified by the locally linear GWR model (10.33), and (ii) the fact that estimators derived under the MQGWR model (10.36) or the MQGWR-LI model (10.39) can be written as linear combinations of the sample values of y. For example, from (10.38) we see that (10.41) can be expressed as a weighted sum of the sample y-values

$$\widehat{\bar{Y}}_d^{MQGWR/CD} = N_d^{-1}w_d^T\mathbf{y}, \tag{10.42}$$

where

$$w_d = \frac{N_d}{n_d}\mathbf{1}_d + \sum_{j\in r_d}\mathbf{H}_{jd}^T\mathbf{x}_j - \frac{N_d-n_d}{n_d}\sum_{j\in s_d}\mathbf{H}_{jd}^T\mathbf{x}_j. \tag{10.43}$$

Here $\mathbf{1}_d$ is the n-vector with j-th component equal to one whenever the corresponding sample unit is in area d and is zero otherwise and

$$\mathbf{H}_{jd} = \left\{\mathbf{X}^TW^*(u_j;\hat{\theta}_d)\mathbf{X}\right\}^{-1}\mathbf{X}^TW^*(u_j;\hat{\theta}_d),$$

where $W^*(u;q)$ is the limit of the weighting matrices $W^{*(t-1)}(u;q)$ defined following (10.38).

If we assume that the weights defining the linear representation (10.42) are fixed, and that the values of y follow a location specific linear model, e.g. (10.33), then an estimator of the prediction variance of (10.42) can be computed following standard methods of heteroskedasticity-robust prediction variance estimation for linear predictors of population quantities (see Royall and Cumberland(1978)). Put $w_d = (w_{jd})$. This estimator is of the form

$$mse(\hat{\bar{Y}}_d^{MQGWR/CD}) = N_d^{-2} \sum_{k:n_k>0} \sum_{j \in s_k} \lambda_{jdk} \left\{ y_j - \hat{Q}_{\hat{\theta}_k}(\mathbf{x}_j; \psi, u_j) \right\}^2, \qquad (10.44)$$

where $\lambda_{jdk} = \left\{ (w_{jd} - 1)^2 + (n_d - 1)^{-1}(N_d - n_d) \right\} I(k = d) + w_{jk}^2 I(k \neq d)$ and $\hat{Q}_{\hat{\theta}_k}(\mathbf{x}_j; \psi, u_j)$ is assumed to define an unbiased estimator of the expected value of y_j given $\mathbf{x}_j$ at location u_j. Since the weights defining (10.43) reproduce the small area mean of $\mathbf{X}$, it also follows that (10.42) is unbiased for this mean in the special case where this expectation does not vary with location within the small area of interest, and so (10.44) then estimates the mean squared error of (10.42) in this special case. More generally, when the expectation of y_i given $\mathbf{x}_i$ varies from location to location within the small area, this unbiasedness holds on average provided sampling within the small area is independent of location, in which case (10.44) is an estimator of a first order approximation to the mean squared error of (10.42).

Note that (10.44) treats the weights (10.43) as fixed and so ignores the contribution to the mean squared error from the estimation of the area level M-quantile coefficients by $\hat{\theta}_d$. This is a pseudo-linearization assumption since for large overall sample sizes the contribution to the overall mean squared error of (10.42) arising from the variability of $\hat{\theta}_d$ will be of smaller order of magnitude than the fixed weights prediction variance of (10.42). As a consequence (10.44) will tend to be almost unbiased. However, the potential underestimation of the MSE of (10.42) implicit in (10.44) needs to be balanced against the bias robustness of this MSE estimator under misspecification of the second order moments of y, and may well lead to (10.44) being preferable to other MSE estimators based on higher order approximations that depend on the model assumptions being true. Empirical results reported in Tzavidis et al. (2010) indicate that the version of the MSE estimator (10.44) for the linear M-quantile predictor performs well both in model-based and design-based studies.

10.4.4 Simulations for M-quantile GWR models

In this Section we present results from a simulation study used to examine the performance of the M-quantile GWR small area estimators. In particular, we report results from model-based simulations where population data are generated at each simulation using a linear mixed model with different parametric assumptions about the distribution of errors and the spatial structure of the data, and a single sample is then taken from this simulated population according to a pre-specified design.

In these simulations, synthetic population values are generated under two versions of a linear mixed model and two distributional specifications for the random area effects and the individual residuals. Each population is of size $N = 10,500$ and contains $d = 30$ equal-sized small areas. More specifically, under the first model, population values of y are generated via $y_{jd} = 1 + 2x_{jd} + \gamma_j + \varepsilon_{jd}$ where $d = 1, \ldots, 350$ and $d = 1, \ldots, 30$. The values x_{jd} in this model are independently generated from the uniform distribution over the interval $[0,1]$, denoted as $x_{jd} \sim U[0,1]$, and the random effects are generated under two different distributional specifications: (a) Gaussian errors with $\gamma_d \sim N(0, 0.04)$ and $\varepsilon_{jd} \sim N(0, 0.16)$ and (b) Chi-squared errors with $\gamma_d \sim \chi^2(1) - 1$ and $\varepsilon_{jd} \sim \chi^2(3) - 3$, i.e. mean corrected Chi-squared variates with 1 and 3 degrees of freedom, respectively. For the second model, random effects are still simulated as in (a) and (b), but in addition the intercept and the slope of the linear model for y are allowed to vary with longitude and latitude. In particular, these simulations are based on the two-level model $y_{jd} = \alpha_{jd} + \beta_{jd} x_{jd} + \gamma_d + \varepsilon_{jd}$ with

$$\alpha_{jd} = 0.2 \times longitude_{jd} + 0.2 \times latitude_{jd},$$

$$\beta_{jd} = -5 + 0.1 \times longitude_{jd} + 0.1 \times latitude_{jd}$$

with the known location coordinates $(longitude_{jd}, latitude_{jd})$ for each population unit independently generated from $U[0,50]$. Note that the reason for using different parametric assumptions for the error terms of the linear mixed model is because we are interested in how the small area predictors perform both when the Gaussian assumptions of the linear mixed model are satisfied and when these assumptions are violated.

This simulation design corresponds to four scenarios (Gaussian stationary, Gaussian non-stationary, Chi-squared stationary, Chi-squared non-stationary). For each of these scenarios $T = 200$ Monte-Carlo populations are generated using the corresponding model specifications. For each generated population and for each area d we select a simple random sample (without replacement) of size $n_d = 20$, leading to an overall sample size of $n = 600$. The sample values of y and the population values of x obtained in each simulation are then used to estimate the small area means. Although a larger number of simulations would be preferable, this is not feasible due to the computer intensive nature of the model-fitting process. Note also that there is no specific motivation behind the choice of equal area specific sample sizes. Repetition of our simulation studies with unequal area-specific sample sizes does not lead to any differences in the conclusions that we draw below. These results of these simulations are not reported here, but are available from the authors.

Four different types of small area linear models are fitted to these simulated data. These are (i) a random intercepts version with uncorrelated random area effects, (ii) the linear M-quantile regression specification (10.35), (iii) the MQGWR model (10.36), and (iv) the MQGWR-LI model (10.39), with the last two models making use of the known location coordinates for the population units. An alternative model specification that can be used with this type of spatial data is one where the longitude and the latitude are included as covariates in the fixed part of the mixed model. This additional model specification is also investigated in the model-based simulations and is denoted by EBLUP+longlat in what follows. We did not include the SAR model in these comparisons because the spatial dependency simulated in the non-stationary scenarios is in the mean structure of the model and so favours the two M-quantile GWR small area models, ensuring that such a comparison would be rather one-sided.

The random intercepts model (i) is fitted using the default REML option of the *lme* function (see Section 10.3 in Venables and Ripley (2002)) in R. The M-quantile linear regression model (ii) is fitted using a modified version of the *rlm* function (see Section 8.3 in Venables and Ripley (2002)) in R and so uses iteratively re-weighted least squares to fit this model (see Chambers and Tzavidis (2006)). An extended version of this R code, available from Salvati and Tzavidis (2010), is used to fit the MQGWR models (iii) and (iv). Both the M-quantile regression and the M-quantile GWR models use the Huber Proposal 2 influence function with $c = 1.345$. Estimated model coefficients obtained from these fits are used to compute the EBLUP, the bias-adjusted M-quantile predictor , denoted by MQ below, and the MQGWR and the MQGWR-LI versions of the corresponding bias-adjusted M-quantile predictor (10.41).

The performance of the different small area estimators is evaluated with respect to three basic criteria: the bias and the root mean squared error of estimates of the small area means and the coverage rate of nominal 95 per cent confidence intervals for these means. The bias for small area d is computed as

$$Bias_d = \frac{1}{T}\sum_{t=1}^{T}(\hat{\bar{Y}}_{dt} - \bar{Y}_{dt}),$$

and the root mean squared error for area d is computed as

$$RMSE_d = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\bar{Y}}_{dt} - \bar{Y}_{dt})^2}.$$

The coverage performance of the confidence intervals is computed as

$$CR\%_d = \frac{1}{T}\sum_{t=1}^{T} I(|\hat{\bar{Y}}_{dt} - \bar{Y}_{dt}| \leqslant 2mse_{dt}^{1/2}) \times 100.$$

Note that the subscript t here indexes the T Monte-Carlo simulations, with $\bar{Y}_{dt}$ denoting the value of the small area d mean in simulation t and $\hat{\bar{Y}}_{dt}$, mse_{dt} denoting the area d estimated value and the corresponding estimated MSE in simulation t.

Key percentiles of the across areas distributions of the prediction biases and root mean squared errors of these estimators over the simulations are set out in Table 10.9. For Gaussian random effects and a spatially stationary regression surface, we see that EBLUP and EBLUP+longlat are the best predictors in terms of RMSE, as one would expect. The MQ, MQGWR and MQGWR-LI predictors all have similar bias and RMSE in this case. In contrast, when the underlying regression function is non-stationary we see that the MQGWR and MQGWR-LI predictors are considerably more efficient than the MQ, EBLUP and EBLUP+longlat predictors, and we also note that the RMSE of EBLUP+longlat is lower than the RMSE of the EBLUP. Under Chi-squared random effects this relative performance is unchanged, although the absolute differences in performance between the various predictors is much smaller. For a non-stationary Chi-squared process the RMSEs of EBLUP+longlat and the MQGWR estimators are similar.

In Table 10.10 we show key percentiles of the across area distributions of the true and estimated mean squared errors (the latter based on expression (10.44) and averaged over the simulations) of the MQGWR and MQGWR-LI predictors, as well as the corresponding area level coverage rates for ‘normal theory’-based nominal 95 per cent prediction intervals. Here coverage is defined by the number of times the interval, defined by the estimate of the small area mean plus or minus twice its estimated MSE, contains the ‘true’ population value. In general the proposed mean squared error estimator (10.44) provides a good approximation to the true mean squared error. These results also show that when M-quantile GWR fits are used in (10.44), then this estimator underestimates the true mean squared error of the corresponding predictor, leading to some undercoverage of associated prediction intervals. This is consistent with both the MQGWR and the MQGWR-LI models overfitting the actual population regression function. However, this bias is not excessive, being more pronounced in the case of the MQGWR model.

		Summary of across areas distribution					
Predictor	Indicator	Min	Q1	Median	Mean	Q3	Max
		Stationary process, Gaussian errors					
EBLUP	Bias	-0.051	-0.034	0.001	-0.001	0.023	0.068
	RMSE	0.068	0.075	0.079	0.081	0.087	0.101
EBLUP+longlat	Bias	-0.053	-0.030	0.001	-0.001	0.020	0.066
	RMSE	0.068	0.075	0.079	0.081	0.087	0.101
MQ	Bias	-0.015	-0.003	0.001	-0.001	0.003	0.012
	RMSE	0.074	0.083	0.088	0.087	0.091	0.100
MQGWR	Bias	-0.016	-0.007	-0.003	-0.002	0.005	0.008
	RMSE	0.067	0.084	0.088	0.087	0.091	0.100
MQGWR-LI	Bias	-0.010	-0.005	0.001	-0.001	0.003	0.012
	RMSE	0.073	0.085	0.087	0.086	0.090	0.097
		Non-stationary process, Gaussian errors					
EBLUP	Bias	-0.034	-0.013	-0.003	-0.002	0.011	0.031
	RMSE	0.169	0.193	0.205	0.220	0.238	0.323
EBLUP+longlat	Bias	-0.122	-0.063	-0.010	-0.001	0.041	0.162
	RMSE	0.061	0.085	0.119	0.124	0.138	0.225
MQ	Bias	-0.036	-0.011	0.000	-0.002	0.009	0.015
	RMSE	0.164	0.181	0.188	0.188	0.193	0.219
MQGWR	Bias	-0.047	-0.013	-0.005	-0.004	0.005	0.027
	RMSE	0.083	0.092	0.098	0.098	0.103	0.119
MQGWR-LI	Bias	-0.065	-0.010	-0.005	-0.004	0.007	0.047
	RMSE	0.088	0.097	0.107	0.112	0.114	0.186
		Stationary process, Chi-squared errors					
EBLUP	Bias	-0.441	-0.097	0.075	-0.011	0.112	0.237
	RMSE	0.399	0.457	0.482	0.489	0.511	0.651
EBLUP+longlat	Bias	-0.432	-0.069	0.061	-0.011	0.105	0.206
	RMSE	0.421	0.461	0.482	0.489	0.511	0.631
MQ	Bias	-0.063	-0.043	-0.021	-0.011	0.014	0.062
	RMSE	0.437	0.496	0.526	0.522	0.542	0.598
MQGWR	Bias	-0.075	0.002	0.035	0.028	0.060	0.113
	RMSE	0.482	0.507	0.539	0.539	0.564	0.633
MQGWR-LI	Bias	-0.067	-0.009	0.009	0.010	0.032	0.062
	RMSE	0.471	0.500	0.525	0.528	0.552	0.618
		Non-stationary process, Chi-squared errors					
EBLUP	Bias	-0.069	-0.046	-0.021	-0.014	0.008	0.069
	RMSE	0.465	0.541	0.560	0.566	0.592	0.675
EBLUP+longlat	Bias	-0.441	-0.071	-0.059	-0.010	0.118	0.209
	RMSE	0.440	0.512	0.538	0.539	0.562	0.678
MQ	Bias	-0.086	-0.048	-0.015	-0.014	0.021	0.051
	RMSE	0.460	0.540	0.554	0.555	0.586	0.641
MQGWR	Bias	-0.083	-0.009	0.022	0.017	0.050	0.124
	RMSE	0.482	0.507	0.534	0.535	0.562	0.619
MQGWR-LI	Bias	-0.085	-0.018	0.004	0.007	0.041	0.080
	RMSE	0.466	0.518	0.541	0.542	0.557	0.641

Table 10.9: Across areas distribution of Bias and RMSE over simulations.

		Percentile of across areas distribution					
Predictor	Indicator	10	25	Median	Mean	75	90
		Stationary process, Gaussian errors					
	True RMSE	0.080	0.084	0.088	0.087	0.091	0.093
MQGWR	Est. RMSE	0.076	0.078	0.081	0.081	0.083	0.085
	CR%	89.51	90.34	91.72	91.88	93.71	94.48
	True RMSE	0.079	0.085	0.087	0.086	0.090	0.090
MQGWR-LI	Est. RMSE	0.077	0.079	0.082	0.082	0.083	0.086
	CR%	90.45	91.13	93.00	92.88	94.50	95.00
		Non-stationary process, Gaussian errors					
	True RMSE	0.090	0.092	0.098	0.098	0.103	0.106
MQGWR	Est. RMSE	0.074	0.076	0.078	0.079	0.081	0.084
	CR%	84.30	85.00	87.00	87.08	89.38	90.50
	True RMSE	0.096	0.097	0.107	0.112	0.114	0.138
MQGWR-LI	Est. RMSE	0.085	0.088	0.098	0.100	0.103	0.122
	CR%	88.50	90.50	91.50	91.25	92.88	93.05
		Stationary process, Chi-squared errors					
	True RMSE	0.489	0.507	0.539	0.539	0.564	0.577
MQGWR	Est. RMSE	0.463	0.489	0.507	0.506	0.529	0.542
	CR%	85.71	89.10	90.38	90.24	92.15	92.44
	True RMSE	0.488	0.500	0.525	0.528	0.552	0.574
MQGWR-LI	Est. RMSE	0.467	0.486	0.505	0.508	0.528	0.543
	CR%	87.00	90.50	91.00	90.88	92.50	93.10
		Non-stationary process, Chi-squared errors					
	True RMSE	0.494	0.507	0.534	0.535	0.562	0.574
MQGWR	Est. RMSE	0.448	0.470	0.488	0.488	0.512	0.524
	CR%	85.50	88.13	90.00	89.40	91.00	92.05
	True RMSE	0.505	0.518	0.541	0.542	0.557	0.588
MQGWR-LI	Est. RMSE	0.485	0.501	0.515	0.514	0.529	0.537
	CR%	88.95	90.63	91.50	91.07	92.38	93.05

Table 10.10: Across areas distribution of true (i.e. Monte Carlo) root mean squared errors (True RMSE), area averages of estimated root mean squared errors (Est. RMSE) and area coverage rates (CR%) for nominal 95% prediction intervals. Intervals are defined by the small area mean estimate plus or minus twice their corresponding estimated root mean squared error.

Note that the construction of confidence intervals for small area parameters requires careful consideration. In our simulations we use the MSE estimation method described in Section 10.4.3 to generate "normal theory" intervals based on M-quantile model-based estimators. This use of the estimated MSE to construct confidence intervals, though widespread, has been criticised. Hall and Maiti (2006) and more recently Chatterjee et al. (2008), discuss the use of bootstrap methods for constructing confidence intervals for small area parameters since there is no guarantee that the asymptotic behaviour underpinning normal theory confidence intervals applies in the context of the small samples that characterise small area estimation. Further research on using the bootstrap techniques described by Tzavidis et al. (2010) to construct more accurate confidence intervals under the M-quantile GWR model is left for the future.

Chapter 11

References

Anselin, L. (1988). *Spatial Econometrics. Methods and Models*. Boston: Kluwer Academic Publishers.

Aragon, Y., Casanova, S., Chambers, R.L., Leconte, E. (2005). Conditional ordering using nonparametric expectiles. *Journal of Official Statistics*, **21**, 617-633.

Ballini, F., Betti, G., Carrette, S. and Neri, L. (2006). Poverty and inequality mapping in the Commonwealth of Dominica. *Estudios Economicos*, **2**, 123-162.

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. New York: Chapman and Hall.

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**, 28-36.

Bell, W. (1997). Models for county and state poverty estimates. Preprint, Statistical Research Division, U. S. Census Bureau.

Betti, G. and Verma, V. (1999). Measuring the degree of poverty in a dynamic and comparative context: a multi-dimensional approach using fuzzy set theory. *Proceedings, ICCS-VI, Lahore, Pakistan, August 27-31* **11**, 289-301.

Betti, G., Ferretti, C., Gagliardi, F., Lemmi, A. and Verma, V. (2009). Proposal for new multidimensional and fuzzy measures of poverty and inequality at national and regional level. *C.R.I.DI.RE. research centre working paper, University of Siena 83*

Bowman, A.W.,Hall, P., Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, **85**, 799-808.

Breckling, J. and Chambers, R.L. (1988). M-quantiles. *Biometrika*, **75**, 761-71.

Brunsdon, C., Fotheringham, A.S., Charlton, M. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, **28**, 281-298.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Linear Functional Model. *Statistics and Probability Letters*, **45**, 11–22.

Castro, P.E., Lawton, W.H. and Sylvestre, E.A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics*, **28**, 329–337.

Chambers, R.L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, **81**, 1063-1069.

Chambers, R.L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, **73**, 597-604.

Chambers, R.L., Dorfman, A.H., Hall, P. (1992). Properties of estimators of the finite population distribution function. *Biometrika*, **79**, 577-82.

Chambers, R.L. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, **93**, 255-268.

Chambers, R.L., Chandra, H., Tzavidis, N. (2008). On robust mean squared error estimation for linear predictors for domains. [Paper submitted for publication. A copy is available upon request].

Chambers, R., Chandra, H. and Tzavidis, N. (2009). On Bias-Robust Mean Squared Error Estimation for Linear Predictors for Domains. *Working Papers, 09-08* Centre for Statistical and Survey Methodology, The University of Wollongong, Australia. (Available from: http://cssm.uow.edu.au/publications).

Chatterjee, S., Lahiri, P. and Huilin, L. (2008). Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Linear Mixed Models. *Annals of Statistics*, **36**, 1221-1245.

Cheli, B. and Lemmi, A. (1995). A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty. *Economic Notes*, 24, 115-134.

Cheli, B. and Betti, G. (1999). Totally Fuzzy and Relative Measures of Poverty in Dynamics Context. *Metron*, **57(1-2)**, 83-104

Cole, R.J. (1988). Fitted smoothed centile curves to reference data. *Journal of the Royal Statistical Society, Series A*, **151**, 385-418.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, **31**, 377-403.

Cressie, N. (1991). Small-area prediction of undercount using the general linear model. *Proceedings of Statistic Symposium 90: Measurement and Improvement of Data Quality, Ottawa: Statistics Canada*, 93-105.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.

Cuadras, C.M. and Fortiana, J. (1995). A continuous metric sacling solution for a random variable. *Journal of Multivariate Amalysis*, **52**, 1-14.

Cuadras, C.M. and Fortiana, J. (1998). Visualizing categorical data with related metric scaling. In J. Blasius and M. Greenacre (Eds.), *Visualization of Categorical Data*, pp. 365-376. Academic Press.

Das, K., Jiang, J. and Rao, J. N. K. (2004). Mean squared error of empirical predictor. *The Annals of Statistics*, **32**, 818-840.

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

Datta, G.S., Lahiri, P. and Maiti, T. (2002). Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, **102**, 83–97.

Datta, G.S., Rao, J.N.K. and Smith, D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika* **92**, 183-196.

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function: some application to statistical inference. *Journal of Multivariate Amalysis*, **12**, 136–154.

Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, **78**, 605-610.

Eilers, P.H. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89-121.

Elbers, C., Lanjouw, J. O. and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, **71**, 355-364.

Fay, R. E. and Herriot, R. A. (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

Farber, S. and Páez, A. (2007). A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, **9**, 371-396.

Fellner, W.H. (1986). Robust Estimation of Variance Components. *Technometrics*, **28**, 51-60.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. New York: Springer-Verlag.

Foster, J., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures, *Econometrica*, **52**, 761–766.

Fotheringham, A.S., Brundson, C., and Charlton M. (2002). *Geographically Weighted Regression - The analysis of spatially varying relationship*. West Sussex, England: John Wiley & Sons Ltd.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55-93.

Ghosh, M., Nangia, N. and Kim, D. (1996). Estimation of median income of four-person families: a Bayesian time series approach. *Journal of the American Statistical Association*, **91**, 1423–1431.

Giusti, C., Pratesi, M., and Salvati, S. (2009a). Estimation of poverty indicators: a comparison of small area methods at LAU1-2 level in Tuscany. *Abstract Book, NTTS - Conferences on New Techniques and Technologies for Statistics*, Brussels, 18-20 Febbraio 2009.

Giusti, C., Pratesi, M., and Salvati, S. (2009b). Small area methods in the estimation of poverty indicators: the case of Tuscany. *Politica Economica*, **3**, 369-380.

Giusti, C., Marchetti, S., Pratesi, M., and Salvati, S. (in preparation). Semiparametric Fay-Herriot model using Penalized Splines.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2007). Estimation of the mean squared error of predictors of small area linear parameters under logistic mixed model. *Computational Statistics and Data Analyis*, **51**, 2720-2733.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2008a). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, **78**, 443–462.

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

Griffith, D. and Csillag, F. (1993). Exploring relationships between semi-variogram and spatial autoregressive. *Papers in Regional Science*, **72**, 283-296.

Hall, P. and Maiti, T. (2006a). On Parametric Bootstrap Methods for Small Area Prediction. *Journal Royal Statistical Society, Series B*, **68**, 221-238.

Hall, P. and Maiti, T. (2006b). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *The Annals of Statistics*, **34**, 1733-1750.

Hayfield, T. and Racine, J. (2008). Nonparametric Econometrics: The *np* Package. *Journal of statistical Software*, **27**, number 5.

Harville, D. and Jeske, D. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, **87**, 724-731.

Haslett, S. and Jones, G. (2005). Small area estimation using surveys and some practical and statistical issues. *Statistics in Transition*, **7**, 541–555.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, **51**, 186-192.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423-447.

Huber, P.J. (1981). *Robust Statistics*. John Wiley & Sons.

Huggins, R.M. (2009). A Robust Approach to the Analysis of Repeated Measures. *Biometrics*, **49**, 255–268.

Jiang, J. (1996). REML estimation: asymptotic behavior and related topics. *Annals of Statistics*, **24**, 255-286.

Jiang, J. and Lahiri, P. (2002). A unified jacknife theory for empirical best prediction with *D*-estimation. *The Annals of Statistics*, **30**, 2720-2733.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1-96.

Kackar, R. and Harville, D. (1984). Approximations for standard errors of fixed and random effects in mixed models. *Journal of the American Statistical Association*, **79**, 853-862.

Kakamu, K. (2005). Bayesian estimation of a distance functional weight matrix model. *Economics Bulletin*, **3**, 1-6.

Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74-89.

Kokic, P., Chambers, R.L., Breckling, J., Beare, S. (1997). A measure of production performance. *Journal of Business and Economic Statistics*, **15**, 445-451.

LeSage, J.P. and Pace, R.K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, **140**, 190-214.

Li, Q. and Racine, J.S. (2007). Nonparametric Econometrics: Theory and Practice. *Princeton University Press*.

Lombardia M.J., Gonzalez-Manteiga W. and Prada-Sanchez J.M. (2003). Bootstrapping the Dorfman-Hall-Chambers-Dunstan estimator of a finite population distribution function. *Journal of Nonparametric. Statistics*, **16**, 63-90.

Molina, I., Salvati, N. and Pratesi, M. (2008). Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics*, **24**, 441-458.

Molina, I. and Rao, J.N.K. (2009). Small Area Estimation of Poverty Indicators. *The Canadian Journal of Statistics* **37**, 1–17.

Molina, I. and Rao, J.N.K. (2009). Small area estimation of poverty indicators. Working Paper 09-15, Statistics and Econometric Series 05, Universidad Carlos III de Madrid, http://halweb.uc3m.es/esp/Personal/per imjrWorkingPaperPoverty.pdf

Neri, L., Ballini, F. and Betti, G. (2005). Poverty and inequality in transition countries. *Statistics in Transition*, **7**, 135–157.

Newey, W.K. and powell, J.L. (1987). Asymmetric least squared estimation and testing. *Econometrica*, **55**, 819-47.

Nychka, D. and Saltzman, N. (1998). Design of air quality monitoring networks. In Kychka, Douglas Piegorsch, Walter and Cox (eds), *Case studies in environmental statistics*.

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G. and Breidt, F. J. (2008). Nonparametric Small Area Estimation Using Penalized Spline Regression. *Journal of the Royal Statistical Society, Series B*, **70**, 265-283.

Petrucci, A. and Salvati, N. (2006). Small Area Estimation considering Spatial Correlation in Watershed Erosion Assessment Survey. *Journal of Agricultural , Biological, and Environmental Statistics*, Vol. 11, 14,169-182.

Pfeffermann, D. and Buck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, **16**, 217—237.

Pfeffermann, D. (2002). Small Area Estimation - New Developments and Directions. *International Statistical Review*, **70**, 1, 125-143.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis*, **26**, 893-916.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163-171.

Pratesi, M., Salvati, N. (2008). Small Area Estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical Methods and Applications*, **17**, 113-141.

Pratesi, M., Salvati, N. (2009). Small Area Estimation in the presence of correlated random area effects. *Journal of Official Statistics*, **25**, 37-53.

Pratesi, M., Ranalli, M.G., Salvati, N. (2006). Nonparametric M-quantile regression via penalized splines. *ASA Proceedings on Survey Research Methods*, Alexandria, VA.

Pratesi, M., Ranalli, M.G., Salvati, N. (2008). Semiparametric M-quantile regression for estimating the proportion of acidic lakes in 8-digit HUCs of the Northeastern US. *Environmetrics*, **19**, 687-701.

Pratesi, M., Ranalli, M.G., Salvati, N. (2009). Nonparametric M-quantile regression using penalized splines. *Journal of Nonparametric Statistics*, **21**, 287-304.

R Development Core Team (2005). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. URL: http://www.R-project.org.

Rao, J.N.K., Kovar, J.G., Mantel, H.J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, **77**, 365-75.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, **25**, 175-186.

Rao, J.N.K. and Yu, M. (1994). Small area estimation by combining time series and cross sectional data. *Canadian Journal of Statistics*, **22**, 511—528.

Rao, J.N.K. (2003). *Small Area Estimation*. John Wiley.

Richardson, A.M. and Welsh, A.H.(1995). Robust Estimation in the Mixed Linear Model. *Biometrics*, **51**, 1429-1439.

Royall, R. M. (1976). The linear least squared prediction approach to two-stage sampling. *Journal of the American Statistical Association*, **71**, 657-664.

Royall, R.M. and Cumberland, W.G. (1978). Variance estimation in finite population sampling. *Journal of the American Statistical Association*, **73**, 351 - 58.

Ruppert, D., Wand, M.P., Carroll, R. (2003). *Semiparametric regression*. Cambridge University Press, Cambridge, New York.

Saei, A. and Chambers, R.L. (2005). Empirical best linear unbiased prediction for out of sample areas. In *S3Ri Methodology Working Papers*. Southampton Statistical Science Research Institute, pp. 1-15.

Salvati, N., Tzavidis, N., Pratesi, M., Chambers, R.L. (2008). Small area estimation via M-quantile geographically weighted regression. [Paper submitted for publication. A copy is available upon request].

Salvati, N. and Tzavidis, N. (2010). M-quantile GWR function. Available from *URL: http://www.dipstat.ec.unipi.it/persone/docenti/salvati/*.

Salvati, N., Chandra, H., Ranalli, M.G. and Chambers, R. (2010a). Small Area Estimation Using a Nonparametric Model Based Direct Estimator. *Computational Statistics and Data Analysis*, **54**, 2159-2171.

Salvati, N., Ranalli, M. G. and Pratesi, M. (2010b). Small area estimation of the mean using non-parametric M-quantile regression: a comparison when a linear mixed model does not hold. To appear in *Journal of Statistical Computation and Simulation*.

Särndal, C.E., Swensson, B. and Wretman J. (1992) *Model assisted survey sampling*, Springer-Verlag.

Sinha, S.K. and Rao, J.N.K. (2009). Robust Small Area Estimation. *Canadian Journal of Statistics*, **37**, 381-399.

Singh, B., Shukla, G. and Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, **31**, 183-195.

Street, J.O., Carroll, R.J. and Ruppert, D. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *American Statistician*, **42**, 152-154.

Tarozzi, A. and Deaton, A. (forthcoming). Using census and survey data to estimate poverty and in-equality for small areas. *Review of Economics and Statistics*.

Tzavidis, N., Marchetti, S., and Chambers, R.L. (2009). Robust prediction of small are means and quantiles. To appear in the Australian and New Zealand Journal of Statistics.

Tzavidis, N. and Chambers, R.L. (2006). Bias adjusted distribution estimation for small areas with outlying values. In *S3Ri Methodology Working Papers*. Southampton Statistical Science Research Institute, pp. 1-30.

Tzavidis, N. and Chambers, R. (2007). Robust prediction of small area means and distributions. *CCSR Working Paper 2007-08, University of Manchester.*

Tzavidis, N., Marchetti, S. and Chambers, R. (2010). Robust estimation of small area means and quantiles. *Australian and New Zealand Journal of Statistics* DOI 10.1111/j.1467-842X.2010.00572.x

Ugarte, M. D., Militino, A.F. and Goicoa, T. (2008). Prediction error estimators in Empirical Bayes disease mapping. *Environmetrics*, **19**, 287-300.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*. New York: Springer.

Wang, S. and Dorfman, A.H. (1996). A new estimator of the finite population distribution function. *Biometrika*, **83**, 639-52.

You, Y. and Rao, J.N.K. (2000). Hierarchical Bayes estimation of small area means using multi-level models. *Survey Methodology*, **26**, 173-181.

You, Y., Rao, J.N.K. and Gambino, J. (2001). Model-based unemployment raye estimation for the Canadian Labour Force Survey: a hierarchical approach. Technical report, Household Survey Method Division. Statistics Canada.

Welsh, A.H. and Ronchetti, E.(1998). Bias-calibrated estimation from sample surveys containing outliers. *Journal of the Royal Statistical Society, Series B*, **60**, 413-428.

Zimmerman, D. and Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, **44**, 27-43.