FIRST SMALL AREA ESTIMATION DEVELOPMENTS

Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Project/Contract No:	EU-FP7-SSH-2007-1
Grant agreement No:	217565
Work Package 2:	Small area estimation of poverty and inequality indicators
Document title:	First small area estimation developments
Date:	October 30^{th}, 2009
Type of document:	Deliverable D8
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
	Nicola Salvati (salvati@ec.unipi.it), UNIPI-DSMAE
	Nikos Tzavidis (nikos.tzavidis@manchester.ac.uk), CCSR
	Isabel Molina (imolina@est-econ.uc3m.es), UC3M
	Mara 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
	J.N.K. Rao (jrao@math.carleton.ca), Carleton University

Contents

Prologue

This report contains some first small area developments of the partners of the WP2 in the SAMPLE project. The target of the report is to illustrate with applications to real data some of the statistical methodology that is being developed within the SAMPLE project. The manuscript is organized in four 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 deals with area-level time 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. 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. An application to the Spanish Living Conditions Survey data is also given. The target of the application is to estimate poverty proportions and gaps in Spanish provinces by gender.

Chapter 3 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 as particular cases of non-linear parameters. The mean squared error of the proposed estimators is obtained by a parametric bootstrap for finite populations. The method is applied to the estimation of FGT poverty measures in Spanish provinces by gender.

Chapter 4 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 a first empirical evaluation for the estimation of the mean squared error for the mean and quantile estimates. Finally Chapter 4 describes the EU-SILC data and the Census data which are used to produce the small area estimates and it presents the first results.

This report has been coordinated by Domingo Morales (UMH). He has also been in charge of writing Chapters 1-2. Isabel Molina (UC3M) has been responsible for the elaboration of Chapter 3. Finally, Nikos Tzavidis (CCSR) and Monica Pratesi (UNIPI-DSMAE) have coordinated the production of the contents of Chapter 4.

Chapter 1

Linear mixed models

1.1 Linear mixed models with known variance

1.1.1 Introduction

We consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{1.1}$$

where $\mathbf{y}_{n\times 1}$ is the vector of observations, $\boldsymbol{\beta}_{p\times 1}$ 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,

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

depending on a parameter $\boldsymbol{\theta}$ containing the variance components. From (1.1) we obtain

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

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

1.1.2 Least squared estimation of $\boldsymbol{\beta}$

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

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

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

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

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

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

We observe that

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

By taking derivatives, we obtain

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

The normal equations are

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

and the solution is

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

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

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

1.1.3 Best linear unbiased prediction of a linear combination of effects

We look at the model (1.1) and define $\tau = \mathbf{a}_r'(\mathbf{X}_r\boldsymbol{\beta} + \mathbf{Z}_r\mathbf{u})$, 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.

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

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

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

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

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

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

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

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

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

In matrix form, the above equations are

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

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

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

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

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

where

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

is the least squared estimator of $\boldsymbol{\beta}$.

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

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

or equivalently

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

1.1.4 Best linear unbiased prediction of u

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

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

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

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

where $\mathbf{y} = (y_1, \ldots, y_n)'$ is the vector of sample observations, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ 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

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

with $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_{u_i}$, $i = 1, \ldots, m$, 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 $\mathbf{G}_0 = \mathbf{\Sigma}_e$ and $\mathbf{G}_i = \mathbf{Z}_i \mathbf{\Sigma}_{u_i} \mathbf{Z}_i'$, $i = 1, \ldots, m$.

(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

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

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

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

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

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

1.3 Maximum likelihood estimation

1.3.1 Description of the method

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

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

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

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

If $\widehat{\boldsymbol{\theta}}$ 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.

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

We know that

$$\frac{\partial \log |\mathbf{V}|}{\partial \sigma_i^2} = \operatorname{tr}\left\{ \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2} \right\} , \qquad (1.9)$$

$$\frac{\partial \mathbf{V}^{-1}}{\partial \sigma_i^2} = -\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1} . \qquad (1.10)$$

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

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

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

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} , \qquad (1.12)$$

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

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 $\boldsymbol{\theta}^0$. In each iteration, the Newton-Raphson method updates the estimator of $\boldsymbol{\theta}$ by using the formula

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

where $\mathbf{S}(\boldsymbol{\theta}^i)$ is the vector of derivatives and $\mathbf{H}(\boldsymbol{\theta}^i)$ is the Hessian matrix of $l(\boldsymbol{\theta})$, both calculated with the estimator obtained at the last iteration $\boldsymbol{\theta}^i$. 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.

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= -\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} , & (1.14) \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_i^2 \partial \boldsymbol{\beta}} &= \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma_i^2} = -\mathbf{X}'\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) , & (1.15) \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_j^2 \sigma_i^2} &= \frac{1}{2}\operatorname{tr}\{\mathbf{V}^{-1}\mathbf{G}_j\mathbf{V}^{-1}\mathbf{G}_i\} - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{G}_j\mathbf{V}^{-1}\mathbf{G}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) , & (1.16) \end{aligned}$$

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

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

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

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

and evaluated in $\boldsymbol{\theta}^i$. Taking expectations in (1.14)-(1.16), changing the sign and using the result

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

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

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

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

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

We get

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

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

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

Finally

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

1.3.2 Maximum likelihood with alternative parametrization

We consider the model (1.5) and the parameters

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

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

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

The likelihood function is

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

The components of the vector of scores are

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

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

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

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

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

Partial derivatives of the log-likelihood function are

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

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

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

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

$\mathbf{y}$ 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 $\boldsymbol{\beta}$. Let $\mathbf{K}_1$ be a matrix such that $\mathbf{K}_1\mathbf{X} = \mathbf{0}$. Therefore

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

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

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

The maximum likelihood estimator of $\boldsymbol{\beta}$ in (1.23) is

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

We define the transformed vector (normalized residual)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The elements of the Fisher information matrix are

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

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

$$\boldsymbol{\sigma}_{k+1} = \boldsymbol{\sigma}_k + F(\boldsymbol{\sigma}_k)^{-1}S(\boldsymbol{\sigma}_k)\,,$$

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

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

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

By taking partial derivatives with respect to $\boldsymbol{\beta}$, and equating to zero, we obtain

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

Therefore

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

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 $\boldsymbol{\beta}$, we get

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

that is the same value of $\mathbf{F}_{\boldsymbol{\beta\beta}}$ 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 $r \le p$. 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 $\boldsymbol{\varphi}' = (\sigma^2, \varphi_1, \ldots, \varphi_m)$, $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \boldsymbol{\varphi}')$ and $\mathbf{V} = \sigma^2\left(\boldsymbol{\Sigma}_e + \sum_{i=1}^{m} \varphi_i \mathbf{G}_i\right) = \sigma^2\boldsymbol{\Sigma}$. For the REML method, the log-likelihood associated to the this parametrization is

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

where $\mathbf{P} = \mathbf{K}(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K})^{-1}\mathbf{K}' = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$. 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

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

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

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

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 $\boldsymbol{\varphi}$ with

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

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 $\boldsymbol{\beta}$ 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 $\boldsymbol{\beta}$ 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, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, in the form

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}, \tag{1.26}$$

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 $\boldsymbol{\beta}$ in two groups of effects $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, without taking into account if they represent fixed or random effects. This issue will be considered later.

We apply the transformation

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

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

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

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

If we fit the model (1.27) under the assumption that $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ 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

$$SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mathbf{y}'\mathbf{M}\mathbf{y}, \tag{1.29}$$

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

$$SSR(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = SST - SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mathbf{y}'\boldsymbol{Q}\mathbf{y},$$

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

If we fit the submodel

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

under the assumption that $\boldsymbol{\beta}_1$ is a fixed effect, the residual sum of squares is

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

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

$$SSR(\boldsymbol{\beta}_1) = SST - SSE(\boldsymbol{\beta}_1) = \mathbf{y}'\boldsymbol{Q}_1\mathbf{y},$$

where $\boldsymbol{Q}_1 = \mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}$. 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

$$SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SSR(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - SSR(\boldsymbol{\beta}_1) = SSE(\boldsymbol{\beta}_1) - SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2).$$

To introduce the Henderson 3 method, we first calculate the expectation of $SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ and $SSR(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. 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 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\boldsymbol{\beta}$ may contain fixed or random effects, we have $E[\mathbf{y}] = \mathbf{X}E[\boldsymbol{\beta}]$ and $\mathrm{var}[\mathbf{y}] = \mathbf{X}\mathrm{var}[\boldsymbol{\beta}]\mathbf{X}' + \sigma_0^2\mathbf{W}^{-1}$. From (1.24), we obtain

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

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

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

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

$$E[SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)] = E[\mathbf{y}'\mathbf{M}\mathbf{y}] = \mathrm{tr}\left(\mathbf{X}'\mathbf{M}\mathbf{X}E[\boldsymbol{\beta}\boldsymbol{\beta}']\right) + \sigma_0^2\mathrm{tr}\left(\mathbf{M}\mathbf{W}^{-1}\right).$$

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

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

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

$$E[SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)] = \sigma_0^2\,[n - \mathrm{r}(\mathbf{X})] \tag{1.32}$$

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

$$E[SSR(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)] = E[SST] - E[SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)] = \operatorname{tr}\left(\mathbf{X}'\mathbf{W}\mathbf{X}E[\boldsymbol{\beta}\boldsymbol{\beta}']\right) + \sigma_0^2\mathrm{r}(\mathbf{X}).$$

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

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

From (1.30) and (1.24) we obtain

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

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

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

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

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

and

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

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

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

We observe that $E\left[SSR(\boldsymbol{\beta}_2|\,\boldsymbol{\beta}_1)\right]$ is simply a function of $E\left[\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\right]$ and σ_0^2. It does not depend on $E\left[\boldsymbol{\beta}_1\boldsymbol{\beta}_1'\right]$ and $E\left[\boldsymbol{\beta}_1\boldsymbol{\beta}_2'\right]$. We also observe that (1.36) has been obtained without doing assumptions about the form of $E\left[\boldsymbol{\beta}\boldsymbol{\beta}'\right]$. Therefore (1.36) holds for any structure of covariance matrix of $\boldsymbol{\beta}$.

Let us consider again the model (1.5)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and so on.

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

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

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 $\boldsymbol{\beta}$ 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 $N(\mathbf{x}_d\,\boldsymbol{\beta}, \sigma_u^2)$, where $\boldsymbol{\beta}$ is a vector of dimension p containing the regression parameters, i.e. $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_D)' \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_u)$ with $\boldsymbol{\Sigma}_u = \sigma_u^2 \mathbf{I}_D$.

3. Let $\overline{\mathbf{y}} = (\overline{y}_{1\cdot}, \ldots, \overline{y}_{D\cdot})'$ be a vector of direct estimators of $\boldsymbol{\mu}$ with distribution $N(\boldsymbol{\mu}, \mathbf{V}_e)$, where $\mathbf{V}_e = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_D^2)$ and the diagonal elements σ_d^2 are assumed to be known.

The area-level Fay-Herriot model is

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

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 $N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ respectively. If we write (1.41) in the form $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, 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 $\mathbf{Z} = \mathbf{I}_D$, $\operatorname{tr}(\mathbf{Z}'\mathbf{Z}) = D$, $r(\boldsymbol{X}, \boldsymbol{Z}) = D$, $Cov[\mathbf{y}, \mathbf{u}] = \mathbf{Z}\boldsymbol{\Sigma}_u$,

$$\mathbf{V} = \operatorname{var}(\mathbf{y}) = \mathbf{Z}\boldsymbol{\Sigma}_u\mathbf{Z}' + \mathbf{V}_e = \boldsymbol{\Sigma}_u + \mathbf{V}_e = \operatorname{diag}(\sigma_u^2 + \sigma_1^2, \ldots, \sigma_u^2 + \sigma_D^2),$$

and

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

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

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

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

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

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

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

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

Proposition 1.6.1. The best predictor of $\boldsymbol{\mu}_d$ is

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

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

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

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

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

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

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.

$$\widehat{\overline{Y}}_d^{FH} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2+\sigma_d^2}\,\overline{y}_{d\cdot} + \frac{\sigma_d^2}{\hat{\sigma}_u^2+\sigma_d^2}\,\mathbf{x}_d\widehat{\boldsymbol{\beta}} \quad (1.43)$$

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

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

with $\hat{\sigma}_d^2 = \widehat{V}\,(\overline{y}_{d\cdot})$, $d = 1,\dots,D$, 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 $\tilde{u}_d = \overline{y}_d - \mathbf{x}_d\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{y}} = \left(\sum_{d=1}^{D}\mathbf{x}_d'\mathbf{x}_d\right)^{-1}\left(\sum_{d=1}^{D}\mathbf{x}_d'\overline{y}_d\right)$.

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 $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, with covariance matrix $\mathbf{V} = \text{diag}_{1\leq d\leq D}(\sigma_u^2+\sigma_d^2)$. The log-likelihood is

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

The partial derivatives of the log-likelihood are

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

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

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

The components of the Fisher information matrix are

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

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

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

The updating formulas of the Fisher-scoring algorithm are

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

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

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

and

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

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

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

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 $\widehat{\boldsymbol{\beta}}_0=\widetilde{\boldsymbol{\beta}}$, where $\widehat{\sigma}_u^2$ and $\widetilde{\boldsymbol{\beta}}$ are the moment estimators given by (1.45).

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

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

whre

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

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

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

Alternatively the following algorithm can be used.

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

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

3. Stop when $|\widehat{\sigma}^2_{u,k}-\widehat{\sigma}^2_{u,k-1}|<\varepsilon$ and $\left[(\widehat{\boldsymbol{\beta}}_k-\widehat{\boldsymbol{\beta}}_{k-1})'(\widehat{\boldsymbol{\beta}}_k-\widehat{\boldsymbol{\beta}}_{k-1})\right]^{1/2}<\varepsilon$. Output: $\widehat{\boldsymbol{\beta}}_{ML}=\widehat{\boldsymbol{\beta}}_k$, $\widehat{u}_d=\widehat{u}_{d,k}$ 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 $\{1,\ldots,N\}$ 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 $\tau=\mathbf{a}'_r(\mathbf{X}_r\boldsymbol{\beta}+\mathbf{Z}_r\mathbf{u})$, 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. $\boldsymbol{\beta}$, $\theta_0, \theta_1, \ldots, \theta_m$ are known,

2. $\theta_0, \theta_1, \ldots, \theta_m$ are known, $\boldsymbol{\beta}$ is unknown,

3. All the model parameters are unknown.

All the model parameters are known

Assume that $\boldsymbol{\beta}$ and $\theta_0, \theta_1, \ldots, \theta_m$ are known. The BLUP of τ is

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

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

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

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 $\boldsymbol{\beta}$ 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

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

where

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

It holds that

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

All the parameters are unknown

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

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

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 $E[\tau(\boldsymbol{\theta})]$ is finite and $\widehat{\boldsymbol{\theta}}$ is an even and translation invariante (as the Henderson 3, Ml and REML estimators are), then $\widehat{\tau}_{eblup} = \tau(\widehat{\boldsymbol{\theta}})$ 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 $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_m)$ of $\boldsymbol{\theta}$ and define $\mathbf{d}(\boldsymbol{\theta}) = (d_0(\boldsymbol{\theta}), d_1(\boldsymbol{\theta}), \ldots, d_m(\boldsymbol{\theta}))'$, where

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

A first order Taylor series expansion of $\tau(\boldsymbol{\gamma})$ around $\boldsymbol{\theta}$ yields to

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

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

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

Let us now assume that $\widehat{\boldsymbol{\theta}}$ 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

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

Further, it holds

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

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

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

Now we have

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

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

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

and therefore

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

The second summand in (1.48) can be written as

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

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

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

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

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

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

Prasad and Rao (1990) gave the new approximation

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

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

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

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

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

1.7.2 Mean squared error estimation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

with

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

In this case the components of model (1.1) can be written in the form $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_m)'$, $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_m)'$, $\mathbf{Z} = \operatorname{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_m)'$, $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_m)'$, $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_m)'$, 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

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

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

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

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

$$\mathbf{V}_i^{(j)} = \frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\theta}_j} = -\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_j}\mathbf{V}_i^{-1} \quad \text{and} \quad \mathcal{I}_{jk}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^m tr\left[\left(\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_j}\right)\left(\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}_k}\right)\right].$$

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 $E[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = 0$. 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

Area-level time models

2.1 Area-level model with time correlated effects

2.1.1 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

$$y_{dt} = \mathbf{x}_{dt}\boldsymbol{\beta} + v_d + u_{dt} + e_{dt}, \quad d = 1, \ldots, D, \quad t = 1, \ldots, T, \tag{2.1}$$

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 $v_1, \ldots, v_D$ are i.i.d. normal, $(u_{d1}, \ldots, u_{dT})$'s follow i.i.d. AR(1) processes (i.e. they follow autoregressive processes of order 1), $e_{11}, \ldots, e_{DT}$ are i.i.d. normal, and the v_d's, the $(u_{d1}, \ldots, u_{dT})$'s and the e_{dt}'s are independent.

In this section we introduce a model that it is related to the model (2.1) 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

$$y_{dt} = \mathbf{x}_{dt}\boldsymbol{\beta} + u_{dt} + e_{dt}, \quad d = 1, \ldots, D, \quad t = 1, \ldots, m_d, \tag{2.2}$$

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 $(u_{d1}, \ldots, u_{dm_d})$, $d = 1, \ldots, D$, 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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{2.3}$$

where $\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$, $\mathbf{y}_d = \underset{1\le t\le m_d}{\text{col}}(y_{dt})$, $\mathbf{u} = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_d)$, $\mathbf{u}_d = \underset{1\le t\le m_d}{\text{col}}(u_{dt})$, $\mathbf{e} = \underset{1\le d\le D}{\text{col}}(\mathbf{e}_d)$, $\mathbf{e}_d = \underset{1\le t\le m_d}{\text{col}}(e_{dt})$, $\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d = \underset{1\le t\le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \underset{1\le i\le p}{\text{col}'}(x_{dti})$, $\boldsymbol{\beta} = \underset{1\le i\le p}{\text{col}}(\beta_i)$, $\mathbf{Z} = \mathbf{I}_{M\times M}$ 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 componentes are known, then the BLUE of $\boldsymbol{\beta}$ and the BLUP of $\mathbf{u}$ are

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}),$$

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 $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ we apply the formulas

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{y}_d\right), \quad \widehat{\mathbf{u}} = \sigma_u^2 \underset{1\le d\le D}{\text{col}}\left(\Omega_d(\rho)\mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\boldsymbol{\beta}})\right).$$

2.1.2 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 $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\sigma_u^2, \rho)$, $\mathbf{V}_1 = \frac{\partial \mathbf{V}}{\partial \sigma_u^2} = \underset{1\leq d\leq D}{\text{diag}}\,(\Omega_d(\rho))$ and $\mathbf{V}_2 = \frac{\partial \mathbf{V}}{\partial \rho} = \sigma_u^2 \underset{1\leq d\leq D}{\text{diag}}\,(\dot{\Omega}_d(\rho))$. 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

$$F_{ab} = \frac{1}{2}\,\text{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a, b = 1, 2.$$

We use the Fisher-scoring algorithm to calculate the REML estimates of $\boldsymbol{\theta}$. The updating formula is

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + \mathbf{F}^{-1}(\boldsymbol{\theta}^k)\mathbf{S}(\boldsymbol{\theta}^k).$$

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 $\boldsymbol{\beta}$ is calculated by applying the formula

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

The asymptotic distributions of the REML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are

$$\hat{\boldsymbol{\theta}} \sim N_2(\boldsymbol{\theta}, \mathbf{F}^{-1}(\boldsymbol{\theta})), \quad \hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

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 $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\kappa}$, $\mathbf{F}^{-1}(\boldsymbol{\theta}^{\kappa}) = (\nu_{ab})_{a,b=1,2}$, $(\mathbf{X}'\mathbf{V}^{-1}(\boldsymbol{\theta}^{\kappa})\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\ldots,p}$, 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.

$$
\begin{aligned}
\mathbf{Q} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D} \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}, \\
\mathbf{P} &= \operatorname*{diag}_{1\le d\le D}(\mathbf{V}_d^{-1}) - \operatorname*{col}_{1\le d\le D}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\operatorname*{col'}_{1\le d\le D}(\mathbf{X}_d'\mathbf{V}_d^{-1}), \\
\mathbf{P}\mathbf{V}_a &= \operatorname*{diag}_{1\le d\le D}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \operatorname*{col}_{1\le d\le D}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\operatorname*{col'}_{1\le d\le D}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}), \\
\operatorname{tr}(\mathbf{P}\mathbf{V}_a) &= \sum_{d=1}^{D}\operatorname{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}) - \sum_{d=1}^{D}\operatorname{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q}), \\
\operatorname{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b) &= \sum_{d=1}^{D}\operatorname{tr}(\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}) - 2\sum_{d=1}^{D}\operatorname{tr}(\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\mathbf{Q}) \\
&+ \operatorname{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{bd}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\right\}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y} &= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)' \\
&- \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{y}_d\right) \\
&+ \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ad}\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'.
\end{aligned}
$$

Finally, the derivative of matrix $\Omega_d(\rho)$ with respect to ρ is

$$
\dot{\Omega}_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix}
0 & 1 & \dots & \dots & (m_d-1)\rho^{m_d-2} \\
1 & 0 & \ddots & & (m_d-2)\rho^{m_d-3} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
(m_d-2)\rho^{m_d-3} & & \ddots & 0 & 1 \\
(m_d-1)\rho^{m_d-2} & \dots & \dots & 1 & 0
\end{pmatrix} + \frac{2\rho\Omega_d(\rho)}{(1-\rho^2)^2}.
$$

2.1.3 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 $\mathbf{a} = \operatorname*{col}_{1\le \ell\le D}(\operatorname*{col}_{1\le k\le m_\ell}(\delta_{d\ell}\delta_{tk}))$ 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

$$
MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3(\boldsymbol{\theta}),
$$

where $\boldsymbol{\theta} = (\sigma_u^2, \rho)$,

$$\begin{aligned} g_1(\boldsymbol{\theta}) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\ g_2(\boldsymbol{\theta}) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}], \\ g_3(\boldsymbol{\theta}) &\approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right]\right\} \end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\boldsymbol{\theta}}) + g_2(\hat{\boldsymbol{\theta}}) + 2g_3(\hat{\boldsymbol{\theta}}).$$

Calculation of $g_1(\boldsymbol{\theta})$

In the formula of $g_1(\boldsymbol{\theta}) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, we have that $\mathbf{Z} = \mathbf{I}_{M\times M}$, and

$$\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \sigma_u^2 \underset{1\leq d\leq D}{\text{diag}}\,(\Omega_d(\rho)) - \sigma_u^4 \underset{1\leq d\leq D}{\text{diag}}\,(\Omega_d(\rho)) \underset{1\leq d\leq D}{\text{diag}}\,(\mathbf{V}_d^{-1}) \underset{1\leq d\leq D}{\text{diag}}\,(\Omega_d(\rho)).$$

Let us write $\Omega_d = \Omega_d(\rho)$ and $\mathbf{a}_d = \underset{1\leq k\leq m_d}{\text{col}}\,(\delta_{tk})$. Then, $g_1(\boldsymbol{\theta})$ can be expressed in the form

$$g_1(\boldsymbol{\theta}) = \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d.$$

Calculation of $g_2(\boldsymbol{\theta})$

We have that $g_2(\boldsymbol{\theta}) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}]$, where

$$\begin{aligned} \mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X} &= \left[\sigma_u^2 \underset{1\leq d\leq D}{\text{diag}}\,(\Omega_d) - \sigma_u^4 \underset{1\leq d\leq D}{\text{diag}}\,(\Omega_d) \underset{1\leq d\leq D}{\text{diag}}\,(\mathbf{V}_d^{-1}) \underset{1\leq d\leq D}{\text{diag}}\,(\Omega_d)\right] \underset{1\leq d\leq D}{\text{diag}}\,(\mathbf{V}_{ed}^{-1}) \underset{1\leq d\leq D}{\text{col}}(\mathbf{X}_d) \\ &= \sigma_u^2 \underset{1\leq d\leq D}{\text{col}}(\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_u^4 \underset{1\leq d\leq D}{\text{col}}(\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d). \end{aligned}$$

Therefore

$$\begin{aligned} g_2(\boldsymbol{\theta}) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_u^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d + \sigma_u^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\right]. \end{aligned}$$

Calculation of $g_3(\boldsymbol{\theta})$

We have that

$$g_3(\boldsymbol{\theta}) \approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right]\right\},$$

where

$$\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \sigma_u^2\mathbf{a}' \underset{1\leq \ell\leq D}{\text{diag}}\,(\Omega_\ell) \underset{1\leq \ell\leq D}{\text{diag}}\,(\mathbf{V}_\ell^{-1}) = \sigma_u^2 \underset{1\leq \ell\leq D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell\Omega_\ell\mathbf{V}_\ell^{-1}).$$

It holds that

$$\begin{aligned}
\frac{\partial \mathbf{b}'}{\partial \sigma_u^2} &= \underset{1\leq \ell \leq D}{\mathrm{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell \mathbf{V}_\ell^{-1}) - \sigma_u^2 \underset{1\leq \ell \leq D}{\mathrm{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell \mathbf{V}_\ell^{-1}\mathbf{V}_{\ell u}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell u} = \frac{\partial \mathbf{V}_\ell}{\partial \sigma_u^2} = \Omega_\ell, \\
\frac{\partial \mathbf{b}'}{\partial \rho} &= \sigma_u^2 \underset{1\leq \ell \leq D}{\mathrm{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\dot{\Omega}_\ell \mathbf{V}_\ell^{-1}) - \sigma_u^2 \underset{1\leq \ell \leq D}{\mathrm{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell \mathbf{V}_\ell^{-1}\mathbf{V}_{\ell \rho}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell \rho} = \frac{\partial \mathbf{V}_\ell}{\partial \rho} = \sigma_u^2\dot{\Omega}_\ell\,.
\end{aligned}$$

We define

$$\begin{aligned}
q_{11} = \frac{\partial \mathbf{b}'}{\partial \sigma_u^2} \underset{1\leq \ell \leq D}{\mathrm{diag}}(\mathbf{V}_\ell)\left(\frac{\partial \mathbf{b}'}{\partial \sigma_u^2}\right)' &= \mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d - 2\sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&+ \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d, \\
q_{12} = \frac{\partial \mathbf{b}'}{\partial \sigma_u^2} \underset{1\leq \ell \leq D}{\mathrm{diag}}(\mathbf{V}_\ell)\left(\frac{\partial \mathbf{b}'}{\partial \rho}\right)' &= \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\
&- \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d, \\
q_{22} = \frac{\partial \mathbf{b}'}{\partial \rho} \underset{1\leq \ell \leq D}{\mathrm{diag}}(\mathbf{V}_\ell)\left(\frac{\partial \mathbf{b}'}{\partial \rho}\right)' &= \sigma_u^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - 2\sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d \\
&+ \sigma_u^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d.
\end{aligned}$$

Finally

$$g_3(\boldsymbol{\theta}) = \mathrm{tr}\left\{\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22}\end{pmatrix}\begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22}\end{pmatrix}^{-1}\right\},$$

where F_{ab} is the element of the REML Fisher information matrix.

2.1.4 Simulations

Simulation 1

For $d = 1, \dots, D$, $t = 1, \dots, m_d$, the explanatory and target variables are

$$\begin{aligned}
x_{dt} &= (b_{dt} - a_{dt})U_{dt} + a_{dt},\ U_{dt} = \frac{t}{m_d + 1},\ a_{dt} = 1,\ b_{dt} = 1 + \frac{1}{D}\left(m_d(d-1) + t\right), \\
y_{dt} &= \beta_1 + \beta_2 x_{dt} + u_{dt} + e_{dt},\ \beta_1 = 0,\ \beta_2 = 1,
\end{aligned}$$

where $e_{dt} \sim N(0, \sigma_{dt}^2)$ and

$$\sigma_{dt}^2 = \frac{(\alpha_1 - \alpha_0)\left(m_d(d-1) + t - 1\right)}{M - 1} + \alpha_0, \quad \alpha_0 = 0.8,\ \alpha_1 = 1.2.$$

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 $d \leq D_A$, $\varepsilon_{dt} \sim N(0, \sigma_B^2)$ 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 $\hat{\tau}^{(k)} \in \{\hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\sigma}_u^{2(k)}, \hat{\rho}^{(k)}\}$ 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 2.1.4.1.

D	50	100	200	300	400	500
m_d	5	5	5	5	5	5
m	250	500	1000	1500	2000	2500

Table 2.1.4.1: Sample sizes.

Table 2.1.4.2 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 2.1.4.2. Results of simulation experiment 1.

Table 2.1.4.2 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 2.1.4.3 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 2.1.4.3. Results of simulation experiment 2.

Tables 2.1.4.3 shows that BIAS and MSE tend to zero as D increases.

2.2 Area-level model with independent time effects

2.2.1 Introduction

This section presents a simplification of model (2.2) 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 (2.2), but assuming that $\rho=0$. Parameter estimates of model (2.4) can also be used as seeds for an iterative fitting method in model (2.2). We assume that

$$y_{dt}=\mathbf{x}_{dt}\boldsymbol{\beta}+u_{dt}+e_{dt},\quad d=1,\dots,D,\quad t=1,\dots,m_d, \tag{2.4}$$

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 (2.4) can be alternatively written in the form

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}+\mathbf{e}, \tag{2.5}$$

where $\mathbf{y}=\underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$, $\mathbf{y}_d=\underset{1\le t\le m_d}{\text{col}}(y_{dt})$, $\mathbf{u}=\underset{1\le d\le D}{\text{col}}(\mathbf{u}_d)$, $\mathbf{u}_d=\underset{1\le t\le m_d}{\text{col}}(u_{dt})$, $\mathbf{e}=\underset{1\le d\le D}{\text{col}}(\mathbf{e}_d)$, $\mathbf{e}_d=\underset{1\le t\le m_d}{\text{col}}(e_{dt})$, $\mathbf{X}=\underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d=\underset{1\le t\le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt}=\underset{1\le i\le p}{\text{col}'}(x_{dti})$, $\boldsymbol{\beta}=\underset{1\le i\le p}{\text{col}}(\beta_i)$, $\mathbf{Z}=\mathbf{I}_M$,

$M = \sum_{d=1}^{D} m_d$. 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

$$\mathbf{V}_u = \sigma_u^2 \mathbf{I}_M, \quad \mathbf{I}_M = \underset{1\leq d\leq D}{\text{diag}}(\mathbf{I}_{m_d}), \quad \mathbf{V}_e = \underset{1\leq d\leq D}{\text{diag}}(\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1\leq t\leq m_d}{\text{col}}(\sigma_{dt}^2),$$

and known variances σ_{dt}^2.

The BLUE of $\boldsymbol{\beta}$ and the BLUP of $\mathbf{u}$ are

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}),$$

where

$$var(\mathbf{y}) = \mathbf{V} = \sigma_u^2 \underset{1\leq d\leq D}{\text{diag}}(\mathbf{I}_{m_d}) + \mathbf{V}_e = \underset{1\leq d\leq D}{\text{diag}}(\sigma_u^2\mathbf{I}_{m_d} + \mathbf{V}_{ed}) = \underset{1\leq d\leq D}{\text{diag}}(\mathbf{V}_d).$$

To calculate $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ we apply the formulas

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{d=1}^{D} \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1} \left(\sum_{d=1}^{D} \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{y}_d\right), \quad \widehat{\mathbf{u}} = \sigma_u^2 \underset{1\leq d\leq D}{\text{col}} \left(\mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\boldsymbol{\beta}})\right).$$

2.2.2 The Henderson 3 method

For the linear mixed model

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

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

$$\begin{aligned} E[SSE(\boldsymbol{\beta}, \mathbf{u})] &= \sigma_e^2[n - \text{rg}(\mathbf{X}, \mathbf{Z})], \\ E[SSE(\mathbf{u}|\boldsymbol{\beta})] &= \text{tr}\left\{\mathbf{Z}'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}']\mathbf{W}\mathbf{Z}\right\}\sigma_u^2 + \sigma_e^2[\text{rg}(\mathbf{X}, \mathbf{Z}) - \text{rg}(\mathbf{X})], \end{aligned}$$

where $SSE(\mathbf{u}|\boldsymbol{\beta}) = SSE(\boldsymbol{\beta}) - SSE(\boldsymbol{\beta}, \mathbf{u})$ y $SSE(\boldsymbol{\beta})$, $SSE(\boldsymbol{\beta}, \mathbf{u})$ are the sum of squares of residuals of the fixed effect models $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ and $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$ respectively. It hold that

$$\begin{aligned} E[SSE(\boldsymbol{\beta})] &= E[SSE(\mathbf{u}|\boldsymbol{\beta})] + E[SSE(\boldsymbol{\beta}, \mathbf{u})] \\ &= \text{tr}\left\{\mathbf{Z}'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}']\mathbf{W}\mathbf{Z}\right\}\sigma_u^2 + \sigma_e^2[n - \text{rg}(\mathbf{X})]. \end{aligned}$$

The Henderson 3 estimators of σ_u^2 is

$$\widehat{\sigma}_{uH}^2 = \frac{SSE(\boldsymbol{\beta}) - \sigma_e^2[n - \text{rg}(\mathbf{X})]}{\text{tr}\left\{\mathbf{Z}'\mathbf{W}[\mathbf{W}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}']\mathbf{W}\mathbf{Z}\right\}},$$

where $SSE(\boldsymbol{\beta}) = \mathbf{y}'\mathbf{P}_2\mathbf{y}$ 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 (2.3) 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 $\mathrm{rg}(\mathbf{X}) = p$. Therefore,

$$\widehat{\sigma}_{uH}^2 = \frac{\mathbf{y}'\mathbf{P}_2\mathbf{y} - (M-p)}{\mathrm{tr}\{\mathbf{P}_2\}},$$

where

$$\begin{aligned}
\mathbf{Q}_2 &= (\mathbf{X}'\mathbf{V}_e^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)^{-1}, \\
\mathbf{P}_2 &= \mathbf{V}_e^{-1} - \mathbf{V}_e^{-1}\mathbf{X}\mathbf{Q}_2\mathbf{X}'\mathbf{V}_e^{-1} = \underset{1\le d\le D}{\mathrm{diag}}\,(\mathbf{V}_{ed}^{-1}) - \underset{1\le d\le D}{\mathrm{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d)\mathbf{Q}_2\underset{1\le d\le D}{\mathrm{col}'}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1}), \\
\mathrm{tr}\{\mathbf{P}_2\} &= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\sigma_{dt}^{-2} - \sum_{d=1}^{D}\mathrm{tr}\{\mathbf{X}_d'\mathbf{V}_{ed}^{-2}\mathbf{X}_d\mathbf{Q}_2\},
\end{aligned}$$

$$\begin{aligned}
\mathbf{y}'\mathbf{P}_2\mathbf{y} &= \underset{1\le d\le D}{\mathrm{col}'}(\mathbf{y}_d)\left[\underset{1\le d\le D}{\mathrm{diag}}\,(\mathbf{V}_{ed}^{-1}) - \underset{1\le d\le D}{\mathrm{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d)\mathbf{Q}_2\underset{1\le d\le D}{\mathrm{col}'}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1})\right]\underset{1\le d\le D}{\mathrm{col}}(\mathbf{y}_d) \\
&= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\sigma_{dt}^{-2}y_{dt}^2 - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)\mathbf{Q}_2\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)'.
\end{aligned}$$

2.2.3 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

$$S = S(\theta) = \frac{\partial l_{REML}}{\partial\theta} = -\frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_u) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_u\mathbf{P}\mathbf{y} = -\frac{1}{2}\mathrm{tr}(\mathbf{P}) + \frac{1}{2}\mathbf{y}'\mathbf{P}^2\mathbf{y}.$$

The minus expectation of the second order derivative of l_{REML} with respect to $\theta = \sigma_u^2$ is

$$F = F(\theta) = \frac{1}{2}\,\mathrm{tr}(\mathbf{P}\mathbf{V}_u\mathbf{P}\mathbf{V}_u) = \frac{1}{2}\,\mathrm{tr}(\mathbf{P}^2). \tag{2.6}$$

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 $\boldsymbol{\beta}$ is

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

The asymptotic distributions of the REML estimators of σ_u^2 and $\boldsymbol{\beta}$ are

$$\hat{\sigma}_u^2 \sim N_2(\boldsymbol{\theta}, F^{-1}(\sigma_u^2)), \quad \hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

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.

$$\begin{aligned}
\mathbf{Q} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D} \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}, \\
\mathbf{P} &= \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}) - \underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{X}_d)\mathbf{Q}\underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_d^{-1}), \\
\text{tr}(\mathbf{P}) &= \sum_{d=1}^{D}\text{tr}(\mathbf{V}_d^{-1}) - \sum_{d=1}^{D}\text{tr}(\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\mathbf{Q}),
\end{aligned}$$

$$\begin{aligned}
\text{tr}(\mathbf{P}^2) &= \sum_{d=1}^{D}\text{tr}(\mathbf{V}_d^{-2}) - 2\sum_{d=1}^{D}\text{tr}(\mathbf{X}_d'\mathbf{V}_d^{-3}\mathbf{X}_d\mathbf{Q}) \\
&+ \text{tr}\left\{\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\right\}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{y}'\mathbf{P}^2\mathbf{y} &= \sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-2}\mathbf{y}_d - 2\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{y}_d\right) \\
&+ \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-2}\mathbf{X}_d\right)\mathbf{Q}\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)'.
\end{aligned}$$

2.2.4 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

$$\begin{aligned} g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\ g_2(\theta) &= [\mathbf{a}'\mathbf{X}-\mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a}-\mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}] \qquad \text{y} \\ g_3(\theta) &\approx \operatorname{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right]\right\} \end{aligned}$$

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

$$\mathbf{T}=\mathbf{V}_u-\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u=\sigma_u^2\mathbf{I}_M-\sigma_u^4\underset{1\le d\le D}{\operatorname{diag}}(\mathbf{V}_d^{-1}).$$

We define $\mathbf{a}_d=\underset{1\le k\le m_d}{\operatorname{col}}(\delta_{tk})$. 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

$$\begin{aligned} \mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X} &= \left[\sigma_u^2\mathbf{I}_M-\sigma_u^4\underset{1\le d\le D}{\operatorname{diag}}(\mathbf{V}_d^{-1})\right]\underset{1\le d\le D}{\operatorname{diag}}(\mathbf{V}_{ed}^{-1})\underset{1\le d\le D}{\operatorname{col}}(\mathbf{X}_d) \\ &= \sigma_u^2\underset{1\le d\le D}{\operatorname{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d)-\sigma_u^4\underset{1\le d\le D}{\operatorname{col}}(\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d). \end{aligned}$$

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
$$g_3(\sigma_u^2) \approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right]\right\},$$
where
$$\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \sigma_u^2\mathbf{a}'\underset{1\le\ell\le D}{\operatorname{diag}}(\mathbf{V}_\ell^{-1}) = \sigma_u^2\underset{1\le\ell\le D}{\operatorname{col}'}(\delta_{d\ell}\mathbf{a}'_\ell\mathbf{V}_\ell^{-1}).$$
It holds that
$$\frac{\partial\mathbf{b}'}{\partial\sigma_u^2} = \underset{1\le\ell\le D}{\operatorname{col}'}(\delta_{d\ell}\mathbf{a}'_\ell\mathbf{V}_\ell^{-1}) - \sigma_u^2\underset{1\le\ell\le D}{\operatorname{col}'}(\delta_{d\ell}\mathbf{a}'_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell u}\mathbf{V}_\ell^{-1}),\quad \mathbf{V}_{\ell u} = \frac{\partial\mathbf{V}_\ell}{\partial\sigma_u^2} = \mathbf{I}_{m_\ell}.$$
We define
$$\begin{aligned} q &= \frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\underset{1\le\ell\le D}{\operatorname{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\right)' = \mathbf{a}'_d\mathbf{V}_d^{-1}\mathbf{a}_d - 2\sigma_u^2\mathbf{a}'_d\mathbf{V}_d^{-2}\mathbf{a}_d + \sigma_u^4\mathbf{a}'_d\mathbf{V}_d^{-3}\mathbf{a}_d \\ &= \frac{1}{\sigma_u^2+\sigma_{dt}^2} - \frac{2\sigma_u^2}{\left(\sigma_u^2+\sigma_{dt}^2\right)^2} + \frac{\sigma_u^4}{\left(\sigma_u^2+\sigma_{dt}^2\right)^3}, \end{aligned}$$
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. (2.6)).

2.2.5 Simulations

Simulation 1

For $d = 1,\ldots,D$, $t = 1,\ldots,m_d$, The explanatory and target variables are
$$\begin{aligned} x_{dt} &= (b_{dt}-a_{dt})U_{dt} + a_{dt},\ U_{dt} = \frac{t}{m_d+1},\ a_{dt} = 1,\ b_{dt} = 1 + \frac{1}{D}\left(m_d(d-1)+t\right), \\ y_{dt} &= \beta_1 + \beta_2 x_{dt} + u_{dt} + e_{dt},\ \beta_1 = 0,\ \beta_2 = 1, \end{aligned}$$
where $u_{dt} \sim N(0,\sigma_u^2)$, $e_{dt} \sim N(0,\sigma_{dt}^2)$, $\sigma_u^2 = 1$ and
$$\sigma_{dt}^2 = \frac{(\alpha_1-\alpha_0)\left(m_d(d-1)+t-1\right)}{M-1} + \alpha_0,\quad \alpha_0 = 0.8,\ \alpha_1 = 1.2.$$
The first simulation experiment has the following steps:

1. Repeat $K = 10^4$ times $(k = 1,\ldots,K)$
   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 $\hat{\mu}_{dt}$, $d = 1, \ldots, D$, $t = 1, \ldots, m_d$, 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 2.2.5.1.

D	50	100	200	300	400	500
m_d	5	5	5	5	5	5
M	250	500	1000	1500	2000	2500

Table 2.2.5.1: Sample sizes.

The Table 2.2.5.2 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 2.2.5.2. Results of simulation experiment 1.

The Table 2.2.5.2 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 ($k = 1, \ldots, K$)

   1.1. Generate the sample $(y_{dt}^{(k)}, \mathbf{x}_{dt})$, $d = 1, \ldots, D$, $t = 1, \ldots, m_d$.

1.2. Calculate $\hat{\sigma}_u^{2(k)}$ and $mse_{dt}^{(k)} = mse_{dt}(\hat{\sigma}_u^{2(k)})$.

2. Calculate the performance measures of estimator mse_{dt}

$$B_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt}), \quad E_{dt} = \frac{1}{K}\sum_{k=1}^{K}(mse_{dt}^{(k)} - MSE_{dt})^2, \quad d = 1, \dots, D,$$

$$B = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} B_{dt}, \quad E = \frac{10^3}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} E_{dt}.$$

The Table 2.2.5.3 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 2.2.5.3. Results of simulation experiment 2.

The Table 2.2.5.3 shows that the BIAS and the MSE tends to zero as D increases.

2.2.6 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 (2.2) with correlation parameter ρ instead of the simplified model (2.4) restricted to $\rho = 0$. 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 $e_{dt} \sim N(0, \sigma_{dt}^2)$, $\sigma_{dt}^2 = \alpha_0 + \frac{(\alpha_1 - \alpha_0)(m_d(d-1)+t-1)}{M-1}$, $\alpha_0 = 0.8$ and $\alpha_1 = 1.2$. For $d = 1, \dots, D$, the random vectors $(u_{d1}, \dots, u_{dm_d})$ are generated as follows:

$$u_{d1} = (1 - \rho^2)^{-1/2}\varepsilon_{d1}, \quad u_{dt} = \rho u_{dt-1} + \varepsilon_{dt}, \ t = 2, \dots, m_d,$$

where $\varepsilon_{dt} \sim N(0, \sigma_u^2)$, $d = 1, \dots, D$, $t = 1, \dots, m_d$, and $\sigma_u^2 = 1$.

The first simulation experiment is dedicated to investigated the gain of efficiency achieved by the EBLUP based on model (2.2) 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 $(k = 1, \dots, K)$

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 (2.4) 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 (2.2).

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 2.2.6.1 (left). Biases $BIAS^{(0)}$ and $BIAS^{(1)}$ are presented in the Table 2.2.6.1 (right). In the Figure 2.2.6.1 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 2.2.6.2 the $BIAS_{dm_d}$-values are plotted for $D = 100$, $m_d = 5$ with the same configuration as in the Figure 2.2.6.1.

When the true model is model (2.4) 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 (2.2) the increment of MSE is almost negligible. This can be appreciated in the two first rows of the Table 2.2.6.1 (left) and on the Figure 2.2.6.1. If we look at the bias, no increment is observed for incorrectly using model (2.2).

		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 2.2.6.1. MSE's (left) and $BIAS$'s (right) of EBLUP0 and EBLUP1 for $D = 100$

When the true model is model (2.2) 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 (2.4). 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.

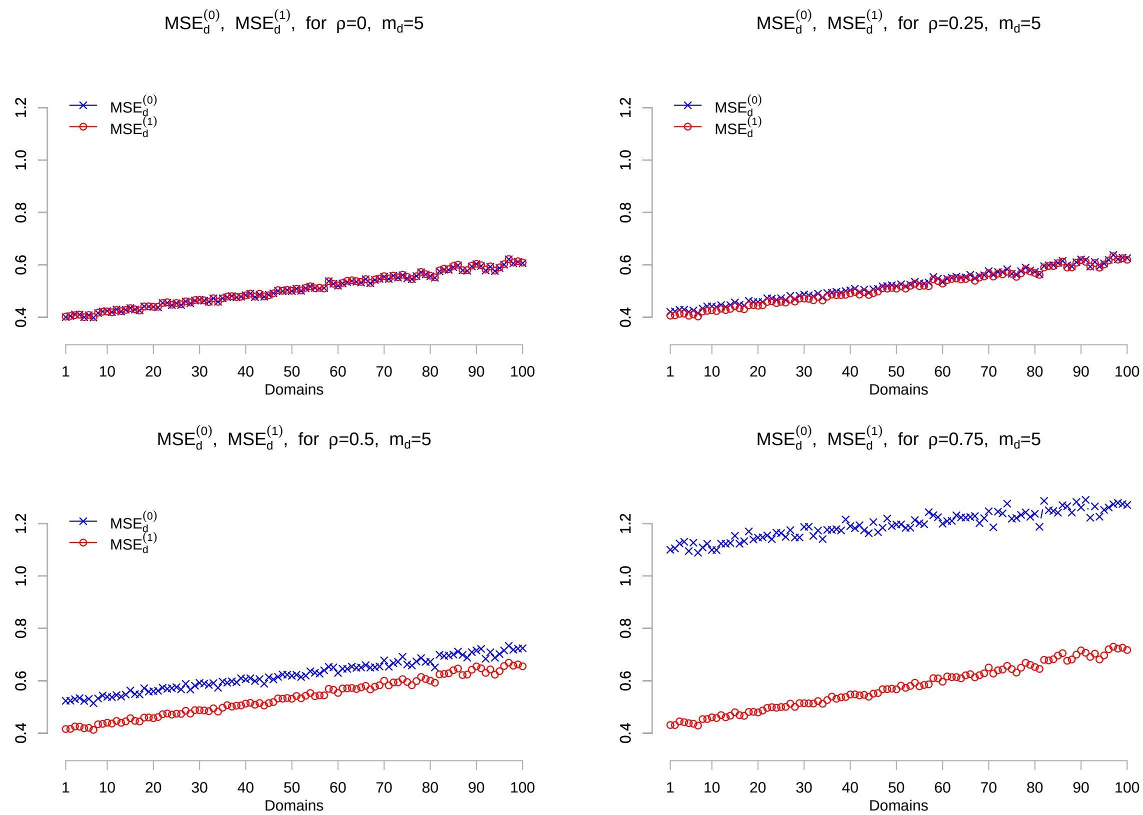

Figure 2.2.6.1. MSE_{dm_d}'s of EBLUP0 and EBLUP1 for $D = 100$, $m_d = 5$.

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 2.2.6.2 (left). Biases $B^{(0)}$ and $B^{(1)}$ are presented in the Table 2.2.6.2 (right). For $D = 100$ and $m_d = 5$, in the Figure 2.2.6.3 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 (2.4) or under the incorrect model (2.2). On the other

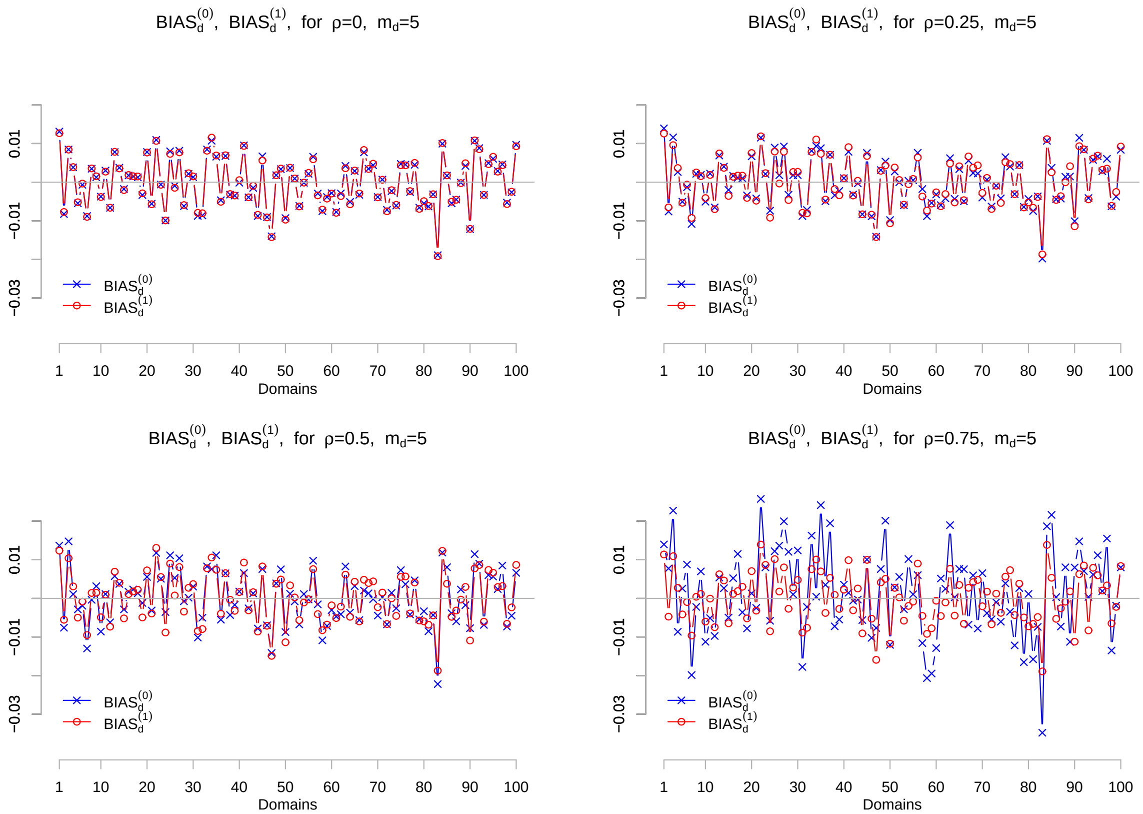

Figure 2.2.6.2. $BIAS_{dm_d}$'s of EBLUP0 and EBLUP1 for $D = 100$, $m_d = 5$.

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 (2.4). Again we conclude that if true model is model (2.2), then there is a loss of efficiency by using model (2.4). 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 2.2.6.2. E's (left) and B's (right) of EBLUP0 and EBLUP1 for $D = 100$

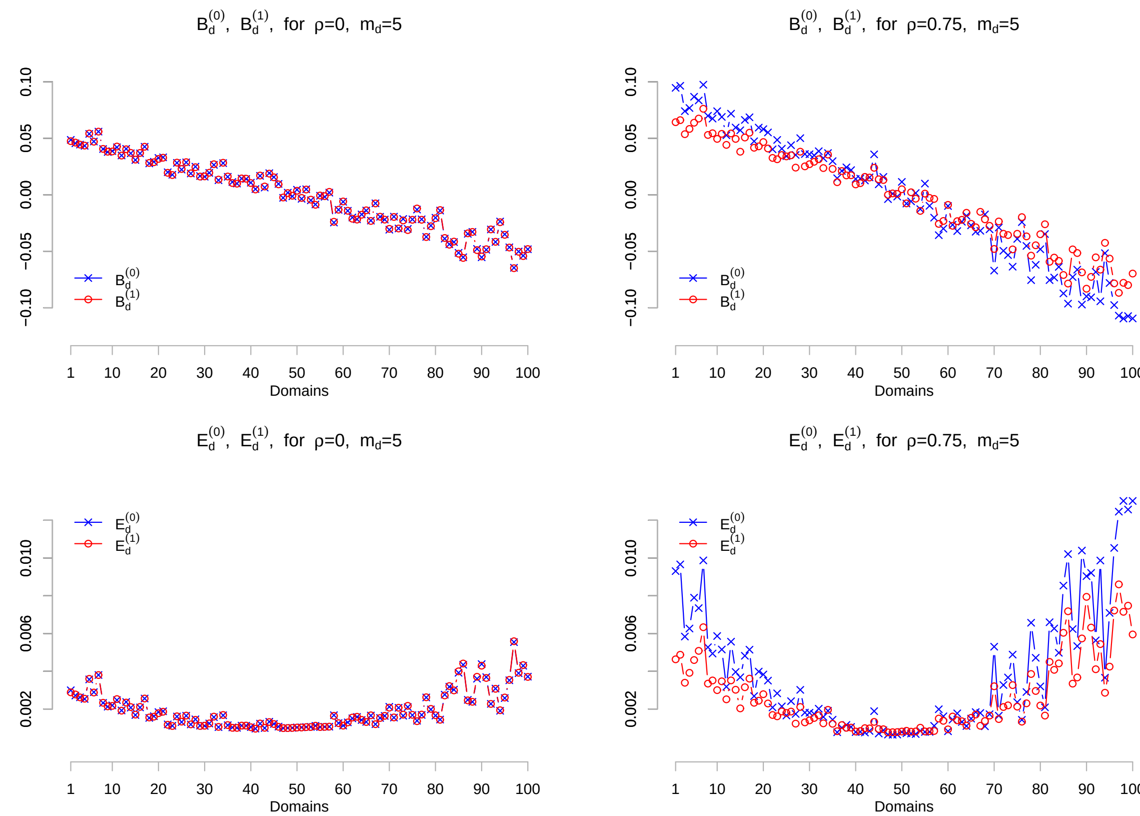

Figure 2.2.6.3. B_{dm_d}'s (top) and E_{dm_d}'s (bottom) of EBLUP0 and EBLUP1 for $D = 100$, $m_d = 5$.

2.3 An application to the Spanish Living Conditions Survey

2.3.1 Estimation of poverty indicators

Let us consider a finite population P_t partitioned into D domains P_{dt} at time period t, and denote their sizes by N_t and N_{dt}, $d = 1, \dots, D$. Let z_{dtj} be an income variable measured in all the units of the population and let z_t be the poverty line, so that units with $z_{dtj} < z_t$ are considered as poor at time period t. The main goal of this section is to estimate the poverty incidence (proportion of individuals under poverty) and the poverty gap in Spanish domains. These two measures belongs to the FGT family proposed by Foster et al. (1984), given by

$$Y_{\alpha;dt} = \frac{1}{N_{dt}} \sum_{j=1}^{N_{dt}} y_{\alpha;dtj}, \quad \text{where } \; y_{\alpha;dtj} = \left(\frac{z_t - z_{dtj}}{z_t} \right)^{\alpha} I(z_{dtj} < z_t), \tag{2.7}$$

$I(z_{dtj} < z_t) = 1$ if $z_{dtj} < z_t$ and $I(z_{dtj} < z_t) = 0$ otherwise. The proportion of units under poverty in the domain d and period t is thus $Y_{0;dt}$ and the poverty gap is $Y_{1;dt}$.

We use data from the Spanish Living Conditions Survey (SLCS) corresponding to years 2004-2006 with sample sizes 44648, 37491, 34694 respectively. The SLCS is the Spanish version of the European Statistics on Income and Living Conditions (EU-SILC), which is one of the statistical operations that have been harmonized for EU countries. The SLCS started in 2004 with an annual periodicity. Its main goal is to provide a reference source on comparative statistics on the distribution of income and social exclusion in the European environment.

The SLCS is an annual survey with a rotating panel design with a sample formed by four independent subsamples, each of which is a four-year panel. Each year the sample is renewed in one of the panels. In order to select each subsample, a two-stage design is implemented independently in each Autonomous Community with first stage unit stratification. The first stage is formed by census sections grouped into strata in agreement with the size of the municipality to which they belong. The second stage is formed by main family dwellings. Within these no sub-sampling is carried out, investigating all dwellings that are their usual residence. The sample includes 16000 dwellings distributed in 2000 census sections.

We consider $D = 104$ domains obtained by crossing 52 provinces with 2 sexes. The quartiles of the distribution of the domain sample sizes are $q_0 = 17$, $q_1 = 170$, $q_2 = 293$, $q_3 = 640$, $q_4 = 2113$ in 2004, 13, 149, 251, 530, 1494 in 2005 and 18, 129, 233, 481, 1494 in 2006, so they are too small to employ direct estimators to estimate the parameters of interest in all the domains.

The SLCS does not produce official estimates at the domain level (provinces $\times$ sex), but the analogous direct estimator of the total $Y_{dt} = \sum_{j=1}^{N_{dt}} y_{dtj}$ is

$$\hat{Y}_{dt}^{dir} = \sum_{j\in S_{dt}} w_{dtj}\, y_{dtj}.$$

where S_{dt} is the domain sample at time period t and the w_{dtj}'s are the official calibrated sampling weights which take into account for non response. In the particular case $y_{dtj} = 1$, for all $j \in P_{dt}$, we get the estimated domain size

$$\hat{N}_{dt}^{dir} = \sum_{j\in S_{dt}} w_{dtj}.$$

Using this quantity, a direct estimator of the domain mean $\bar{Y}_{dt}$ is $\bar{y}_{dt} = \hat{Y}_{dt}^{dir}/\hat{N}_{dt}^{dir}$. The direct estimates of the domain means are used as responses in the area-level time model. The design-based variances of these estimators can be approximated by

$$\hat{V}_{\pi}(\hat{Y}_{dt}^{dir}) = \sum_{j\in S_{dt}} w_{dtj}(w_{dtj}-1)\,(y_{dtj}-\bar{y}_{dt})^2 \quad \text{and} \quad \hat{V}_{\pi}(\bar{y}_{dt}) = \hat{V}\left(\hat{Y}_{dt}^{dir}\right)/\hat{N}_{dt}^2. \tag{2.8}$$

The last formulas are obtained from Särndal et al. (1992), pp. 43, 185 and 391, with the simplifications $w_{dtj} = 1/\pi_{dtj}$, $\pi_{dtj,dtj} = \pi_{dtj}$ and $\pi_{dti,dtj} = \pi_{dti}\pi_{dtj}$, $i \neq j$ in the second order inclusion probabilities.

As we are interested in the cases $y_{dtj} = y_{\alpha;dtj}$, $\alpha = 0, 1$, we select the direct estimates of the poverty incidence and poverty gap at domain d and time period t (i.e. $\bar{y}_{0;dt}$ and $\bar{y}_{1;dt}$ respectively) as target variables for the time dependent area-level models 0 and 1. The considered auxiliary variables are the known domain means of the category indicators of the following variables:

- INTERCEPT: $\bar{X}_{dt}^{(0)} = 1$.
- AGE: Age groups are *age1*-*age5* for the intervals ≤ 15, $16-24$, $25-49$, $50-64$ and ≥ 65. The corresponding auxiliary variables (domain proportions) are denoted by $\bar{X}_{dt,1}^{(1)}, \ldots, \bar{X}_{dt,5}^{(1)}$.
- EDUCATION: Highest level of education completed, with 4 categories denoted by *edu0* for Less than primary education level, *edu1* for Primary education level, *edu2* for Secondary education level and *edu3* for University level. Auxiliary variables are $\bar{X}_{dt,0}^{(2)}, \bar{X}_{dt,1}^{(2)}, \bar{X}_{dt,2}^{(2)}, \bar{X}_{dt,3}^{(2)}$.
- CITIZENSHIP: with 2 categories denoted by *cit1* for Spanish and *cit2* for Not Spanish. Auxiliary variables are $\bar{X}_{dt,1}^{(3)}, \bar{X}_{dt,2}^{(3)}$.
- LABOR: Labor situation with 4 categories taking the values *lab0* for Below 16 years, *lab1* for Employed, *lab2* for Unemployed and *lab3* for Inactive. Auxiliary variables are $\bar{X}_{dt,0}^{(4)}, \bar{X}_{dt,1}^{(4)}, \bar{X}_{dt,2}^{(4)}, \bar{X}_{dt,3}^{(4)}$.

Following the standards of the Spanish Statistical Office, the Poverty Threshold is fixed as the 60% of the median of the normalized incomes in Spanish households. The aim of normalizing the household income is to adjust for the varying size and composition of households. The definition of the total number of normalized household members is the modified OECD scale used by EUROSTAT. This scale gives a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over and 0.3 to each child aged under 14 in the household. The *normalized size* of a household is the sum of the weights assigned to each person. So the total number of normalized household members is

$$H_{dti} = 1 + 0.5(H_{dti\geq 14} - 1) + 0.3H_{dti<14}$$

where $H_{dti\geq 14}$ is the number of people aged 14 and over and $H_{dti<14}$ is the number of children aged under 14. The normalized net annual income of a household is obtained by dividing its net annual income by its normalized size. The Spanish poverty thresholds (in euros) in 2004-06 are $z_{2004} = 6098.57$, $z_{2005} = 6160.00$ and $z_{2006} = 6556.60$ respectively. These are the z_t-values used in the calculation of the direct estimates of the poverty incidence and gap.

We first consider the linear model

$$\overline{y}_{dt} = \bar{\mathbf{X}}_{dt}\boldsymbol{\beta} + u_{dt} + e_{dt}, \quad d = 1, \ldots, D$$

where $\bar{\mathbf{X}}_d$ is the $1 \times p$ vector containing the population (aggregated) mean values of all the categories (except the last one) of the explanatory variables. The first position of $\bar{\mathbf{X}}_d$ contains a "1", so that $p = 1 + 4 + 3 + 1 + 3 = 12$. Random effects errors are assumed to follow the distributional assumptions of model (2.2) either restricted to $\rho = 0$ (model 0) or without this restriction (model 1). As some of the explanatory variables where not significative, the starting models where simplified to include only the auxiliary variables appearing in the Table 2.3.1.1. As the estimates of ρ are 0.8585 and 0.7124 for $\alpha = 0$ and $\alpha = 1$, we recommend model 1 and present its regression parameters in the Table 2.3.1.1.

$\alpha = 0$	*constant*	*age3*	*age4*	*edu1*	*edu2*	*cit1*	*lab2*
	0.5439	-0.5823	-2.6071	0.0895	-0.0556	0.3308	0.2040
$\alpha = 1$	*constant*	*edu0*	*edu1*	*edu2*	*cit1*	*lab1*	
	-0.2613	0.7767	0.2184	0.1286	0.1165	-0.0538	

Table 2.3.1.1. Auxiliary variables for selected type-1 models

By observing the signs of the regression parameters in model 1 for $\alpha = 0$, we interpret that poverty proportion tends to be smaller in those domains with larger proportion of population in the subset defined by age in the interval 25-64 (age interval with greater incomes), education in the category of secondary studies completed, and non Spanish citizenship (may be because immigrants tends to go to regions with greater richness where it is easier to find job), and with lower proportion of unemployed people. By doing the same exercise with the signs of the regression parameter in model 1 for $\alpha = 1$, we interpret that poverty gap tends to be smaller in those domains with larger proportion of population characterized by university education completed, non Spanish citizenship and employed.

Residuals $\hat{e}_{dt} = \bar{y}_{dt} - \bar{\mathbf{X}}_{dt}\hat{\boldsymbol{\beta}} - \hat{u}_{dt}$ of model 1 are plotted against the observed values $\bar{y}_{dt}$ in the Figure 2.3.1.1 for $\alpha = 0$ (top-right) and $\alpha = 1$ (top-left). The dispersion graph shows that EBLUP1 estimates are over and below direct estimates, so that the design unbiased property of the direct estimator is not completely lost by using the model 1. On the right part of the figure we observe that residuals tend to be positive, which means that the model is smoothing the value of the direct estimator larger values. We find that this is an interesting property because it protects us from the presence of outliers in the collection of direct domain estimates.

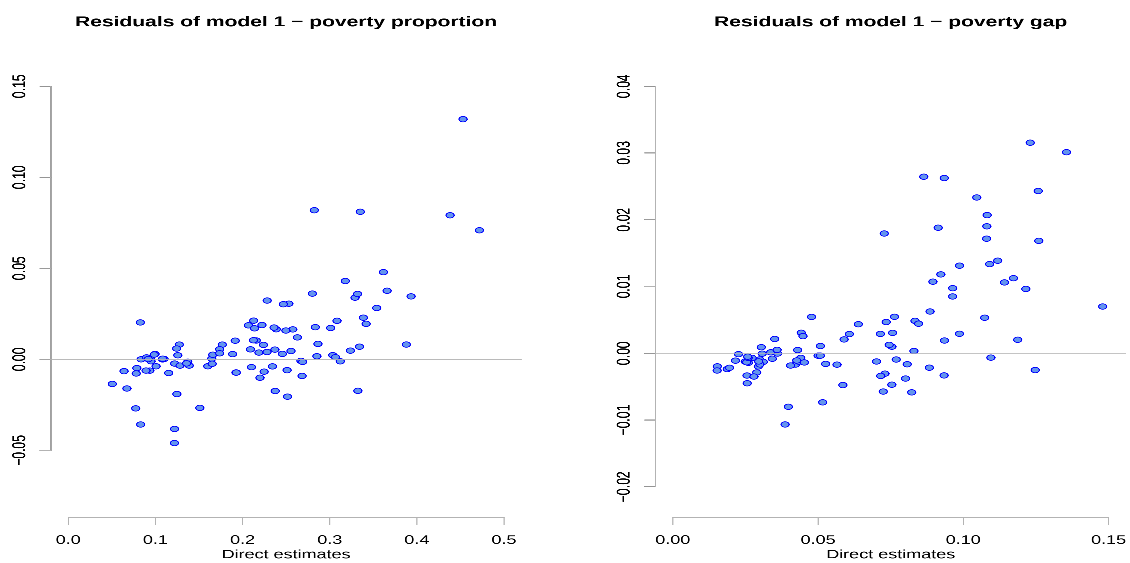

Figure 2.3.1.1. Residuals versus direct estimates.

The three considered estimators of the poverty proportion and gap (direct, EBLUP0 and EBLUP1) are plotted in the Figure 2.3.1.2 for $\alpha = 0$ (top-left) and $\alpha = 1$ (top-right). Their root mean squared error estimates are plotted in the Figure 2.3.1.2 for $\alpha = 0$ (bottom-left) and

$\alpha = 1$ (bottom-right). We observe that the EBLUP1 is the one presenting the best results and it is thus the one we recommend. Finally full numerical information is presented in the Table 2.3.3.1 for the poverty proportion and in the Table 2.3.3.2 for the poverty gap. In these tables direct, EBLUP0 and EBLUP1 estimates are labeled by dir, eb0 and eb1 respectively.

The Spanish provinces are 1 Álava, 2 Albacete, 3 Alicante, 4 Almería, 5 Ávila, 6 Badajoz, 7 Baleares, 8 Barcelona, 9 Burgos, 10 Cáceres, 11 Cádiz, 12 Castellón, 13 Ciudad Real, 14 Córdoba, 15 Coruña La, 16 Cuenca, 17 Gerona, 18 Granada, 19 Guadalajara, 20 Guipúzcoa, 21 Huelva, 22 Huesca, 23 Jaén, 24 León, 25 Lérida, 26 La Rioja, 27 Lugo, 28 Madrid, 29 Málaga, 30 Murcia, 31 Navarra, 32 Orense, 33 Asturias (Oviedo), 34 Palencia, 35 Palmas Las, 36 Pontevedra, 37 Salamanca, 38 Santa Cruz de Tenerife, 39 Cantabria (Santander), 40 Segovia, 41 Sevilla, 42 Soria, 43 Tarragona, 44 Teruel, 45 Toledo, 46 Valencia, 47 Valladolid, 48 Vizcaya, 49 Zamora, 50 Zaragoza, 51 Ceuta, 52 Melilla.

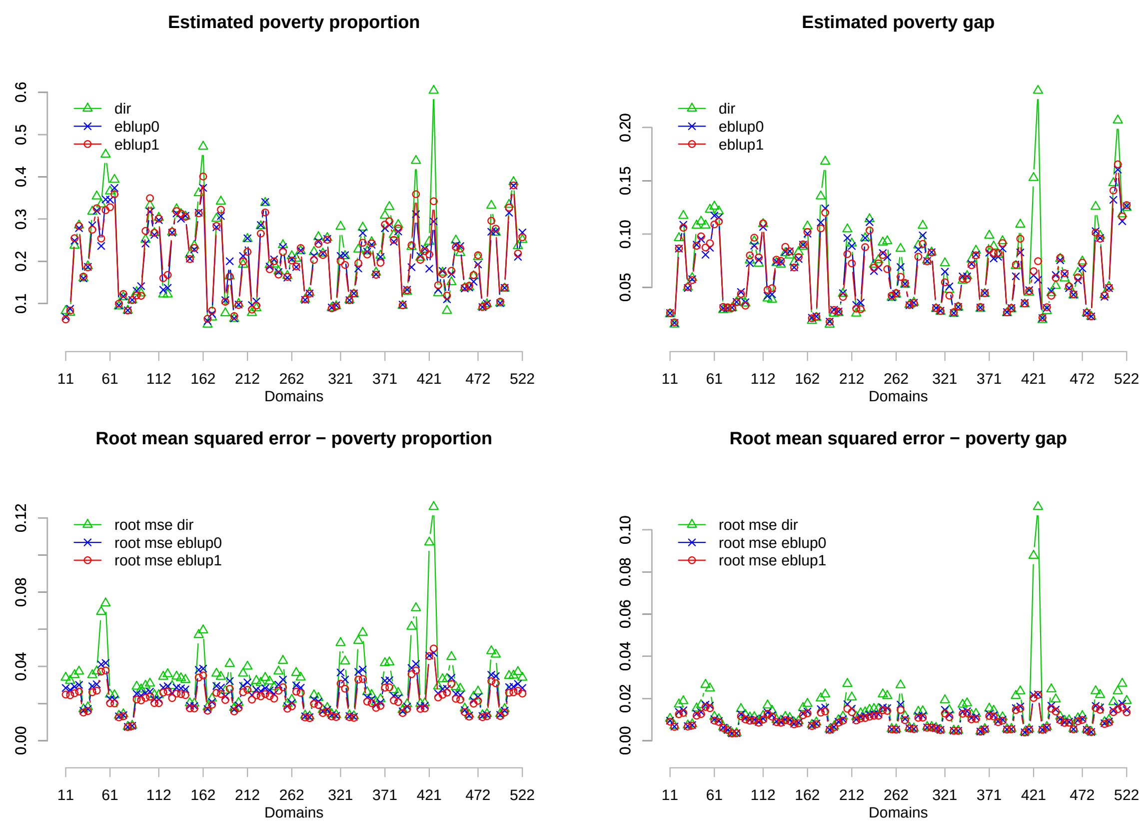

Figure 2.3.1.2. Estimates of poverty proportions and gaps (top) and squared roots of their estimated MSEs (bottom).

In the Table 2.3.1.2 the Spanish provinces are classified in 4 categories depending on the values of the EBLUP1 estimates in % of the poverty proportions and the gaps, i.e. $p_d = 100 \cdot \hat{Y}^{eblup1}_{0;d,2006}$ and $g_d = 100 \cdot \hat{Y}^{eblup1}_{1;d,2006}$. The same results are plotted in Figure 2.3.1.3. We observe that the Spanish regions where the proportion of the population under the poverty line is smallest are those situated in the north and east, like Cataluña, Aragón, Navarra, País Vasco, Cantabria and Baleares. On the other hand the Spanish regions with higher poverty proportion are those situated in the center-south, like Andalucía, Extremadura, Murcia, Castilla La Mancha, Canarias, Canarias, Ceuta and Melilla. In an intermediate position we can find regions that are in the center-north of Spain, like Galicia, La Rioja, Castilla León, Asturias, Comunidad Valenciana and Madrid. If we investigate how far the annual net incomes of population under the poverty line z_{2006} are from z_{2006}, we observe that in the Spanish regions situated in the center-north there exist a distance that is generally lower than the 6% of z_{2006}. However, the cited distance is in general greater than 6% of z_{2006} in the center-south.

It is somehow surprising how large is the amount of Spanish provinces having a proportion greater than 30% of population with annual net incomes below z_{2006}. So it would be desirable that the Spanish Government implements policies to reduce poverty proportion in the center-south of Spain. A criticisms to the use of employed FGT poverty indicators is that they are defined by using only the income variable and do not consider the cost of living. In the case of Spain the south is poorer than the north and this is visualized in the obtained results. On the other hand to live in the south is in general cheaper than to live in the north. As the poverty line is officially calculated for the whole Spain, it is smaller than it should be in the north and greater than it should be in the south. Nevertheless, these comments may moderate but not annul the validity of the given conclusions. The obtained results are valuable tools for taking decisions as they show the basic poverty situation per provinces and sex in Spain.

men	$p_d < 10$	$1, 17, 20, 8, 22, 31, 48, 39, 7$
	$10 < p_d < 20$	$19, 50, 33, 28, 44, 9, 46, 43, 12, 26, 3, 47, 36, 25, 24, 27, 34, 21, 32$
	$20 < p_d < 30$	$29, 41, 15, 30, 35, 42, 52, 45, 40, 38, 5, 2, 23, 11, 13, 10, 4, 18, 37, 49$
	$p_d > 30$	$14, 16, 51, 6$
women	$p_d < 10$	$17, 1, 31, 22, 48, 20$
	$10 < p_d < 20$	$8, 9, 7, 28, 33, 39, 50, 46, 19, 12, 43, 44, 3, 32, 36, 24$
	$20 < p_d < 30$	$26, 47, 25, 21, 41, 45, 27, 15, 29, 35, 34, 30, 52, 49, 38, 2, 37$
	$p_d > 30$	$11, 14, 23, 13, 5, 18, 4, 42, 10, 6, 40, 51, 16$
men	$g_d < 3$	$19, 17, 43, 1, 48, 39, 33, 20, 22$
	$3 < g_d < 6$	$31, 8, 36, 7, 28, 41, 44, 26, 50, 9, 12, 3, 46, 32, 34, 47, 27$
	$6 < g_d < 10$	$42, 24, 15, 35, 40, 30, 13, 45, 11, 29, 10, 25, 21, 38, 37, 2, 5, 23, 14, 4, 16$
	$g_d > 10$	$49, 18, 6, 52, 51$
women	$g_d < 3$	$1, 17, 48, 31, 19, 22$
	$3 < g_d < 6$	$39, 7, 43, 33, 9, 28, 8, 20, 32, 46, 26, 36, 41, 12, 50, 27, 3, 34, 44$
	$6 < g_d < 10$	$45, 25, 21, 24, 47, 42, 13, 15, 35, 37, 30, 14, 29, 5, 38, 40, 49, 10, 4$
	$g_d > 10$	$16, 23, 2, 11, 6, 18, 52, 51$

Table 2.3.1.2. Spanish provinces classified by poverty proportion (up) and gap (bottom) in %.

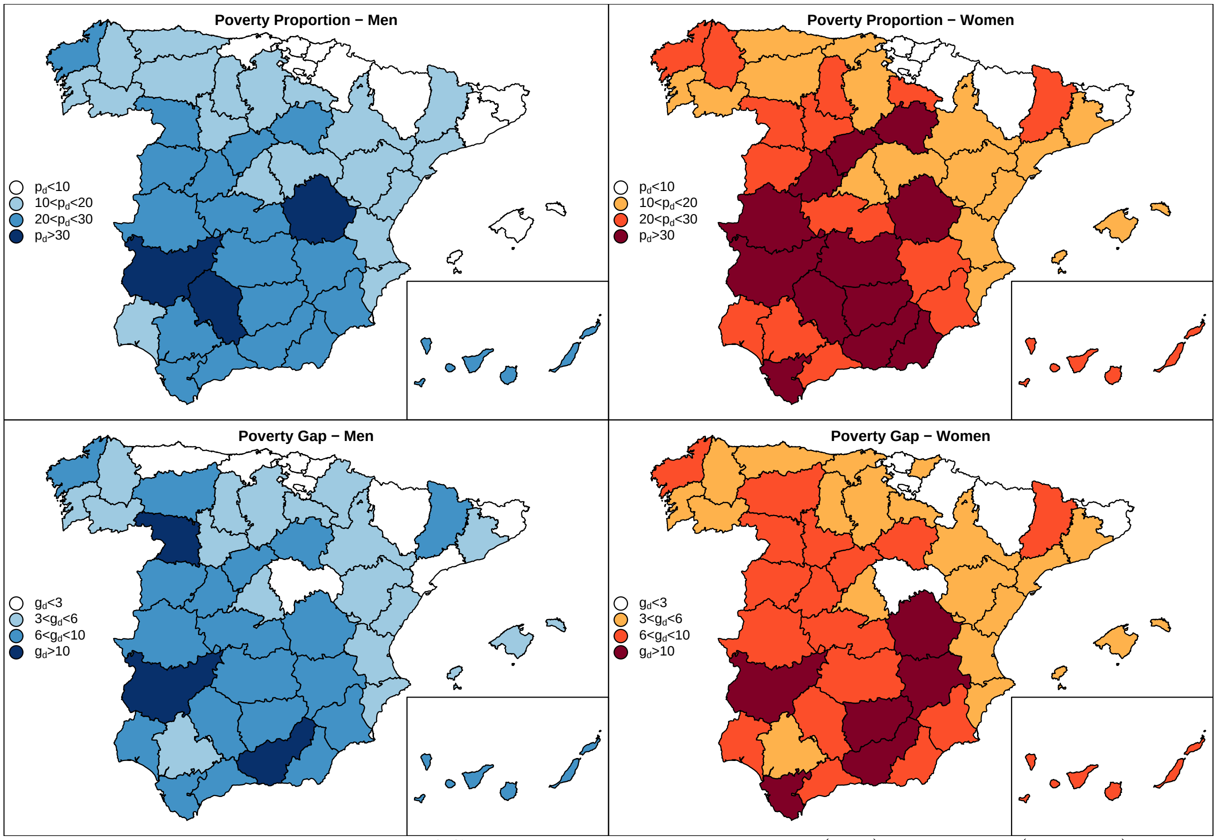

Figure 2.3.1.3. Estimates of Spanish poverty proportions (top) and gaps (bottom) for men (left) and women (right).

2.3.2 Conclusions

As poverty indicators are nonlinear, unit-level model-based estimation approaches cannot always be used. Instead, area-level models provide an easy-to-apply solution. To reinforce the predictability of the area-level models, we propose the use of temporal models that borrow strength from time. Two models are introduced and simulation studies have been carried out to investigate when it is worthwhile to introduce a time correlation parameter. Finally the methodology has been applied to Spanish EU-SILC data and poverty mapping pre provinces and sex has been given.

Acknowledgements

The authors are grateful to the European Commission and to the Spanish Government for their economical support under the grants EU-FP7-SSH-2007-1 and MTM2006-05693 respectively. The authors also thanks the Instituto Nacional de Estadística for providing the Spanish EU-SILC data.

2.3.3 Tables

	men / poverty proportions / women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	ebl0	eb1
1	0.083	0.068	0.062	0.079	0.088	0.083	0.034	0.028	0.025	0.032	0.027	0.024
2	0.237	0.249	0.255	0.285	0.279	0.284	0.035	0.029	0.026	0.037	0.030	0.026
3	0.160	0.160	0.164	0.189	0.187	0.186	0.017	0.016	0.015	0.018	0.017	0.016
4	0.318	0.286	0.275	0.354	0.320	0.326	0.035	0.029	0.026	0.037	0.030	0.027
5	0.335	0.236	0.254	0.453	0.346	0.321	0.069	0.041	0.037	0.074	0.042	0.038
6	0.366	0.346	0.328	0.393	0.374	0.359	0.025	0.022	0.020	0.025	0.022	0.020
7	0.094	0.096	0.100	0.115	0.117	0.123	0.014	0.013	0.013	0.014	0.014	0.013
8	0.083	0.084	0.083	0.108	0.109	0.108	0.008	0.007	0.007	0.008	0.008	0.008
9	0.127	0.130	0.119	0.124	0.141	0.118	0.029	0.025	0.022	0.028	0.024	0.022
10	0.252	0.242	0.272	0.332	0.317	0.350	0.030	0.026	0.023	0.031	0.026	0.024
11	0.267	0.263	0.268	0.303	0.298	0.301	0.025	0.022	0.020	0.025	0.022	0.020
12	0.122	0.133	0.160	0.122	0.137	0.168	0.034	0.029	0.026	0.036	0.029	0.026
13	0.269	0.268	0.270	0.324	0.313	0.319	0.030	0.026	0.023	0.035	0.029	0.025
14	0.312	0.301	0.313	0.307	0.308	0.305	0.034	0.028	0.025	0.033	0.028	0.025
15	0.216	0.206	0.205	0.237	0.228	0.232	0.020	0.019	0.017	0.020	0.019	0.017
16	0.362	0.315	0.314	0.472	0.374	0.401	0.057	0.038	0.034	0.059	0.039	0.035
17	0.050	0.059	0.064	0.067	0.076	0.083	0.018	0.017	0.016	0.023	0.021	0.019
18	0.301	0.281	0.284	0.342	0.307	0.322	0.036	0.029	0.026	0.034	0.029	0.025
19	0.077	0.108	0.104	0.165	0.200	0.164	0.027	0.024	0.022	0.041	0.032	0.028
20	0.064	0.065	0.070	0.100	0.098	0.097	0.018	0.017	0.016	0.020	0.019	0.018
21	0.192	0.213	0.200	0.253	0.255	0.223	0.036	0.030	0.026	0.040	0.031	0.027
22	0.078	0.096	0.086	0.089	0.105	0.095	0.028	0.024	0.022	0.032	0.027	0.024
23	0.283	0.286	0.266	0.339	0.340	0.316	0.031	0.027	0.024	0.034	0.028	0.025
24	0.192	0.188	0.181	0.193	0.205	0.200	0.032	0.027	0.024	0.029	0.025	0.023
25	0.177	0.175	0.169	0.239	0.234	0.222	0.037	0.030	0.027	0.043	0.033	0.029
26	0.166	0.161	0.163	0.212	0.205	0.202	0.020	0.019	0.017	0.022	0.020	0.018
27	0.207	0.186	0.188	0.225	0.225	0.231	0.037	0.030	0.026	0.034	0.028	0.026
28	0.110	0.109	0.110	0.126	0.124	0.123	0.014	0.013	0.013	0.013	0.013	0.012
29	0.222	0.213	0.203	0.258	0.247	0.241	0.025	0.022	0.020	0.023	0.021	0.019
30	0.219	0.218	0.215	0.256	0.253	0.251	0.017	0.016	0.015	0.018	0.017	0.016
31	0.090	0.091	0.089	0.094	0.096	0.094	0.014	0.014	0.013	0.014	0.013	0.013
32	0.282	0.213	0.200	0.213	0.215	0.191	0.053	0.037	0.031	0.043	0.033	0.028
33	0.108	0.108	0.108	0.122	0.124	0.124	0.014	0.013	0.013	0.013	0.013	0.012
34	0.228	0.183	0.196	0.280	0.267	0.244	0.054	0.037	0.033	0.058	0.038	0.033
35	0.224	0.223	0.216	0.246	0.239	0.243	0.026	0.023	0.021	0.025	0.022	0.020
36	0.174	0.169	0.168	0.214	0.210	0.197	0.021	0.019	0.018	0.022	0.021	0.019
37	0.308	0.278	0.287	0.329	0.286	0.295	0.042	0.032	0.029	0.042	0.033	0.029
38	0.263	0.246	0.251	0.286	0.271	0.278	0.027	0.024	0.022	0.026	0.023	0.021
39	0.095	0.098	0.096	0.128	0.133	0.132	0.017	0.016	0.015	0.020	0.018	0.017
40	0.234	0.186	0.238	0.438	0.312	0.359	0.061	0.039	0.036	0.071	0.041	0.038
41	0.209	0.213	0.204	0.228	0.230	0.224	0.020	0.019	0.017	0.020	0.019	0.017
42	0.247	0.183	0.216	0.604	0.294	0.342	0.107	0.046	0.046	0.126	0.047	0.050
43	0.125	0.133	0.144	0.174	0.177	0.170	0.029	0.025	0.023	0.033	0.028	0.025
44	0.083	0.110	0.119	0.151	0.170	0.178	0.033	0.028	0.026	0.045	0.034	0.031
45	0.250	0.238	0.234	0.220	0.236	0.230	0.029	0.025	0.023	0.028	0.025	0.022
46	0.137	0.136	0.138	0.139	0.138	0.142	0.017	0.016	0.015	0.014	0.014	0.013
47	0.165	0.152	0.168	0.210	0.193	0.214	0.024	0.022	0.020	0.027	0.024	0.021
48	0.092	0.092	0.092	0.099	0.100	0.096	0.014	0.013	0.013	0.014	0.014	0.013
49	0.332	0.270	0.296	0.268	0.271	0.277	0.048	0.035	0.032	0.046	0.035	0.031
50	0.101	0.101	0.105	0.136	0.137	0.138	0.014	0.014	0.013	0.017	0.016	0.015

Table 2.3.3.1. Estimated domain poverty proportions and their estimated sqrt MSE's by sex.

	men / poverty gaps / women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	ebl0	eb1
1	0.025	0.026	0.026	0.015	0.017	0.017	0.010	0.009	0.009	0.007	0.007	0.007
2	0.096	0.086	0.087	0.117	0.109	0.106	0.017	0.014	0.013	0.019	0.014	0.013
3	0.050	0.049	0.050	0.059	0.058	0.057	0.007	0.007	0.007	0.008	0.007	0.007
4	0.108	0.090	0.089	0.112	0.095	0.098	0.015	0.013	0.012	0.017	0.014	0.013
5	0.108	0.081	0.087	0.123	0.087	0.091	0.027	0.017	0.016	0.025	0.017	0.015
6	0.126	0.117	0.109	0.121	0.116	0.112	0.011	0.010	0.010	0.010	0.009	0.009
7	0.029	0.030	0.032	0.029	0.031	0.031	0.006	0.006	0.006	0.005	0.005	0.005
8	0.031	0.031	0.031	0.036	0.036	0.036	0.003	0.003	0.003	0.004	0.004	0.004
9	0.042	0.046	0.044	0.035	0.036	0.033	0.015	0.013	0.012	0.012	0.011	0.010
10	0.075	0.073	0.080	0.093	0.089	0.097	0.011	0.010	0.009	0.011	0.010	0.010
11	0.072	0.076	0.078	0.109	0.107	0.110	0.010	0.009	0.009	0.012	0.011	0.010
12	0.040	0.042	0.048	0.039	0.044	0.049	0.017	0.013	0.013	0.014	0.012	0.011
13	0.073	0.074	0.076	0.072	0.074	0.075	0.010	0.009	0.009	0.010	0.009	0.008
14	0.082	0.083	0.088	0.080	0.084	0.084	0.011	0.010	0.010	0.011	0.010	0.009
15	0.073	0.069	0.069	0.083	0.077	0.078	0.009	0.008	0.008	0.009	0.009	0.008
16	0.088	0.090	0.090	0.107	0.100	0.102	0.016	0.013	0.012	0.018	0.014	0.013
17	0.019	0.022	0.021	0.022	0.023	0.023	0.008	0.007	0.007	0.009	0.008	0.008
18	0.135	0.111	0.105	0.168	0.124	0.120	0.020	0.015	0.013	0.022	0.016	0.014
19	0.015	0.018	0.018	0.026	0.029	0.029	0.005	0.005	0.005	0.007	0.007	0.006
20	0.026	0.027	0.027	0.044	0.045	0.041	0.010	0.009	0.009	0.011	0.010	0.009
21	0.105	0.096	0.081	0.091	0.089	0.073	0.027	0.017	0.015	0.021	0.015	0.014
22	0.026	0.034	0.030	0.030	0.036	0.030	0.011	0.010	0.010	0.013	0.011	0.011
23	0.096	0.096	0.088	0.114	0.111	0.104	0.013	0.012	0.011	0.015	0.012	0.011
24	0.071	0.065	0.069	0.076	0.069	0.073	0.015	0.013	0.012	0.015	0.013	0.012
25	0.092	0.082	0.080	0.093	0.078	0.067	0.022	0.016	0.014	0.021	0.015	0.014
26	0.041	0.041	0.042	0.043	0.044	0.045	0.006	0.005	0.005	0.005	0.005	0.005
27	0.086	0.069	0.060	0.053	0.053	0.054	0.026	0.017	0.015	0.012	0.010	0.010
28	0.034	0.033	0.034	0.036	0.036	0.035	0.006	0.006	0.006	0.006	0.006	0.005
29	0.090	0.086	0.079	0.108	0.099	0.091	0.014	0.012	0.011	0.014	0.012	0.011
30	0.075	0.075	0.074	0.083	0.083	0.083	0.007	0.006	0.006	0.007	0.006	0.006
31	0.030	0.031	0.031	0.027	0.028	0.028	0.006	0.006	0.006	0.005	0.005	0.005
32	0.073	0.065	0.055	0.048	0.051	0.042	0.019	0.015	0.013	0.014	0.012	0.011
33	0.025	0.026	0.027	0.031	0.032	0.032	0.005	0.005	0.005	0.005	0.005	0.005
34	0.056	0.060	0.058	0.061	0.060	0.058	0.017	0.014	0.013	0.018	0.014	0.013
35	0.076	0.072	0.071	0.085	0.081	0.080	0.012	0.011	0.010	0.013	0.011	0.010
36	0.030	0.031	0.032	0.044	0.045	0.045	0.004	0.004	0.004	0.006	0.005	0.005
37	0.099	0.083	0.086	0.089	0.077	0.082	0.015	0.013	0.012	0.014	0.012	0.011
38	0.081	0.079	0.082	0.093	0.087	0.092	0.010	0.009	0.009	0.011	0.010	0.010
39	0.026	0.027	0.027	0.030	0.031	0.031	0.006	0.005	0.005	0.006	0.005	0.005
40	0.070	0.060	0.071	0.109	0.083	0.096	0.021	0.015	0.015	0.023	0.016	0.015
41	0.034	0.035	0.035	0.045	0.047	0.047	0.004	0.004	0.004	0.006	0.005	0.005
42	0.153	0.061	0.065	0.235	0.057	0.075	0.088	0.022	0.020	0.111	0.022	0.022
43	0.019	0.021	0.022	0.028	0.031	0.031	0.005	0.005	0.005	0.007	0.006	0.006
44	0.045	0.046	0.042	0.052	0.062	0.059	0.024	0.017	0.015	0.020	0.015	0.014
45	0.077	0.075	0.078	0.059	0.063	0.063	0.011	0.010	0.010	0.009	0.009	0.008
46	0.051	0.049	0.051	0.043	0.043	0.044	0.010	0.009	0.008	0.006	0.005	0.005
47	0.064	0.057	0.059	0.074	0.068	0.073	0.011	0.010	0.009	0.012	0.010	0.010
48	0.026	0.026	0.026	0.023	0.023	0.023	0.005	0.005	0.005	0.004	0.004	0.004
49	0.126	0.103	0.101	0.099	0.096	0.096	0.024	0.016	0.015	0.022	0.016	0.015
50	0.043	0.041	0.042	0.051	0.049	0.050	0.009	0.008	0.008	0.010	0.009	0.009

Table 2.3.3.2. Estimated domain poverty gaps and their estimated squared root MSE's by sex.

Chapter 3

EB prediction of non-linear domain parameters with a unit-level model

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. The method is applied to the estimation of FGT poverty measures in Spanish provinces by gender. Thus, Section 3.1 describes the empirical best predictor of a non-linear population parameter. Section 3.2 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 3.3 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 3.4 describes the parametric bootstrap for mean squared error estimation. Section 3.5 particularizes de proposed method to the estimation of domain FGT poverty measures. Section 3.6 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. Section 3.7 describes the results of model-based and design-based simulation experiments 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. Finally, Section 3.8 summarizes the results of an application with Spanish data from the SILC, with full figures included in Appendix.

3.1 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

$$MSE(\hat{\delta}) = E_{\mathbf{y}}[(\hat{\delta} - \delta)^2], \tag{3.1}$$

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 (3.1). 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

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

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

$$\hat{\delta}^B = \delta^0 = E_{\mathbf{y}_r}(\delta|\mathbf{y}_s). \tag{3.2}$$

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 $\boldsymbol{\theta}$. 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 (3.2), with the expectation taken with respect to the distribution of $\mathbf{y}_r|\mathbf{y}_s$ with $\boldsymbol{\theta}$ replaced by an estimator $\hat{\boldsymbol{\theta}}$. The EBP is not exactly unbiased, but the bias coming from the estimation of the parameter $\boldsymbol{\theta}$ is typically negligible.

Observation 3.1.1. *Assume that* $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$ *follows a Normal distribution with mean vector* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, *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

$$\hat{\delta}^B = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\left[\mathbf{X}_r\boldsymbol{\beta} + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})\right]. \tag{3.3}$$

Replacing $\boldsymbol{\beta}$ by the weighted least squares estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s$ in in (3.3), we obtain the best linear unbiased predictor (BLUP) of $\delta = \mathbf{a}'\mathbf{y}$ as defined by Royall (1976).

3.2 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

$$\hat{\delta}_d^B = E_{\mathbf{y}_{dr}}(\delta_d|\mathbf{y}_{ds}). \tag{3.4}$$

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 (3.4) can be easily derived. Thus, we consider that

$$\mathbf{y}_d \sim \text{ind } N(\boldsymbol{\mu}_d, \mathbf{V}_d), \quad d = 1, \ldots, D, \tag{3.5}$$

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

$$\boldsymbol{\mu}_d = \begin{pmatrix} \boldsymbol{\mu}_{ds} \\ \boldsymbol{\mu}_{dr} \end{pmatrix}, \quad \mathbf{V}_d = \begin{pmatrix} \mathbf{V}_{ds} & \mathbf{V}_{dsr} \\ \mathbf{V}_{drs} & \mathbf{V}_{dr} \end{pmatrix}. \tag{3.6}$$

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

$$\mathbf{y}_{dr}|\mathbf{y}_{ds} \sim N(\boldsymbol{\mu}_{dr|s}, \mathbf{V}_{dr|s}), \tag{3.7}$$

where

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

For complex non-linear domain parameters $\delta_d = h(\mathbf{y}_d)$, the expectation in (3.19) 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 (3.7). 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,

$$\hat{\delta}_d^B = E_{\mathbf{y}_r}[h(\mathbf{y}_d)|\mathbf{y}_{ds}] \approx \frac{1}{L}\sum_{\ell=1}^{L} h(\mathbf{y}_d^{(\ell)}). \tag{3.8}$$

Typically, the mean vectors and covariance matrices in (3.5) involve an unknown parameter vector $\boldsymbol{\theta}$, that is, $\boldsymbol{\mu}_d = \boldsymbol{\mu}_d(\boldsymbol{\theta})$ and $\mathbf{V}_d = \mathbf{V}_d(\boldsymbol{\theta})$. An estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is replaced in (3.7). 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 $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\theta}}$, and applying (3.8).

3.3 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}:

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

where 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

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

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

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

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 $\boldsymbol{\mu}_d = E(\mathbf{y}_d)$ and $\mathbf{V}_d = \text{Var}(\mathbf{y}_d)$ as in (3.6). The distribution of the out-of-sample vector $\mathbf{y}_{dr}$ given the sample data $\mathbf{y}_{ds}$ is given by (3.7) where, for this particular model, the conditional mean vector and covariance matrix are given by

$$\boldsymbol{\mu}_{dr|s} = \mathbf{X}_{dr}\boldsymbol{\beta} + \sigma_u^2\mathbf{1}_{N_d-n_d}\mathbf{1}_{n_d}'\mathbf{V}_{ds}^{-1}(\mathbf{y}_{ds} - \mathbf{X}_{ds}\boldsymbol{\beta}), \tag{3.10}$$

$$\mathbf{V}_{dr|s} = \sigma_u^2(1-\gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}_{N_d-n_d}' + \sigma_e^2\mathbf{I}_{N_d-n_d}, \tag{3.11}$$

with $\gamma_d = \sigma_u^2(\sigma_u^2 + \sigma_e^2/n_d)^{-1}$. Observe that the application of the Monte Carlo approximation (3.8) involves simulation of D multivariate Normal vectors of sizes $N_d - n_d$, $d = 1, \ldots, D$, from (3.7). 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 (3.7), corresponds to the covariance matrix of a vector $\mathbf{y}_{dr}$ generated by the model

$$\mathbf{y}_{dr} = \boldsymbol{\mu}_{dr|s} + v_d \mathbf{1}_{N_d-n_d} + \boldsymbol{\epsilon}_{dr}, \tag{3.12}$$

with new random effects v_d and errors $\boldsymbol{\epsilon}_{dr}$ that are independent and satisfy

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

Using model (3.12), 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 $\epsilon_{dj} \sim N(0, \sigma_\epsilon^2)$ independently, for $j \in r_d$, and then obtain the corresponding out-of-sample elements Y_{dj} from (3.12) using as means the corresponding elements of $\boldsymbol{\mu}_{dr|s}$ given by (3.10). As mentioned before, in practice the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_u^2, \sigma_e^2)'$ are replaced by suitable estimators $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)'$, and then the variables Y_{dj} are generated from (3.12) with $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\theta}}$.

3.4 Parametric bootstrap for MSE estimation

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

$$\text{MSE}(\hat{\delta}_d^{EB}) = E\left[(\hat{\delta}_d^{EB} - \delta_d)^2\right], \tag{3.13}$$

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, (3.13) can be decomposed as

$$\text{MSE}(\hat{\delta}_d^{EB}) = \left[E(\hat{\delta}_d^{EB} - \delta_d)\right]^2 + V(\hat{\delta}_d^{EB} - \delta_d). \tag{3.14}$$

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 (3.14) 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 (3.14).

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

2. Generate bootstrap random domain effects as $u_d^* \sim$ iid $N(0, \hat\sigma_u^2)$, $d = 1, \dots, D$.

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

4. Construct a bootstrap population vector $\mathbf{y}^* = ((\mathbf{y}_1^*)', \dots, (\mathbf{y}_D^*)')'$ using the estimated model,
$$Y_{dj}^* = \mathbf{x}_{dj}\hat{\boldsymbol\beta} + u_d^* + e_{dj}^*, \quad j = 1, \dots, N_d, \quad d = 1, \dots, D, \tag{3.15}$$
and calculate the true domain quantities for this bootstrap population, $\delta_d^* = h(\mathbf{y}_d^*)$, $d = 1, \dots, D$.

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 (3.9) again to bootstrap data $\mathbf{y}_s^*$, obtaining new model parameter estimates $\hat{\boldsymbol\beta}^*$, $\hat\sigma_u^{2*}$ 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 3.2 and calculate bootstrap EBPs, $\hat\delta_d^{EB*}$, $d = 1, \dots, D$.

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 (3.15) given sample data $\mathbf{y}_s$. Then bootstrap random effects u_d^* and errors e_{dj}^* are iid with
$$E_*(u_d^*) = 0, \quad Var_*(u_d^*) = \hat\sigma_u^2, \quad E_*(e_{dj}^*) = 0, \quad Var_*(e_{dj}^*) = \hat\sigma_e^2. \tag{3.16}$$
Observe also that the mean vectors and covariance matrices of the bootstrap domain vectors $\mathbf{y}_d^*$ are given by
$$E_*(\mathbf{y}_d^*) = \mathbf{X}_d\hat{\boldsymbol\beta} \quad \text{and} \quad Var_*(\mathbf{y}_d^*) = \hat\sigma_u^2 \mathbf{1}_{N_d}\mathbf{1}_{N_d}' + \hat\sigma_e^2 \mathbf{I}_N.$$
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 $MSE(\hat\delta_d^{EB})$ is the bootstrap MSE of the bootstrap EBP, that is
$$MSE_*(\hat\delta_d^{EB*}) = E_*\left[(\hat\delta_d^{EB*} - \delta_d^*)^2\right].$$

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 $\delta_d^{*(b)}$ and $\hat\delta_d^{EB*(b)}$ 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
$$mse(\hat\delta_d^{EB}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat\delta_d^{EB*(b)} - \delta_d^{*(b)}\right)^2. \tag{3.17}$$

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.

3.5 Empirical best estimators of small domain FGT poverty measures

Consider the FGT family of poverty measures for domain d

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z), \quad \alpha = 0, 1, 2, \tag{3.18}$$

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 (3.9). 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

$$\hat{F}^{B}_{dj} = E_{\mathbf{y}_{dr}} \left[h_{\alpha}(\mathbf{y}_d) | \mathbf{y}_{ds} \right] = \int_{\mathbb{R}} h_{\alpha}(\mathbf{y}_d) f(\mathbf{y}_{dr} | \mathbf{y}_{ds})\, d\mathbf{y}_{dr}, \tag{3.19}$$

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 (3.7). Due to the complexity of the function $h_\alpha(\cdot)$, there is not explicit expression for the expectation in (3.19), but this expectation can be approximated by Monte Carlo as explained in Section 3.2. Then, an approximation to the best predictor of $F_{\alpha d}$ is

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

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

3.6 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 (3.9) 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 3.4. 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,

$$\begin{aligned} &Y_{dj} = \mathbf{x}_{dj}\boldsymbol{\beta} + u_c + e_{dj}, \quad j=1,\dots,N_d, \quad d=1,\dots,D, \quad c=1,\dots,C,\\ &u_c \sim \text{iid } N(0,\sigma_c^2), \quad e_{dj} \sim \text{iid } N(0,\sigma_e^2). \end{aligned} \tag{3.20}$$

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

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

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

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

$$Y_{dj}^* = \mathbf{x}_{dj}\boldsymbol{\beta} + u_c^* + e_{dj}^*, \quad j=1,\dots,N_d, \quad d=1,\dots,D, \quad c=1,\dots,C. \tag{3.21}$$

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

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 3.2.

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

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

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

$$\hat{\bar{Y}}_d^{ELL} = E_*(\bar{Y}_d^*), \tag{3.22}$$

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

$$Y_{dj} = \mathbf{x}_{dj}\boldsymbol{\beta} + e_{dj}, \quad e_{dj} \sim \text{iid } N(0, \sigma_e^2), \quad j = 1, \ldots, N_d, \quad d = 1, \ldots, D, \tag{3.23}$$

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

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

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 $\bar{\mathbf{X}}_d\boldsymbol{\beta}$, 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 (3.23) by $\hat{Y}_{dj} = \mathbf{x}_{dj}\boldsymbol{\beta}$ 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 (3.22), 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 (3.16) of the bootstrap method. On the other hand, the EB estimator of $\bar{Y}_d$ under the linear model (3.23) 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 (3.23) over the elements in d-th domain, we can express the true mean as

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

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 $\hat{\bar{Y}}_d^{ELL} = \bar{\mathbf{X}}_d\boldsymbol{\beta}$. Then, it holds that

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

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 (3.23) 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

$$mse(\hat{\bar{Y}}_d^{ELL}) = E_*[(\bar{Y}_d^* - E_*(\bar{Y}_d^*))^2] = E_*[(\bar{E}_d^*)^2] = \frac{Var_*(e_{dj}^*)}{N_d} = \frac{\sigma_e^2}{N_d}, \tag{3.24}$$

which is the true MSE of the ELL estimator under model (3.23). 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 (3.23). 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 (3.9). However, when there are no clusters, the ELL method fits model (3.23). In this case, the true mean for d-th domain is given by

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

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

$$MSE(\hat{\bar{Y}}_d^{ELL}) = E[(u_d + \bar{E}_d)^2] = \sigma_u^2 + \frac{\sigma_e^2}{N_d}. \tag{3.25}$$

Summarizing, when the true model is (3.9), 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 (3.24) with (3.25). 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 (3.16), 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 (3.25). 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 (3.9), 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 $Var(\bar{Y}_d)$.

3.7 Simulation experiments

3.7.1 Model-based simulation experiment

A model-based simulation study has been carried out to study the performance of the proposed EBPs of small domain FGT poverty measures (3.18) with $\alpha = 0$ (poverty incidence) and $\alpha = 1$ (poverty gap). For this, 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 for the population units Y_{dj} were generated from the model (3.9) taking as auxiliary variables two dummies $X_1 \in \{0, 1\}$ and $X_2 \in \{0, 1\}$ plus an intercept. The values of these two dummies for the population units were generated from Bernouilli distributions with success probabilities increasing with the area index for X_1 and constant for X_2; that is, with probabilities

$$p_{1d} = 0.3 + 0.5\, d/80; \quad p_{2d} = 0.2, \quad d = 1, \dots, D,$$

respectively. Here the welfare variables E_{dj} are the exponential of the model responses Y_{dj}; that is, the transformation $T(\cdot)$ defined in Section 3.5 is $T(x) = \log(x)$. A set of sample indices s_d with $n_d = 50$ 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 all Monte Carlo simulations.

The intercept and the regression coefficients associated with the two auxiliary variables used to generate populations were $\boldsymbol{\beta} = (3, 0.03, -0.04)'$. In this way, the mean welfare increases when moving from the case $(X_1 = 0, X_2 = 0)$ to $(X_1 = 1, X_2 = 0)$, but decreases when moving from $(X_1 = 0, X_2 = 0)$ to $(X_1 = 0, X_2 = 1)$. This implies that the "richest" individuals are those with values $X_1 = 1$ and $X_2 = 0$. Since the probability p_{1d} of $X_1 = 1$ increases with the area index but that 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. 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 roughly equal to 0.6 times the median of the welfare variables E_{dj} for a population generated as mentioned above. In this way, the poverty incidence for the simulated populations is approximately 16%.

Under this setup, $I = 10^4$ population vectors $\mathbf{y}^{(i)}$ were generated from the true model. For each population i, we carried out the following steps:

(a) The true area poverty incidences and gaps (FGT measures for $\alpha = 0$ and $\alpha = 1$ respectively) were obtained for each area $d = 1, \dots, D$ and each population i 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 (3.9) was fitted to sample data $(\mathbf{y}_s^{(i)}, \mathbf{X}_s)$. Then, substituting the estimated model parameters in (3.10) and (3.11), $L = 50$ out-of-sample vectors $\mathbf{y}_r^{(i\ell)}$, $\ell = 1, \ldots, L$ were generated from the conditional distribution (3.7) using (3.12) 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) ELL estimators of the poverty measures were also calculated. For this, first model (3.9) was fitted to sample data $\mathbf{y}_s$ and then $A = 50$ populations or censuses were generated using the parametric bootstrap described in Section 3.4. For each population, the poverty measures were calculated and finally, the results were averaged over the $A = 50$ populations to calculate the ELL estimator $\hat{F}_{\alpha d}^{ELL(i)}$ for each i, as described in Section 3.6.

Observation 3.7.1. *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 $i = 1, \ldots, I$ of the true values of the FGT measures of order $\alpha = 0, 1$ were computed as

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

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 3.1 a) and b) report respectively the biases and the MSEs of the estimators for the poverty gap ($\alpha = 1$). Figure 3.1 a) shows that the EB estimator has the smallest absolute bias followed by ELL estimator, but compared to the corresponding values of MSE (Figure 3.1 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 3.4. It is clear from Figure 3.1 b) that the EB estimator is significantly more efficient than ELL and direct estimators. Surprisingly, Figure 3.1 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.

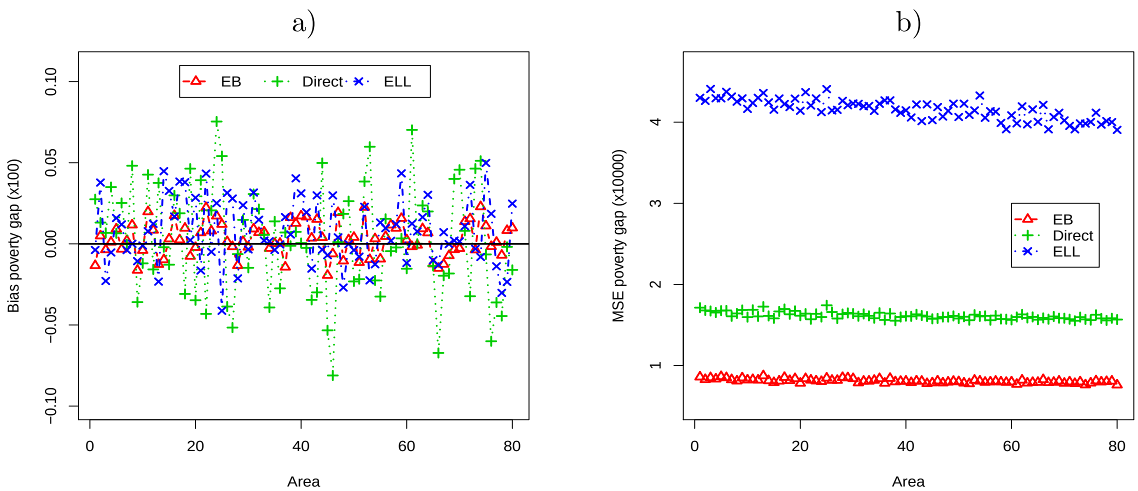

Figure 3.1: a) Bias ($\times 100$) and b) MSE ($\times 10^4$) over simulated populations of EB, direct and ELL estimators of the poverty gap F_{1d} for each area d.

Turning to MSE estimation, the parametric bootstrap procedure described in Section 3.4 was implemented with $B = 500$ replicates and the results are plotted in Figure 3.2 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 3.4 shows that the bootstrap MSE estimator tracks the pattern of the true MSE values. Similar results were obtained for the poverty incidence ($\alpha = 0$).

3.7.2 Design-based simulation experiment

A design-based simulation experiment was also carried out to study the performance of estimators over repeated samples drawn from a fixed population. Only one population was generated with the same population and sample sizes, and using the same values of model parameters as described in Section 3.7.1. Then, in each replication out of $I = 1000$, a new sample was drawn from this fixed population according to SRS without replacement within each area. From each

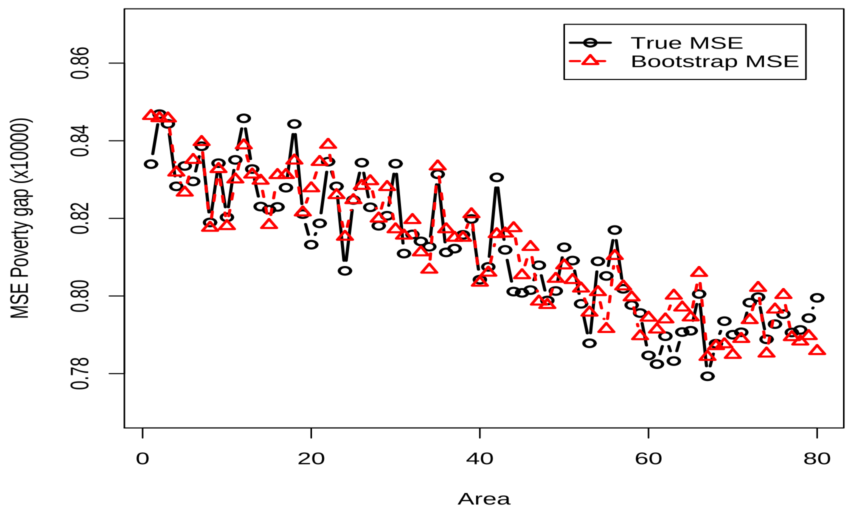

Figure 3.2: True MSE ($\times 10^4$) of EB predictor of poverty gap ($\alpha = 1$) and bootstrap MSE estimate with $B = 500$ for each area d.

sample, the three types of estimators of poverty measures, namely EBP, direct and ELL were obtained.

Results on the design bias and design MSE of the estimators for poverty gap ($\alpha = 1$) are reported in Figures 3.3 a) and b) respectively. As expected, Figure 3.3 shows that the Monte Carlo design bias of the direct estimator is practically zero, followed by EB estimator.

In terms of MSE, Figure 3.3 b) shows 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 remain small for all areas. For most areas, the MSE of EB estimator is smaller than that of the direct estimator.

3.8 Application with Spanish SILC data

The EB method was applied to estimate poverty incidences and poverty gaps by gender in Spanish provinces. For this, data from the Spanish Survey on Income and Living Conditions (SILC) from the year 2006 was used. The welfare variable for the individuals is the equivalised annual net income calculated following the standard procedure of the Spanish Statistical Institute (INE). This variable has been transformed by adding a fixed quantity to make it always positive and then taking logarithm. This transformed variable acts as the response in the nested-error regression model. As auxiliary variables, we have considered the indicators of the 5 quinquennial groupings of the variable age, the indicator of having Spanish nationality, the indicators of the 3 levels of the variable education level, and the indicators of the 3 categories of the variable employment, with categories “unemployed”, “employed” and “inactive”. For

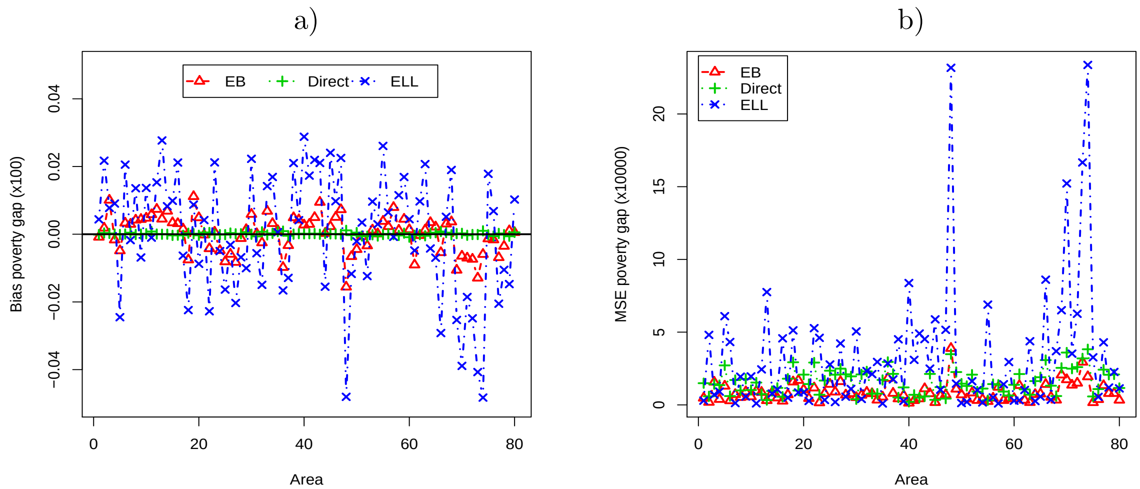

Figure 3.3: a) Bias ($\times 100$) and b) MSE ($\times 10^4$) of EB, direct and ELL estimators of the poverty gap F_{1d} for each area d under the design-based setup.

each auxiliary variable, one of the categories was considered as base reference, omitting the corresponding indicator and then including an intercept in the model.

The values of the dummy indicators are not known for the out-of-sample units, but the EB method requires only the knowledge of the total number of people with the same x-values. These totals were estimated using the sampling weights attached to the sample units in the SILC.

The MSEs of the poverty measures were estimated by using the parametric bootstrap estimator $mse(\hat{F}_{\alpha d}^{EB})$, given by (3.17), with $B = 500$ replicates. Values of EB estimators, $\hat{F}_{\alpha d}^{EB}$, and associated coefficients of variation (CVs) for the poverty incidence ($\alpha = 0$) and the poverty gap ($\alpha = 1$) are listed respectively in Tables 3.3 and 3.4 for a few representative small domains (provinces $\times$ gender), where $cv(\hat{F}_{\alpha d}^{EB}) = \{mse(\hat{F}_{\alpha d}^{EB})\}^{1/2}/\hat{F}_{\alpha d}^{EB}$. Direct estimators and their estimated variances were also calculated following standard formulas in sampling theory but taking as observations the quantities

$$F_{\alpha dj} = \left(\frac{z - E_{dj}^{(i\ell)}}{z}\right)^{\alpha} I(E_{dj}^{(i\ell)} < z), \quad j \in s_d$$

and using the SILC sampling weights attached to the units w_{dj}, namely

$$\hat{F}_{\alpha d}^{w} = \frac{1}{N_d}\sum_{j\in s_d} w_{dj}F_{\alpha dj} \quad \text{and} \quad var(\hat{F}_{\alpha d}^{w}) = \frac{1}{N_d^2} w_{dj}(w_{dj}-1)F_{\alpha dj}^2,$$

Direct estimators along with their estimated variances are also shown in Tables 3.3 and 3.4. Full results for all domains are included in Molina and Rao (2009).

The CVs of EB estimators are much smaller than those of direct estimators for all except few domains, in which the CVs are similar for both estimators. This improvement in efficiency is larger for the poverty gap than for the poverty incidence for most domains. Moreover, the reduction in CV tends to be greater for domains with smaller sample sizes. National statistical offices usually establish a maximum publishable CV. For these data, the estimated CVs of direct estimators of poverty incidences exceeded the level of 10% for 78 (out of the 104) domains while those of the EB estimators exceeded this level for only 28 domains. If we increase the level to 20%, then the direct estimators have greater CV for 17 domains but the CV of EB estimators exceeded 20% only for one domain.

Prov:Gender	n_d	$\hat{F}_{0d}^{w}$	$\hat{F}_{0d}^{EB}$	$var(\hat{F}_{0d}^{w})$	$mse(\hat{F}_{0d}^{EB})$	$cv(\hat{F}_{0d}^{w})$	$cv(\hat{F}_{0d}^{EB})$
Soria:F	17	60.41	12.84	15.8671	0.6805	20.85	16.52
Tarragona:M	129	12.46	12.50	0.8570	0.5761	23.50	16.15
Crdoba:F	230	30.66	29.22	1.0760	0.5025	10.70	6.73
Badajoz:M	472	36.58	33.74	0.6185	0.1703	6.80	3.57
Barcelona:F	1483	10.82	19.45	0.0661	0.0494	7.51	5.37

Table 3.1: Sample sizes, direct and EB estimates of poverty incidences ($\times$ 100), estimated MSEs of direct and EB estimators and CVs of direct and EB estimators ($\times$ 100) for the Spanish domains with sample size closest to minimum, first quartile, median, third quartile and maximum.

Prov:Gender	n_d	$\hat{F}_{1d}^{w}$	$\hat{F}_{1d}^{EB}$	$var(\hat{F}_{1d}^{w})$	$mse(\hat{F}_{1d}^{EB})$	$cv(\hat{F}_{1d}^{w})$	$cv(\hat{F}_{1d}^{EB})$
Soria:F	17	2.49	3.75	122.9756	5.598	47.27	19.99
Tarragona:M	129	1.53	3.65	0.2800	1.0977	27.15	23.14
Crdoba:F	230	9.63	10.53	1.1694	1.1819	13.50	8.87
Badajoz:M	472	11.72	12.68	1.2979	0.3086	9.05	3.94
Barcelona:F	1483	5.00	6.29	0.1297	0.1027	10.00	8.17

Table 3.2: Sample sizes, direct and EB estimates of poverty gaps ($\times$ 100), estimated MSEs of direct and EB estimators and CVs of direct and EB estimators ($\times$ 100) for the Spanish domains with sample size closest to minimum, first quartile, median, third quartile and maximum.

Cartograms of the estimated poverty incidences and the poverty gaps in Spanish provinces for males and females have been constructed using the EB estimates, see Figures 3.4 and 3.5. In these maps we can see that the poorer provinces concentrate mainly in the south and west parts of Spain. Provinces with critical poverty incidences (over 30%) for men are, in the south: Almera and Crdoba; west: Badajoz, vila, Salamanza and Zamora and then Cuenca, situated east of Madrid. For women the poverty incidences increase in most provinces, becoming critical also, in the south: Granada, Jan, Albacete and Ciudad Real, and in the north: Palencia and Soria. The poverty level for Lrida (north-east) seems unexpected considering that this province belongs to the region of Catalonia, which is commonly considered as a rich region.

The poverty gap measures the degree of poverty instead of the quantity of people under poverty. For a region with many people whose income is under the poverty line but very close to it, the poverty gap will be close to zero. Observe that the provinces with an income of over 12.5% under the poverty line are also among those provinces with critical values of poverty incidence, except for the northern provinces such as Lrida, which do not have significant gaps in comparison with the rest of the provinces.

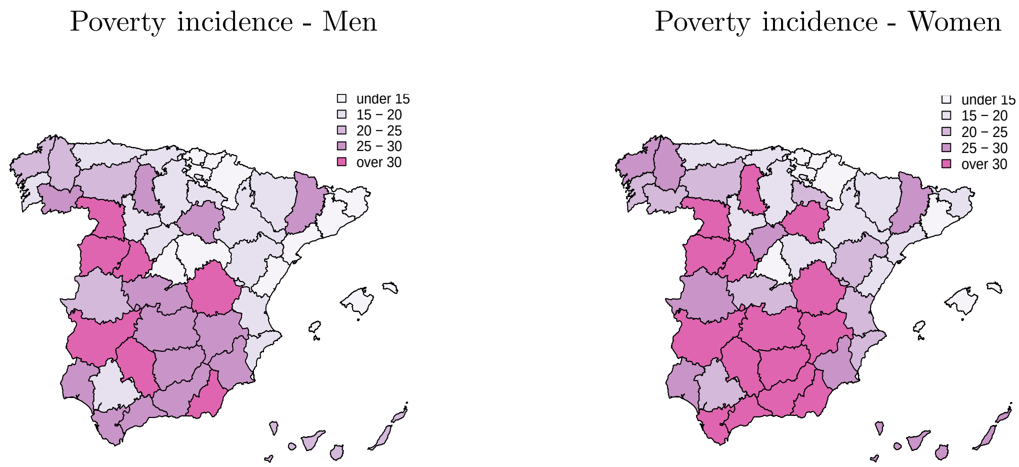

Figure 3.4: Cartograms of estimated percent poverty incidences in Spanish provinces for Men and Women.

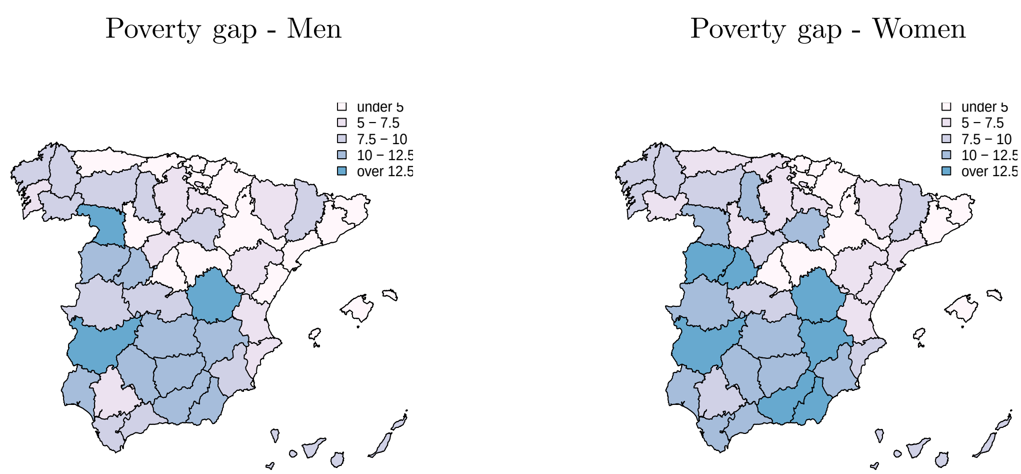

Figure 3.5: Cartograms of estimated percent poverty gaps in Spanish provinces for Men and Women.

Appendix: Application results

Table 1. Results on poverty incidence: Spanish SILC data.

Province	Sex	N_d	n_d	$\hat{F}_{0d}^{w}$	$\hat{F}_{0d}^{EB}$	$var(\hat{F}_{0d}^{w})$	$mse(\hat{F}_{0d}^{EB})$	CV($\hat{F}_{0d}^{w}$)	CV($\hat{F}_{0d}^{EB}$)	Ratio
lava	M	99354	95	8,27	12,84	1,1548	0,6805	36,60	20,32	1,80
lava	F	108422	96	7,87	12,50	1,0438	0,6000	27,08	19,60	1,38
Albacete	M	184058	163	23,74	29,22	1,2529	0,4617	10,92	7,35	1,48
Albacete	F	186503	183	28,52	33,74	1,3896	0,4618	11,12	6,37	1,75
Alicante	M	929288	526	16,00	19,45	0,2886	0,1466	9,31	6,23	1,49
Alicante	F	931405	552	18,85	22,59	0,3267	0,1601	8,78	5,60	1,57
Almera	M	341228	204	31,78	32,88	1,2489	0,3642	10,47	5,80	1,80
Almera	F	318857	193	35,39	35,72	1,3979	0,5020	10,04	6,27	1,60
vila	M	56601	56	33,50	31,48	4,8234	1,2061	24,03	11,03	2,18
vila	F	61708	60	45,29	38,51	5,4750	1,3285	24,19	9,46	2,56
Badajoz	M	351985	472	36,58	36,56	0,6185	0,1703	7,12	3,57	1,99
Badajoz	F	346810	515	39,33	39,13	0,6026	0,1947	6,38	3,57	1,79
Baleares	M	477561	609	9,36	11,55	0,1833	0,1042	11,88	8,84	1,34
Baleares	F	472843	660	11,52	14,05	0,2038	0,1130	9,59	7,57	1,27
Barcelona	M	2617681	1358	8,35	10,49	0,0568	0,0524	7,68	6,90	1,11
Barcelona	F	2752431	1483	10,82	13,10	0,0661	0,0494	6,67	5,37	1,24
Burgos	M	215155	168	12,73	16,72	0,8495	0,3736	18,16	11,56	1,57
Burgos	F	211240	167	12,43	18,33	0,7637	0,4097	16,28	11,04	1,47
Cceres	M	169833	261	25,15	24,69	0,8812	0,2099	9,04	5,87	1,54
Cceres	F	184785	302	33,23	28,24	0,9493	0,2689	7,32	5,81	1,26
Cdiz	M	642053	373	26,69	26,88	0,6100	0,2013	6,83	5,28	1,29
Cdiz	F	681522	422	30,34	31,63	0,6061	0,2316	6,62	4,81	1,38
Castelln	M	201428	113	12,19	14,79	1,1895	0,6489	18,73	17,22	1,09
Castelln	F	197726	123	12,18	17,35	1,2876	0,6008	20,21	14,13	1,43
Ciudad Real	M	265393	260	26,88	28,39	0,8722	0,3060	10,60	6,16	1,72
Ciudad Real	F	254508	239	32,37	30,18	1,1960	0,3598	9,34	6,29	1,49
Crdoba	M	356218	217	31,21	30,16	1,1389	0,3636	8,55	6,32	1,35
Crdoba	F	364583	230	30,66	33,32	1,0760	0,5025	7,52	6,73	1,12
La Corua	M	509141	457	21,57	24,66	0,4097	0,1549	8,67	5,05	1,72
La Corua	F	563190	533	23,70	25,36	0,4027	0,1789	8,78	5,27	1,66
Cuenca	M	92275	96	36,16	35,26	3,2476	0,6676	15,82	7,33	2,16
Cuenca	F	86760	87	47,17	35,35	3,5398	0,9183	12,57	8,57	1,47
Gerona	M	307975	145	5,05	13,29	0,3261	0,4512	21,83	15,98	1,37
Gerona	F	245519	138	6,72	15,38	0,5171	0,5672	21,25	15,48	1,37
Granada	M	371735	188	30,10	29,16	1,3077	0,3423	8,99	6,35	1,42
Granada	F	424598	229	34,17	36,34	1,1875	0,3340	8,09	5,03	1,61
Guadalajara	M	87591	92	7,73	12,74	0,7280	0,6339	22,55	19,76	1,14
Guadalajara	F	79560	86	16,46	15,83	1,7161	0,9055	20,43	19,01	1,08
Guipzcoa	M	323719	279	6,39	11,30	0,3138	0,2488	17,79	13,96	1,27
Guipzcoa	F	348524	291	9,95	14,56	0,4182	0,2250	15,52	10,30	1,51
Huelva	M	223158	121	19,24	29,06	1,3104	0,4976	12,03	7,68	1,57
Huelva	F	214587	123	25,31	29,13	1,5971	0,5606	12,85	8,13	1,58
Huesca	M	96617	125	7,79	17,11	0,7790	0,6126	25,64	14,47	1,77
Huesca	F	91147	105	8,92	18,99	1,0412	0,8048	23,76	14,94	1,59
Jan	M	380752	233	28,34	28,60	0,9750	0,2878	12,22	5,93	2,06
Jan	F	356344	230	33,86	32,31	1,1490	0,4010	11,79	6,20	1,90
Len	M	204462	209	19,15	22,60	1,0197	0,3772	14,51	8,59	1,69
Len	F	225753	228	19,28	24,17	0,8360	0,3639	13,62	7,89	1,72
Lrida	M	214123	127	17,67	25,74	1,3918	0,6116	20,87	9,61	2,17

Lrida	F	218051	133	23,86	27,36	1,8547	0,5777	20,98	8,79	2,39
La Rioja	M	149238	519	16,57	18,57	0,3980	0,1383	11,36	6,33	1,79
La Rioja	F	147554	500	21,25	21,45	0,4890	0,1557	10,27	5,82	1,77
Lugo	M	175462	169	20,66	24,51	1,3365	0,3914	11,74	8,07	1,45
Lugo	F	167892	177	22,47	26,87	1,1597	0,4034	12,32	7,47	1,65
Madrid	M	2816184	893	10,98	12,06	0,1843	0,0619	9,01	6,52	1,38
Madrid	F	3011923	996	12,57	13,91	0,1718	0,0704	7,76	6,04	1,29
Mlaga	M	693871	361	22,22	27,95	0,6008	0,2031	12,73	5,10	2,50
Mlaga	F	702667	397	25,76	32,45	0,5328	0,2190	10,13	4,56	2,22
Murcia	M	668714	868	21,87	25,35	0,2774	0,1027	6,60	4,00	1,65
Murcia	F	660107	902	25,55	28,70	0,3128	0,1087	6,50	3,63	1,79
Navarra	M	286947	525	8,98	9,13	0,2038	0,1405	12,21	12,98	0,94
Navarra	F	289947	603	9,40	11,40	0,1920	0,1211	11,39	9,66	1,18
Orense	M	120257	118	28,23	25,07	2,7728	0,5993	14,92	9,77	1,53
Orense	F	137587	140	21,27	22,12	1,8244	0,4809	15,57	9,92	1,57
Oviedo	M	511169	742	10,82	16,01	0,1885	0,0824	11,01	5,67	1,94
Oviedo	F	546817	864	12,20	16,59	0,1755	0,0893	9,63	5,70	1,69
Palencia	M	75638	71	22,82	26,16	2,9000	1,0455	17,72	12,36	1,43
Palencia	F	72558	72	28,00	30,13	3,3849	1,0907	16,46	10,96	1,50
Las Palmas	M	592262	458	22,39	24,65	0,6853	0,1615	9,45	5,16	1,83
Las Palmas	F	580265	485	24,57	25,40	0,6090	0,1520	7,75	4,85	1,60
Pontevedra	M	494161	434	17,36	19,15	0,4301	0,1620	11,97	6,64	1,80
Pontevedra	F	525627	462	21,37	22,66	0,5031	0,1865	10,67	6,03	1,77
Salamanca	M	151335	166	30,83	31,46	1,7522	0,3862	12,88	6,25	2,06
Salamanca	F	152234	162	32,90	33,56	1,7820	0,4030	11,59	5,98	1,94
Tenerife	M	366253	370	26,29	24,14	0,7483	0,1590	11,34	5,22	2,17
Tenerife	F	376690	392	28,64	26,36	0,6643	0,2006	10,28	5,37	1,91
Santander	M	267290	424	9,49	16,00	0,2762	0,1398	14,33	7,39	1,94
Santander	F	279191	443	12,82	16,93	0,3877	0,1678	12,09	7,65	1,58
Segovia	M	62518	57	23,43	19,24	3,7742	1,0910	15,63	17,17	0,91
Segovia	F	63217	58	43,80	26,74	5,1026	1,2032	12,97	12,97	1,00
Sevilla	M	816795	472	20,90	19,61	0,3971	0,1575	7,50	6,40	1,17
Sevilla	F	853057	491	22,80	24,04	0,4111	0,1493	6,35	5,08	1,25
Soria	M	26431	24	24,67	26,33	11,4108	2,0666	37,89	17,26	2,19
Soria	F	17211	17	60,41	31,48	15,8671	2,7052	40,37	16,52	2,44
Tarragona	M	264627	129	12,46	14,86	0,8570	0,5761	19,85	16,15	1,23
Tarragona	F	255490	139	17,36	19,28	1,0995	0,5197	15,75	11,82	1,33
Teruel	M	53380	66	8,30	17,13	1,1117	0,8420	25,53	16,94	1,51
Teruel	F	65002	78	15,09	22,26	2,0390	1,0112	22,48	14,29	1,57
Toledo	M	288335	278	24,96	26,22	0,8392	0,1871	10,40	5,22	1,99
Toledo	F	305241	272	21,99	22,50	0,7889	0,2784	11,08	7,42	1,49
Valencia	M	1169258	686	13,70	17,89	0,3019	0,0940	8,06	5,42	1,49
Valencia	F	1197478	742	13,88	20,78	0,1978	0,1162	7,16	5,19	1,38
Valladolid	M	305496	292	16,51	15,34	0,5649	0,2216	14,52	9,70	1,50
Valladolid	F	322530	306	21,02	18,29	0,7031	0,2352	11,75	8,38	1,40
Vizcaya	M	576042	515	9,18	10,01	0,1901	0,1267	13,30	11,24	1,18
Vizcaya	F	590094	532	9,86	11,57	0,2077	0,1175	12,91	9,37	1,38
Zamora	M	101433	109	33,19	34,67	2,3298	0,7388	14,30	7,84	1,82
Zamora	F	98337	100	26,82	32,84	2,1544	0,7964	13,29	8,59	1,55
Zaragoza	M	466651	555	10,07	15,42	0,2089	0,1232	13,55	7,20	1,88
Zaragoza	F	462937	574	13,57	15,34	0,2920	0,0989	11,72	6,48	1,81
Ceuta	M	35705	223	33,41	30,26	1,2221	0,3482	9,63	6,17	1,56
Ceuta	F	40426	247	38,79	33,15	1,2324	0,3804	8,92	5,88	1,52
Melilla	M	30595	179	23,61	19,27	1,3732	0,3783	10,50	10,09	1,04

Melilla	F	27498	180	25,10	25,45	1,1551	0,5579	11,09	9,28	1,19

Columns respectively denote province, gender, population size, sample size, direct estimate of poverty incidence, EB estimate, estimated variance of direct estimator, estimated MSE of EB estimator, CV of direct estimator, CV of EB estimator and ratio of CVs of direct estimators over EB estimators. Estimated poverty incidences and CVs in percentage.

Table 2. Results on poverty gap: Spanish SILC data.

Province	Sex	N_d	n_d	$\hat{F}_{1d}^{w}$	$\hat{F}_{1d}^{EB}$	$var(\hat{F}_{1d}^{w})$	$mse(\hat{F}_{1d}^{EB})$	CV($\hat{F}_{1d}^{w}$)	CV($\hat{F}_{1d}^{EB}$)	Ratio
lava	M	99354	95	2.49	3.75	1.0904	1.1907	41.94	29.09	1.44
lava	F	108422	96	1.53	3.65	0.4942	1.1315	45.97	29.12	1.58
Albacete	M	184058	163	9.63	10.53	2.9626	0.9286	17.87	9.15	1.95
Albacete	F	186503	183	11.72	12.68	3.5333	0.9250	16.03	7.59	2.11
Alicante	M	929288	526	5.00	6.29	0.5269	0.2937	14.53	8.62	1.68
Alicante	F	931405	552	5.89	7.55	0.6127	0.3407	13.28	7.73	1.72
Almera	M	341228	204	10.81	12.30	2.3507	0.6839	14.19	6.73	2.11
Almera	F	318857	193	11.18	13.64	2.8714	1.0880	15.16	7.65	1.98
vila	M	56601	56	10.82	11.64	7.0443	2.3237	24.54	13.09	1.87
vila	F	61708	60	12.30	15.40	6.1424	2.8775	20.15	11.01	1.83
Badajoz	M	351985	472	12.59	14.11	1.2979	0.3086	9.05	3.94	2.30
Badajoz	F	346810	515	12.15	15.46	1.0543	0.4007	8.45	4.09	2.06
Baleares	M	477561	609	2.88	3.34	0.4130	0.1955	22.28	13.24	1.68
Baleares	F	472843	660	2.94	4.23	0.2716	0.2227	17.72	11.17	1.59
Barcelona	M	2617681	1358	3.07	3.00	0.1224	0.0997	11.38	10.53	1.08
Barcelona	F	2752431	1483	3.60	3.92	0.1297	0.1027	10.00	8.17	1.22
Burgos	M	215155	168	4.22	5.21	2.2735	0.6704	35.72	15.70	2.27
Burgos	F	211240	167	3.50	5.81	1.4983	0.8377	34.93	15.75	2.22
Cceres	M	169833	261	7.54	8.52	1.2188	0.3704	14.65	7.14	2.05
Cceres	F	184785	302	9.33	10.13	1.2620	0.5620	12.03	7.40	1.63
Cdiz	M	642053	373	7.24	9.38	0.9284	0.4024	13.31	6.76	1.97
Cdiz	F	681522	422	10.95	11.65	1.4154	0.5101	10.87	6.13	1.77
Castelln	M	201428	113	3.97	4.48	2.8120	1.2242	42.19	24.68	1.71
Castelln	F	197726	123	3.86	5.51	2.0386	1.3159	36.97	20.83	1.77
Ciudad Real	M	265393	260	7.30	10.07	0.9995	0.6338	13.70	7.91	1.73
Ciudad Real	F	254508	239	7.15	10.86	0.9134	0.7758	13.36	8.11	1.65
Crdoba	M	356218	217	8.22	10.82	1.2822	0.6983	13.77	7.72	1.78
Crdoba	F	364583	230	8.01	12.26	1.1694	1.1819	13.50	8.87	1.52
La Corua	M	509141	457	7.34	8.47	0.7480	0.2867	11.78	6.32	1.86
La Corua	F	563190	533	8.33	8.72	0.8716	0.3791	11.20	7.06	1.59
Cuenca	M	92275	96	8.83	13.41	2.4195	1.4071	17.62	8.84	1.99
Cuenca	F	86760	87	10.73	13.36	3.0724	2.0791	16.33	10.80	1.51
Gerona	M	307975	145	1.87	3.95	0.5700	0.7954	40.35	22.56	1.79
Gerona	F	245519	138	2.15	4.67	0.7537	1.0857	40.30	22.29	1.81
Granada	M	371735	188	13.55	10.56	4.0423	0.6923	14.84	7.88	1.88
Granada	F	424598	229	16.81	14.02	4.8343	0.7568	13.08	6.20	2.11
Guadalajara	M	87591	92	1.52	3.80	0.2823	1.2206	34.88	29.10	1.20
Guadalajara	F	79560	86	2.55	4.90	0.4615	1.7868	26.63	27.28	0.98
Guipzcoa	M	323719	279	2.60	3.25	0.9591	0.4277	37.69	20.09	1.88
Guipzcoa	F	348524	291	4.42	4.38	1.3093	0.4294	25.90	14.95	1.73
Huelva	M	223158	121	10.46	10.37	7.2743	0.9412	25.78	9.36	2.75
Huelva	F	214587	123	9.13	10.40	4.2187	1.0980	22.49	10.07	2.23
Huesca	M	96617	125	2.56	5.39	1.2775	1.2615	44.18	20.86	2.12

Huesca	F	91147	105	3.04	6.06	1.7064	1.5781	42.92	20.72	2.07
Jan	M	380752	233	9.63	10.28	1.8186	0.5968	14.01	7.51	1.86
Jan	F	356344	230	11.41	11.94	2.1644	0.8385	12.89	7.67	1.68
Len	M	204462	209	7.14	7.58	2.2850	0.7474	21.16	11.41	1.85
Len	F	225753	228	7.56	8.31	2.2879	0.8014	20.00	10.77	1.86
Lrida	M	214123	127	9.22	9.08	4.8531	1.2797	23.88	12.46	1.92
Lrida	F	218051	133	9.34	9.77	4.5156	1.2979	22.75	11.66	1.95
La Rioja	M	149238	519	4.05	5.97	0.3139	0.2546	13.83	8.46	1.64
La Rioja	F	147554	500	4.34	7.14	0.2958	0.3245	12.52	7.98	1.57
Lugo	M	175462	169	8.64	8.40	6.9390	0.7199	30.50	10.10	3.02
Lugo	F	167892	177	5.26	9.40	1.3626	0.8026	22.20	9.54	2.33
Madrid	M	2816184	893	3.37	3.58	0.3812	0.1145	18.33	9.45	1.94
Madrid	F	3011923	996	3.59	4.26	0.3350	0.1442	16.14	8.92	1.81
Mlaga	M	693871	361	8.95	9.90	1.9024	0.4162	15.41	6.52	2.37
Mlaga	F	702667	397	10.80	12.04	1.9554	0.4561	12.95	5.61	2.31
Murcia	M	668714	868	7.54	8.74	0.4296	0.2175	8.69	5.34	1.63
Murcia	F	660107	902	8.30	10.31	0.4393	0.2373	7.99	4.73	1.69
Navarra	M	286947	525	2.99	2.53	0.3732	0.2389	20.45	19.28	1.06
Navarra	F	289947	603	2.73	3.31	0.2752	0.2450	19.23	14.96	1.29
Orense	M	120257	118	7.28	8.66	3.6924	1.1440	26.41	12.36	2.14
Orense	F	137587	140	4.77	7.44	2.0973	0.9954	30.34	13.41	2.26
Oviedo	M	511169	742	2.54	4.95	0.2335	0.1618	19.02	8.12	2.34
Oviedo	F	546817	864	3.11	5.14	0.2300	0.1857	15.41	8.38	1.84
Palencia	M	75638	71	5.65	9.10	2.9335	2.1179	30.32	15.99	1.90
Palencia	F	72558	72	6.08	10.92	3.1612	2.5262	29.27	14.56	2.01
Las Palmas	M	592262	458	7.63	8.40	1.5179	0.3192	16.15	6.73	2.40
Las Palmas	F	580265	485	8.46	8.78	1.6326	0.3139	15.11	6.38	2.37
Pontevedra	M	494161	434	3.00	6.09	0.1949	0.3100	14.73	9.14	1.61
Pontevedra	F	525627	462	4.40	7.57	0.3114	0.3888	12.68	8.24	1.54
Salamanca	M	151335	166	9.87	11.50	2.3273	0.7451	15.46	7.51	2.06
Salamanca	F	152234	162	8.85	12.74	2.0322	0.8510	16.11	7.24	2.22
Tenerife	M	366253	370	8.07	8.20	1.0073	0.2909	12.44	6.58	1.89
Tenerife	F	376690	392	9.35	9.18	1.2484	0.4136	11.95	7.00	1.71
Santander	M	267290	424	2.59	4.94	0.3045	0.2634	21.30	10.38	2.05
Santander	F	279191	443	2.95	5.29	0.3210	0.3526	19.20	11.22	1.71
Segovia	M	62518	57	7.01	6.30	4.5293	2.1717	30.36	23.41	1.30
Segovia	F	63217	58	10.90	9.64	5.5058	2.9114	21.52	17.71	1.22
Sevilla	M	816795	472	3.42	6.34	0.1609	0.2819	11.72	8.38	1.40
Sevilla	F	853057	491	4.53	8.14	0.3045	0.3112	12.19	6.85	1.78
Soria	M	26431	24	15.28	9.13	76.9805	3.9189	57.42	21.68	2.65
Soria	F	17211	17	23.46	11.84	122.9756	5.5980	47.27	19.99	2.37
Tarragona	M	264627	129	1.95	4.53	0.2800	1.0997	27.15	23.14	1.17
Tarragona	F	255490	139	2.79	6.16	0.4304	1.1317	23.52	17.26	1.36
Teruel	M	53380	66	4.48	5.49	5.9649	1.6368	54.54	23.29	2.34
Teruel	F	65002	78	5.16	7.38	3.8559	1.9972	38.09	19.15	1.99
Toledo	M	288335	278	7.69	9.18	1.3151	0.3438	14.92	6.39	2.33
Toledo	F	305241	272	5.85	7.58	0.8997	0.6082	16.21	10.28	1.58
Valencia	M	1169258	686	5.08	5.70	0.9538	0.1722	19.24	7.28	2.64
Valencia	F	1197478	742	4.26	6.78	0.3187	0.2456	13.25	7.31	1.81
Valladolid	M	305496	292	6.38	4.71	1.1430	0.4269	16.75	13.87	1.21
Valladolid	F	322530	306	7.45	5.82	1.3767	0.4782	15.76	11.89	1.33
Vizcaya	M	576042	515	2.57	2.80	0.2783	0.2338	20.49	17.27	1.19
Vizcaya	F	590094	532	2.26	3.35	0.1756	0.2177	18.56	13.92	1.33
Zamora	M	101433	109	12.58	13.10	5.5333	1.5147	18.71	9.40	1.99

Zamora	F	98337	100	9.86	12.18	4.7252	1.7185	22.04	10.76	2.05
Zaragoza	M	466651	555	4.29	4.77	0.7891	0.2377	20.69	10.23	2.02
Zaragoza	F	462937	574	5.08	4.72	0.9837	0.1956	19.53	9.38	2.08
Ceuta	M	35705	223	14.79	11.09	3.3694	0.7296	12.41	7.70	1.61
Ceuta	F	40426	247	20.68	12.52	5.5107	0.8832	11.35	7.50	1.51
Melilla	M	30595	179	11.87	6.22	7.3207	0.7442	22.80	13.86	1.64
Melilla	F	27498	180	12.47	8.82	3.5770	1.1392	15.16	12.10	1.25

Columns respectively denote province, gender, population size, sample size, direct estimate of poverty gap, EB estimate, estimated variance of direct estimator, estimated MSE of EB estimator, CV of direct estimator, CV of EB estimator and ratio of CVs of direct estimators over EB estimators. Estimated poverty gaps and CVs in percentage.

Chapter 4

Estimation of the cumulative distribution function of income at small area level

4.1 Introduction

The structure of this chapter is as follows. In section 4.2 we present M-quantile regression, nonparametric M-quantile regression and M-quantile Geographically Weighted regression. In section 4.3 we describe how quantile or M-quantile models can be employed for measuring area effects and estimators of cumulative distribution function. Its estimation is often an important objective in survey practice. The distribution function allows to identify subgroups in the population whose values for a particular variable lie below or above a given limit. In sections 4.4 and 4.5 we discuss mean squared error estimation for M-quantile small area predictors. In section 4.6 we report a first empirical evaluation for the estimation of the mean squared error for the mean and quantile estimates. In section 4.7 we describe the EU-SILC data and the Census data which are used to produce the small area estimates and we present the first results.

4.2 Parametric and nonparametric M-quantile regression models

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 (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). This model is presented in section 4.2.1.

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, 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 4.2.2.

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 4.2.3.

4.2.1 Linear M-quantile regression models

A recently proposed approach to small area estimation is based on the use of M-quantile models (Chambers and Tzavidis, 2006). M-quantile regression provides a "quantile-like" generalization of regression based on influence functions (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

$$Q_q(x;\psi) = \mathbf{X}\boldsymbol{\beta}_\psi(q) \tag{4.1}$$

where $\psi_q(r_{iq\psi}) = 2\psi\{s^{-1}r_{iq\psi}\}\ \{qI(r_{jq\psi}>0)+(1-q)I(r_{jq\psi}\leq 0)\}$ and s is a suitable robust estimate of scale, e.g. the MAD estimate $s = median\,|r_{jq\psi}|\,/0.6745$. 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 $\hat{\boldsymbol{\beta}}_\psi(q)$ of $\boldsymbol{\beta}_\psi(q)$ is obtained via iterative weighted least squares. Note that there is a different set of regression parameters for each q.

4.2.2 Nonparametric M-quantile regression models

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 $\boldsymbol{\beta}$ coefficients. Pratesi et al. (2008) 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

$$m_{\psi,q}[x_1; \boldsymbol{\beta}_{\psi}(q), \boldsymbol{\beta}\gamma_{\psi}(q)] = \beta_{0\psi}(q) + \beta_{1\psi}(q)x_1 + \ldots + \beta_{p\psi}(q)x_1^p + \sum_{k=1}^{K} \gamma_{k\psi}(q)(x_1 - \kappa_k)_+^p, \quad (4.2)$$

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, $\boldsymbol{\beta}_{\psi}(q) = (\beta_{0\psi}(q), \beta_{1\psi}(q), \ldots, \beta_{p\psi}(q))^t$ is the coefficient vector of the parametric portion of the model and $\boldsymbol{\beta}\gamma_{\psi}(q) = (\gamma_{1\psi}(q), \ldots, \gamma_{K\psi}(q))^t$ 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 (4.2) 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).

An algorithm based on iteratively reweighted penalized least squares is proposed in Pratesi et al. (2008) to effectively compute the parameter estimates. Once those estimates are obtained, $\hat{m}_{\psi,q}[x_1] = m_{\psi,q}[x_1; \hat{\boldsymbol{\beta}}_{\psi}(q), \hat{\boldsymbol{\gamma}}_{\psi}(q)]$ can be computed as an estimate for $Q_q(x_1, \psi)$.

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 as environment, economic and public health. It has particular relevance when referenced responses need to be converted to maps. The use of bivariate p-spline approximations to fit nonparametric unit level nested error and M-quantile regression models allows for reflecting spatial variation in the data and then uses these nonparametric models for small area estimation.

4.2.3 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 (Petrucci and Salvati, 2006; Pratesi and Salvati, 2008; 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) (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.

The GWR model is a model for the conditional expectation of $\mathbf{y}$ given $\mathbf{X}$ at location u. This is easily generalised to a model for the M-quantile of order q of the conditional distribution of $\mathbf{y}$ given $\mathbf{X}$ at u. That is, we write

$$Q_q(\mathbf{X}; \psi, u) = \mathbf{X}\boldsymbol{\beta}_\psi(u; q) \tag{4.3}$$

where $\boldsymbol{\beta}_\psi(u; q)$ varies with u as well as with q. That is, model (4.3) allows the entire conditional distribution (not just the mean) of y given $\mathbf{X}$ to vary from location to location. The parameter $\boldsymbol{\beta}_\psi(u; q)$ in (4.3) can be estimated by solving normal equations by an iteratively re-weighted least squares algorithm that combines the iteratively re-weighted least squares algorithm used to fit a "spatially stationary" M-quantile model and the weighted least squares algorithm used to fit a GWR model.

The model (4.3) was then used to define a predictor of the small area characteristic of interest that accounts for spatial structure of the data. The M-quantile GWR small area model integrates the concepts of robust small area estimation and borrowing strength over space within a unified modeling framework. Extending further the M-quantile GWR for poverty measures will enable the comparison of alternative robust models for borrowing strength over space in small area estimation and will signicantly improve the collection of small area estimation tools.

4.3 Estimating the small area CDF and poverty indicators using M-quantile models

In this section we describe approaches to estimating the small area distribution function using the different models described in section 4.2. In doing so, we follow a unified estimation framework for estimating any small area target parameter that was defined by Tzavidis et al. (2009).

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

$$F_d = N_d^{-1}\Big[\sum_{j \in s_d} \mathrm{I}(y_j \leqslant t) + \sum_{j \in r_d} \mathrm{I}(y_j \leqslant t)\Big]. \tag{4.4}$$

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 (4.4) by their predicted values $\hat{y}_j$ under an appropriate model, leading to a plug-in estimator of (4.4) of the form

$$\hat{F}_d = N_d^{-1}\Big[\sum_{j \in s_d} \mathrm{I}(y_j \leqslant t) + \sum_{j \in r_d} \mathrm{I}(\hat{y}_j \leqslant t)\Big]. \tag{4.5}$$

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 (4.5). This leads to the usual plug-in estimator of the mean,

$$\hat{\bar{Y}}_d = \int_{-\infty}^{\infty} td\hat{F}_d(t) = N_d^{-1}\left(\sum_{j\in s_d} y_j + \sum_{j\in r_d} \hat{y}_j\right).$$

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 $\boldsymbol{\beta}_\psi(q)$ a sufficiently smooth function of q, this suggests an estimator of the distribution function

$$\hat{F}_d^{MQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{j\in r_d} I(\mathbf{x}_j\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d) \le t)\right\} \tag{4.6}$$

where $\mathbf{x}_j\boldsymbol{\beta}_\psi(\theta_d)$ 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 $\hat{\theta}_d = 0.5$.

An nonparametric extension of the M-quantile small area model was proposed proposed by Pratesi et al. (2008). Under the nonparametric M-quantile small area model,

$$\hat{F}_d^{NPMQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{j\in r_d} I(\mathbf{x}_j\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d) + \mathbf{z}_j\hat{\boldsymbol{\gamma}}_\psi(\hat{\theta}_d) \le t)\right\} \tag{4.7}$$

where $\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d)$ and $\hat{\boldsymbol{\gamma}}_\psi(\hat{\theta}_d)$ are the coefficient vectors of the parametric and spline proportion of the fitted p-splines M-quantile regression function at $\hat{\theta}_d$. Using the empirical distribution function and the linear or nonparametric M-quantile small area models one can defined two estimators of the small area mean

$$\hat{\bar{Y}}_d^{MQ}(t) = \int_{-\infty}^{\infty} td\hat{F}_d^{MQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} y_j + \sum_{j\in r_d} \mathbf{x}_j\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d)\right\} \tag{4.8}$$

and

$$\hat{\bar{Y}}_d^{NPMQ}(t) = \int_{-\infty}^{\infty} td\hat{F}_d^{NPMQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} y_j + \sum_{j\in r_d}\left(\mathbf{x}_j\hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d) + \mathbf{z}_j\hat{\boldsymbol{\gamma}}_\psi(\hat{\theta}_d)\right)\right\}. \tag{4.9}$$

We refer to small area estimators that can be expressed as functionals of (4.5), with non-sample predictions derived as estimates of expected values.

Chambers and Tzavidis (2006) observed that the naive M-quantile mean estimator (4.8) can be biased. The distribution function estimator (4.5) underlying (4.6) and (4.7) is not consistent in general. Thus, when the non-sample predicted values in (4.5) are estimated expectations that converge in probability to the actual expected values, we see that

$$\sum_{j\in r_d} \mathrm{I}(\hat{y}_j \leqslant t) = \sum_{j\in r_d} \mathrm{I}(y_j - (y_j - \hat{y}_j) \leqslant t) = \sum_{j\in r_d} \mathrm{I}(y_j \leqslant t + \epsilon_j) \neq \sum_{j\in r_d} \mathrm{I}(y_j \leqslant t),$$

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 (4.5) 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 (4.4), and assuming that the residuals are homoskedastic within the small area of interest, this is of the form

$$\hat{F}_d^{CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \leq t) + \sum_{k\in r_d} n_d^{-1} \sum_{j\in s_d} I(\hat{y}_k + (y_j - \hat{y}_j) \leq t)\right\}. \tag{4.10}$$

It can be shown that under the CD estimator of the small area distribution function the mean functional defined by (4.10) takes the value

$$\hat{\bar{Y}}_d^{CD} = \int_{-\infty}^{\infty} t d\hat{F}_d^{CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} y_j + \sum_{j\in r_d} \hat{y}_j + (f_d^{-1} - 1)\sum_{j\in s_d}(y_j - \hat{y}_j)\right\} \tag{4.11}$$

where $f_d = n_d N_d^{-1}$ is the sampling fraction in area d, $\hat{y}_j = \mathbf{x}_j \hat{\boldsymbol{\beta}}_\psi(\hat{\theta}_d)$, where $\hat{y}_j$ can be obtained either under the linear or the nonparametric M-quantile small area models. We refer to (4.11) as the bias adjusted M-quantile mean predictor. Due to the bias correction in (4.11), this predictor will have higher variability than (4.8) or (4.9) and so it should only be used when (4.6) or (4.7) are 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 (4.11) by using robust (huberized) residuals instead of raw residuals. In particular,

$$\hat{F}_d^{CDRob}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \leq t) + \sum_{k\in r_d} n_d^{-1} \sum_{j\in s_d} I\left(\hat{y}_k + \nu_j\psi\{y_j - \hat{y}_j\} \leqslant t\right)\right\} \tag{4.12}$$

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 (4.10) 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

$$\hat{F}_d^{RKM}(t) = n_d^{-1} \left\{ \sum_{j \in s_d} I(y_j \le t) + N_d^{-1} \sum_{k \in r_d} n^{-1} \sum_{j \in s_d} I(y_j - \hat{y}_j \le t - \hat{y}_k) \right.$$
$$\left. -(n_d^{-1} - N_d^{-1}) \sum_{k \in s_d} n_d^{-1} \sum_{j \in s_d} I(y_j - \hat{y}_j \le t - \hat{y}_k) \right\}. \tag{4.13}$$

Chambers et al. (1992) compared the large-sample mean squared errors of (4.10) and (4.13) and concluded that neither dominates the other. When the model is correctly specified we expect (4.10) to outperform (4.13). However RKM demonstrated that (4.10) can be substantially biased when model assumptions fail, while (4.13) 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 (4.11) 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

$$\int_{-\infty}^{\hat{\mu}_{qd}} d\hat{F}_d(t) = q, \tag{4.14}$$

where $\hat{F}_d(t)$ is suitable estimator of the area d distribution of y such as the CD or the RKM estimators. 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 either by an M-quantile linear or nonparametric small area model.

In the final part of this section we would like to discuss the estimation of poverty indicators at the small area level. As a starting point, in this report we discuss only the incidence of poverty or *Head Count Ratio* (HCR) F_0 as defined by Foster et al. (1984). Denoting by t the poverty line, the incidence of poverty is defined as

$$F_{0d} = N_d^{-1} \sum_{j=1}^{N_d} \mathrm{I}(y_{jd} \leqslant t). \tag{4.15}$$

Using the decomposition for sample and out of sample units we have that

$$F_{0d} = N_d^{-1} \Big[\sum_{j \in s_d} F_{0d} + \sum_{j \in r_d} F_{0d} \Big].$$

The aim then is to estimate the conditional expectation of F_{0d} for out of sample units given the sample data under the M-quantile model. An estimator of F_{0d} under the M-quantile model is given by

$$\hat{F}_{0j}^{MQ} = N_j^{-1}\Big[\sum_{i\in s_j} F_{0j} + \sum_{i\in r_j} \hat{F}_{0j}^{MQ}\Big], \tag{4.16}$$

where $\hat{F}_{0j}^{MQ} = E_{y_r}(F_{0dj}|y_s)$. Under the M-quantile model an empirical estimator of this expectation can be obtained using the following Monte-Carlo approximation.

- Fit the M-quantile small area model using the raw y sample values and obtain estimates of $\boldsymbol{\beta}$ and q_d.
- Draw a vector of $N_d - n_d$ errors, ϵ_{jd}^*, from the Empirical Distribution Function (EDF) of the estimated M-quantile regression residuals.
- Draw a vector of $N_d - n_d$ out of sample units using

$$y_{dr}^* = \mathbf{x}_{jd}\hat{\boldsymbol{\beta}}(\hat{q}_d) + \epsilon_{jd}^*.$$

- Repeat the process H times and each time combine the sample data y with y_{dr}^* for estimating the target

$$\hat{F}_{0d}^{MQ} = N_d^{-1}\Big[\sum_{j\in s_d} \mathrm{I}(y_j \leqslant t) + \sum_{j\in r_d} \mathrm{I}(\mathbf{x}_{jd}\hat{\boldsymbol{\beta}}(\hat{q}_d) + \epsilon_{jd}^* \leqslant t)\Big].$$

- Average the results over H simulations.

Before closing this section we should elaborate on some aspects of the approach used for estimating the incidence of poverty under the M-quantile small area model, described above. To start we note that one can use different approaches for drawing ϵ_{jd}^*. One can draw conditional (upon the small area) or unconditional residuals from the EDF or from a smoothed version of the EDF. These alternatives will be studied in future work of this project. The outlined approach for estimating the incidence of poverty, although nonparametric, is similar in spirit to the EBP approach proposed by Molina and Rao (2009, see also Chapter 3 of this Report). Note for example that y_{dr}^* is drawn using $\mathbf{x}_{jd}\hat{\boldsymbol{\beta}}(\hat{q}_d)$ i.e. from the conditional M-quantile model, where $\hat{q}_d$ play the role of random effects in the M-quantile small area model. 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 comparative analysis of the different approaches to poverty estimation will be performed as part of the work for this project.

4.4 Mean Squared Error (MSE) estimation for estimators of small area means

A robust mean squared error estimation method for the naive M-quantile estimator (4.8) 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 (4.10) 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

$$\hat{\boldsymbol{\beta}}_{\psi}(\hat{\theta}_d) = (\mathbf{X}_s'\mathbf{W}_{s_d}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{W}_{s_d}\mathbf{y}_s$$

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 (4.8) can be written

$$\hat{\overline{Y}}_d^{MQ/CD} = \mathbf{w}_{s_d}'\mathbf{y}_s, \tag{4.17}$$

where $\mathbf{w}_{s_d} = (w_{jd}) = n_d^{-1}\Delta_{s_d} + (1 - N_d^{-1}n_d)\mathbf{W}_d\mathbf{X}_s(\mathbf{X}_s'\mathbf{W}_d\mathbf{X}_s)^{-1}\{\overline{\mathbf{x}}_{r_d} - \overline{\mathbf{x}}_{s_d}\}$ with Δ_{sd} denoting the n-vector that 'picks out' the sample units from area d. Here $\overline{\mathbf{x}}_{s_d}$ and $\overline{\mathbf{x}}_{r_d}$ 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 (4.17) then treats the weights as fixed and applies standard methods of robust mean squared error estimation for linear estimators of population quantities (Royall and Cumberland, 1978). With this approach, the prediction variance of $\hat{\overline{Y}}_d^{CD}$ is estimated by

$$\widehat{Var}(\hat{\overline{Y}}_d^{CD}) = \sum_{g=1}^{d}\sum_{j \in s_g} \lambda_{jdg}\left(y_j - \mathbf{x}_j\hat{\beta}_{\psi}(\hat{\theta}_g)\right)^2, \tag{4.18}$$

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 $\hat{\overline{Y}}_d^{CD}$ are locally calibrated on x, it immediately follows that (4.10) 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 $\hat{\overline{Y}}_d^{MQ}$ described in Chambers and Tzavidis (2006), which includes a squared bias term.

Following the approach described in Chambers et al. (2008), for fixed q and λ, the $\hat{\overline{Y}}_d^{NPMQ}$ in (4.9) can be written as linear combination of the observed y_j like expression (4.17) where the weights can be written as

$$\mathbf{w}_{s_d} = (w_{jd}) = n_d^{-1}\Delta_{s_d} + (1 - N_d^{-1}n_d)\mathbf{W}(\hat{\theta}_d)\,[\mathbf{X}\ \mathbf{Z}]\left([\mathbf{X}\ \mathbf{Z}]'\,\mathbf{W}(\hat{\theta}_d)\,[\mathbf{X}\ \mathbf{Z}] + \lambda\mathbf{G}\right)^{-1}\{\overline{\mathbf{x}}_{r_d} - \overline{\mathbf{x}}_{s_d}\}$$

with $\mathbf{W}(\hat{\theta}_d)$ 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, $\mathbf{G} = \mathtt{diag}\{\mathbf{0}_P, \mathbf{1}_K\}$ with P the number of columns of $\mathbf{X}$ and K the number of columns of $\mathbf{Z}$. The derived weights are treated as fixed and a plug-in estimator of the mean squared error of estimator (4.9) given by

$$\widehat{Var}(\hat{\bar{Y}}_d^{NPMQ}) = \sum_{g=1}^{d} \sum_{j \in s_g} \lambda_{jdg} \left(y_j - \mathbf{x}_j \hat{\beta}_\psi(\hat{\theta}_g) - \mathbf{z}_j \hat{\gamma}_\psi(\hat{\theta}_g) \right)^2 \tag{4.19}$$

where $\lambda_{jdg} = \{(w_{jd} - 1)^2 + (n_d - 1)^{-1}(N_d - n_d)\}\mathrm{I}(g = d) + w_{jg}^2 \mathrm{I}(g \neq d)$.

4.5 Mean Squared Error (MSE) estimation for estimators of small area quantiles

The linearization-based prediction variance estimator (4.18) 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 (4.14). In this section we describe a nonparametric bootstrap approach to MSE estimation of small area quantiles that was described in Tzavidis et al. (2009) and is based on the approach of Lombardia 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 $\epsilon_{jd} = y_{jd} - \hat{\beta}_\psi(\hat{\theta}_d)$.

A bootstrap finite population $U^* = (y^*_{jd}, \mathbf{x}_{jd}), j \in U, d = 1, \cdots, D$ with

$$y^*_{jd} = \mathbf{x}^*_{jd} \hat{\beta}_\psi(\hat{\theta}_d) + \epsilon^*_{jd}$$

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

$$\hat{G}(u) = n^{-1} \sum_{d=1}^{D} \sum_{j \in s_d} \mathrm{I}(\epsilon_{jd} - \bar{\epsilon}_s \leq u)$$

where $\bar{\epsilon}_s$ is the sample mean of the ϵ_{jd}. Similarly, the empirical conditional distribution of these residuals in area d is

$$\hat{G}_d(u) = n_d^{-1} \sum_{j \in s_d} \mathrm{I}(\epsilon_j - \bar{\epsilon}_{sd} \leq u)$$

where $\bar{\epsilon}_{sd}$ is the sample mean of the ϵ_{jd} in area d. A smoothed estimator of the unconditional distribution is

$$\hat{G}(u) = n^{-1} \sum_{d=1}^{D} \sum_{j \in s_d} K\Big(\frac{u - (\epsilon_{jd} - \bar{\epsilon}_s)}{h}\Big)$$

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

$$\hat{G}_d(u) = n_d^{-1} \sum_{j \in s_d} K\Big(\frac{u - (\epsilon_{jd} - \bar{\epsilon}_s)}{h_d}\Big)$$

where $h_d > 0$ and K are the same as above and 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

$$CV(h) = n^{-1} \sum_{d=1}^{D} \sum_{j \in s_d} \int \Big(\mathrm{I}\Big((\epsilon_{jd} - \bar{\epsilon}_s) \leq u\Big) - \hat{G}_{-j}(u)\Big)^2 du,$$

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 (Lee and Racine, 2007, section 1.5) 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 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 (4.10) 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

$$\widehat{Var}_d = B^{-1}L^{-1}\sum_{b=1}^{B}\sum_{l=1}^{L}\left(\hat{F}_d^{*bl,CD}(t) - \hat{\bar{F}}_d^{*bl,CD}(t)\right)^2,$$

where

$$\hat{\bar{F}}_d^{*bl,CD}(t) = L^{-1}\sum \hat{F}_d^{*bl,CD}(t)$$

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

$$\widehat{MSE}_d\left(\hat{F}_d^{CD}(t)\right) = \widehat{Var}_d + \widehat{Bias}_d^2. \tag{4.20}$$

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. Finally, this bootstrap MSE estimator can be used for MSE estimation of the estimates of the incidence of poverty/HCR. However, this will be considered in future work of this project.

4.6 First Empirical Evaluations

In this section we present some first results from simulation studies that were used to compare the performance of the different small area estimators discussed in the preceding sections. In particular, we have designed 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 population data. In both, $N = 232{,}500$ population values of x and y in 30 small areas were generated 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 method of population simulation (scenario 1) generated population values of x in small area h as $x_{jd} = N(\mu_d, \frac{\mu_d^2}{36})$, where $\gamma_d = N(0, 1)$, $\epsilon_{jd} = N(0, 64)$, and with $\mu_d = U(40, 120)$ held fixed over the simulations. The second (scenario 2) generated these values as $x_{jd} = \chi^2(Z_d)$,

$\epsilon_{jd} = \chi^2(3) - 3$, $\gamma_d = \chi^2(1) - 1$, with $Z_d = U(1, 200)$ held fixed over the simulations. The purpose of scenario 2 was to examine the effect of misspecification of the Gaussian assumptions of a mixed model. Population values of y in small area d in both scenarios were then generated from $y_{jd} = 5 + x_{jd} + \gamma_d + \epsilon_{jd}$.

A linear M-quantile regression model was fitted to the sample data obtained in the simulations. The M-quantile linear regression fit was obtained using a modified version of the *rlm* function (Venables and Ripley, 2002, section 8.3) in `np` package of the `R` software environment. Estimated model coefficients obtained from these fits were then used to compute naive, CD and RKM-based versions of the 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 4.1 (scenario 1) and in Table 4.2 (scenario 2). Under scenario 1 all approaches perform reasonably well. The differences between the naive, CD and RKM versions of the M-quantile regression-based estimators are much more pronounced under scenario 2 (area effects distributed as Chi-square). Here we see that use of naive estimators leads to substantial biases as far as quantiles are concerned. In contrast, the CD and RKM-based estimators are 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 either the CD or the RKM distribution function estimators are preferable if there is concern about misspecification of the distribution of area effects.

In order to evaluate the performance of the analytic mean squared error (4.18) and of the bootstrap mean squared estimator (4.20), we carried out a further model based simulation study. For the purposes of this simulation study we focus on mean squared error estimation for the 25th, 50th and 75th percentiles using the bootstrap mean squared error estimator (4.20) and for the mean using either the analytic mean squared error (4.18) or the bootstrap mean squared error estimator (4.20). A total of 200 Monte-Carlo simulations were carried out for the percentile and 100 Monte-Carlo simulations for the mean, with bootstrap mean squared error estimation implemented by generating a single bootstrap population at each Monte Carlo simulation with $L = 500$ bootstrap samples taken from this population. The bootstrap population was generated unconditionally, with bootstrap population values then obtained by sampling from the smoothed residual distribution generated by the Monte Carlo sample data. We note that although it would have been theoretically preferable to have generated multiple bootstrap populations from each Monte Carlo sample, computing limitations restricted our investigation to $B = 1$. Also, since the estimates generated by the bootstrap procedure are then averaged over the 200 Monte Carlo simulations in our evaluation, this limitation is not as severe as it might appear to be at first, since the Monte Carlo simulations then proxy for the bootstrap populations. Simulation results for evaluating the mean squared error estimators are set out in Tables 4.3 and 4.4. Table 4.3 reports the 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 (4.10). It also includes the estimated mean squared errors based on (4.20) using the smoothed unconditional approach (Estimated Bootstrap) or (4.18) (Estimated Analytic). Intervals were defined as the M-quantile/CD estimator (4.10) plus or minus twice its estimated standard error, calculated as the square root of (4.18) or (4.20). Table 4.4 reports the 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 (4.14). The estimated mean squared error for quantiles is based on (4.20) using the smoothed unconditional approach. Focusing first on Table 4.3 we note that under both parametric scenarios, the analytic and the bootstrap mean squared error estimators track the true mean squared error of the small area mean estimators very well and provide coverage rates that are close to the nominal 95%. Focusing next on Table 4.4 we also note that the bootstrap mean squared error estimator offers a good approximation to the true mean squared error of the small area quantile estimators and also provides coverage rates that are close to the nominal 95%.

Table 4.1: Model-based simulation results for Scenario 1 (Gaussian area effects) averaged over 30 small areas.

Method	Target					
	10^{th}	25^{th}	50^{th}	Mean	75^{th}	90^{th}
	Relative Bias (%)					
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 MSE (%)					
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 4.2: Model-based simulation results for Scenario 2 (Chi-square area effects) averaged over 30 small areas.

Method	Target					
	10^{th}	25^{th}	50^{th}	Mean	75^{th}	90^{th}
	Relative Bias (%)					
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 MSE (%)					
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 4.3: Evaluation of MSE estimators (4.18) and (4.20).

MSE	Percentiles of across areas distribution					
	Min	25^{th}	50^{th}	Mean	75^{th}	Max
	Gaussian area effects					
True	0.27	0.331	0.411	0.419	0.481	0.783
Estimated Analytic MSE	0.289	0.317	0.400	0.416	0.500	0.680
Estimated Bootstrap MSE	0.282	0.319	0.401	0.418	0.504	0.715
Coverage Analytic MSE	0.88	0.93	0.95	0.94	0.97	0.99
Coverage Bootstrap MSE	0.88	0.94	0.96	0.96	0.97	0.99
	Chi square area effects					
True	0.344	0.453	0.549	0.589	0.736	1.051
Estimated Analytic	0.411	0.453	0.552	0.592	0.689	0.980
Estimated Bootstrap	0.398	0.444	0.559	0.589	0.706	1.003
Coverage Analytic MSE	0.87	0.89	0.92	0.93	0.96	0.98
Coverage Bootstrap MSE	0.92	0.95	0.96	0.96	0.97	1.00

Table 4.4: Evaluation of MSE estimator of 0.25, 0.50 and 0.75 quantiles using (4.20).

MSE		Percentiles of across areas distribution					
		Min	25^{th}	50^{th}	Mean	75^{th}	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 square 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	T rue	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

4.7 Applications to the Italian Survey on Income and Living Conditions

In Italy, the European Survey on Income and Living Conditions (EU-SILC) is conducted yearly by ISTAT to produce estimates on the living conditions of the population at national and regional (NUTS-2) levels.

Regions are planned domains for which EU-SILC estimates are published, while the Provinces are unplanned domains. These are the administrative areas (LAU-1 level) constituted by a different number of Municipalities (LAU-2 level) and whose boundaries do not cut across the Municipalities themselves. The regional samples are based on a stratified two stage sample design: in each Province the Municipalities are the Primary Sampling Units (PSUs), while the households are the Secondary Sampling Units (SSUs). The PSUs are divided into strata according to their dimension in terms of population size; the SSUs are selected by means of systematic sampling in each PSU. All the members of each sampled household are interviewed through an individual questionnaire, and one individual in each household (usually, the head of the household) is interviewed through an household questionnaire. It is useful to note that some Provinces - generally the smaller ones - may have very few sampled Municipalities; furthermore, many Municipalities are not even included in the sample at all. Direct estimates may therefore have large errors at provincial level or they may not even be computable at municipality level, thereby requiring resort to small area estimation techniques.

In this section we present two different applications using data sources from the 2006 and 2007 EU-SILC surveys. These datasets, with data coming from the 2001 Population Census of Italy, represent a complete and valuable source of information to produce poverty and living condition estimates in Italy.

It is useful to underline that the two case studies presented here are different in many aspects, and thus the results are not directly comparable, at least at this stage of the work. First of all, the survey data sources differ: in the first application, based on 2007 EU-SILC, we use income data referring to year 2006, in the second application to year 2005. Second, while in the first case study the small areas of interest are represented by the Provinces of three Italian Regions (Lombardia, Toscana and Campania, for a total of 29 areas), in the second case we focus only on the Provinces of the Toscana Region (with 11 areas). The more restrictive prospective of the second case study is motivated by the different small area model we specified to produce the poverty measures. Indeed, the preliminary study on the covariates to be used as predictors in the small area models suggested the selection of different covariates in the two studies; moreover, in the second application we found out a non linear relationship between one continuous covariate and the response variable, the equivalised household income. This suggested the specification of a nonparametric small area model, which is more computational intensive with respect to the models we specified in the first case study. Thus, the results of the second application we present in this section are only a starting point for further analyses, but they are useful for understanding the potential uses of nonparametric small area models.

4.7.1 First case study

In the first application our target is the estimation of the mean income, of the *Head Count Ratio* (HCR) or incidence of poverty and of some income quantiles for the Provinces of three Italian Regions: Lombardia (Northern Italy), Toscana (Central Italy) and Campania (Southern Italy). Data on the household equivalised income, on some household characteristics and on individual characteristics of the head of the household in the three Regions are available from the EU-SILC survey 2007. The same covariate information is available from the Census 2001 for all the households living in Lombardia, Toscana and Campania. The aim is not only to evaluate the distribution of the income inside the three Regions, but also to get a picture of the poverty and living conditions inequalities characterizing the Italian territory.

In this application we use an M-quantile CD estimator (4.11) with the following covariates: the marital status of the head of the family (six levels), the working position of the head of the household (four levels), the education of the head of the household (ten levels), the gender of the head of the household (male/female) and the mean house surface at area level (in square meters). The small areas are the 12 in Lombardia (11 Provinces plus the Municipality of Milano), 11 in Toscana (10 Provinces plus the Municipality of Firenze) and six in Campania (5 Provinces plus the Municipality of Napoli), for a total of 29 small areas. Log-transformation of household income has not yet been considered at this stage of the work, to avoid the possible bias and the complications of the back-transformation on the MSE estimation of the small area estimators (Chambers and Dorfman, 2003).

The estimates of the HCR and of the mean household equivalised income for Lombardia, Toscana and Campania are represented in Figures 4.1 and 4.2, 4.3 and 4.4, 4.5 and 4.6 respectively. In all these figures the darker color of an area always correspond to a worst situation of poverty, that is to an higher HCR or to a lower mean income.

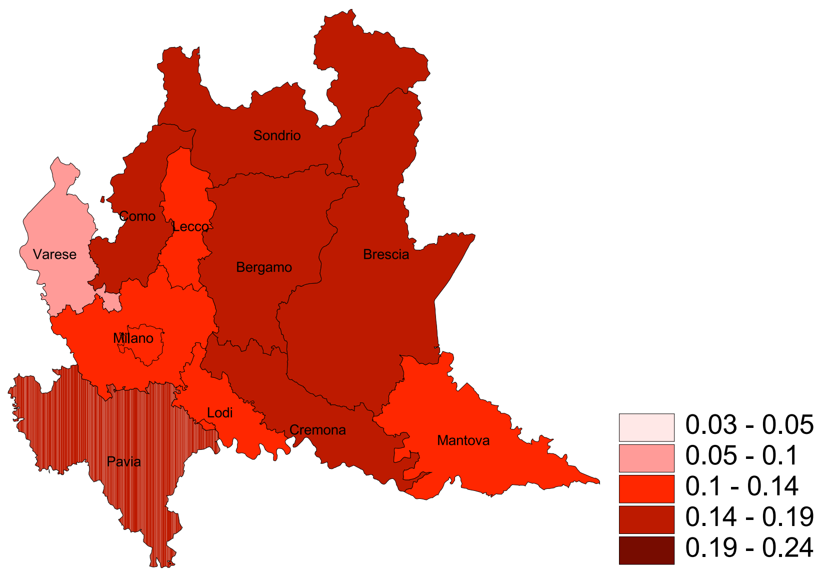

Figure 4.1: Estimate of the Head Count Ratio (% of individuals below the poverty line) - M-quantile CD Estimator, Lombardia Region.

The first evident results of our analyses is the higher incidence of poverty in the areas of Campania, a Region in Southern Italy: for this Region the estimates of the HCR, the percentage of households below the poverty line (9504 Euros, corresponding to the 60% of the median income), are in the range 26-44%, while for Lombardia (Northern Italy) and Toscana (Central Italy) the ranges of the HCR are 5-19% and 9-27% respectively. Also the estimated values for the mean income suggest a gap between these three Italian Regions.

Nevertheless, for both the estimates of interest we can notice a certain variability inside the three Regions. For example (Figure 4.1), several areas in Lombardia have an estimated

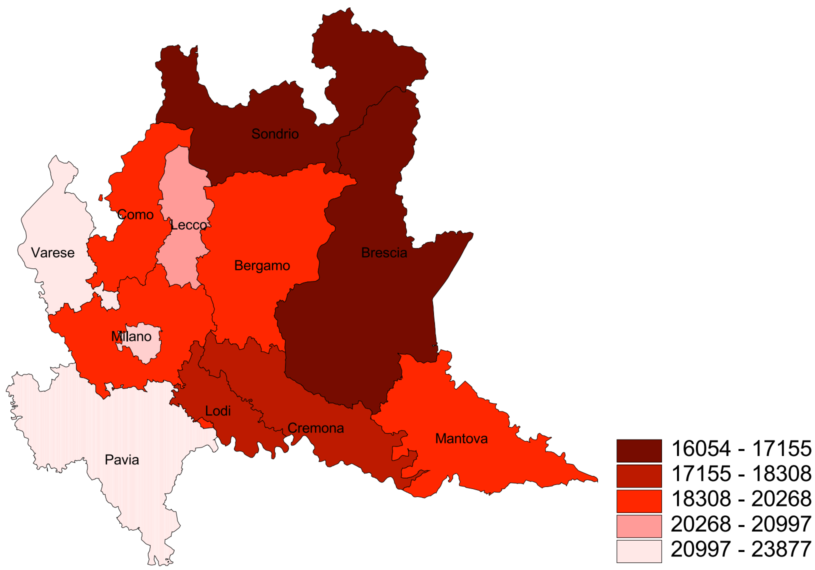

Figure 4.2: Estimate of the mean equivalised household income - M-quantile CD Estimator, Lombardia Region.

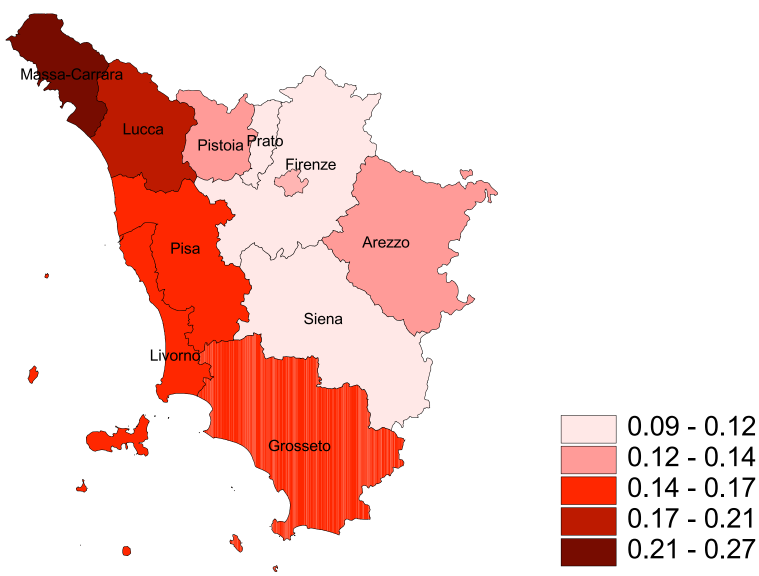

Figure 4.3: Estimate of the Head Count Ratio (% of individuals below the poverty line) - M-quantile CD Estimator, Toscana Region.

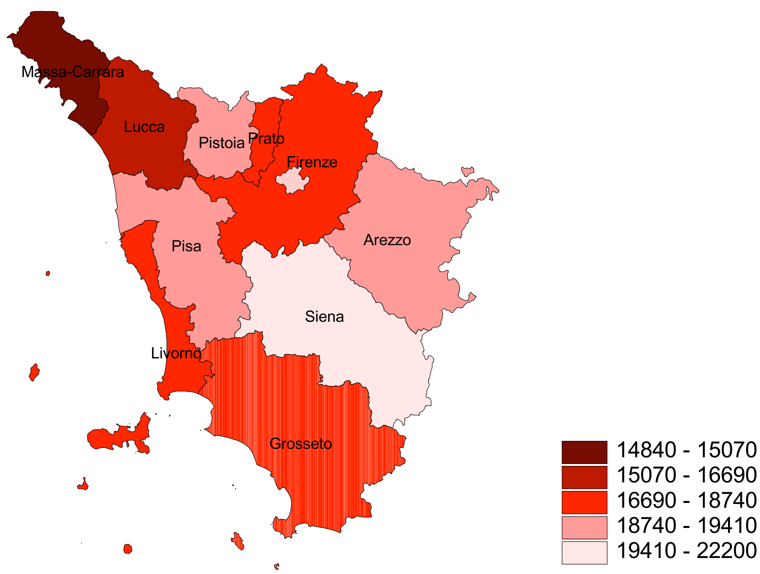

Figure 4.4: Estimate of the mean equivalised household income - M-quantile CD Estimator, Toscana Region.

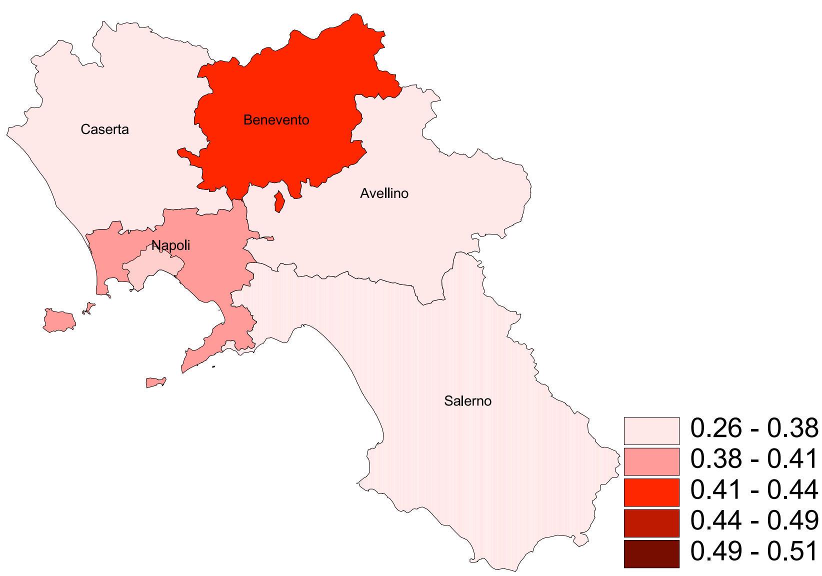

Figure 4.5: Estimate of the Head Count Ratio (% of individuals below the poverty line) - M-quantile CD Estimator, Campania Region.

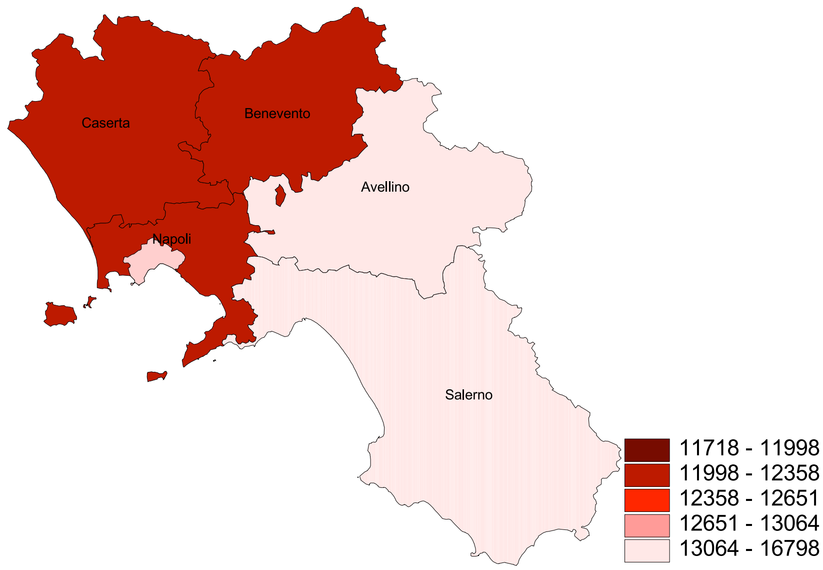

Figure 4.6: Estimate of the mean equivalised household income - M-quantile CD Estimator, Campania Region.

HCR in the class 0.14-0.19, and also the value for the Municipality of Milano is not in the lower class, which characterize the Province of Varese. In Campania we can notice a gap in the income estimates referring to the Northern areas (Caserta, Benevento and Napoli Provinces), with lower income estimates with respect to the Southern areas of the Region; however the same gap characterize only the HCR estimate referring to the Province of Benevento. For the Toscana Region we can notice that the higher incidence of poverty as well as the lower mean income estimate refer to the Province of Massa-Carrara, in the North of the Region. For this area the value of the HCR is comparable to the lower HCR values we observe in the Campania Region.

The Municipality of Milano, though characterized by a relative high HCR in the Lombardia Region, is in the class of higher estimated mean income, and it also has the median and the upper quantile estimates in Lombardia (see Tables 4.5 and 4.6). This suggest that for income values over the poverty line the cumulative distribution function of this area is above all the other estimated cumulative distribution functions in Lombardia. A similar behaviour characterized the other two big Municipalities, Firenze in Toscana and Napoli in Campania, though in these cases the estimated HCRs are always in the lower class of the corresponding Region. If we compare the cumulative distribution functions estimated for the three Municipalities (Figure 4.7 and Table 4.5) we can appreciate the gap between the Municipality of Napoli and the other two areas, and we can see that the quantiles of the Munipality of Milano are slightly higher than those of the Municipality of Firenze for income values above the poverty line. The direct estimates computed through the Horvitz-Thompson estimator (dashed lines in Figure 4.7) are not always consistent with the model based ones, especially in the centre of the income distributions.

Note that in the present application the estimates of interest (Table 4.5) are all accompanied by a measure of variability. In particular, the standard error of the mean income estimates was computed using formula (4.18), while the lower and upper limits of the HCR estimates were computed using a bootstrap estimator.

Table 4.5: Small areas estimates using the M-quantile CD model: population and small area sizes, Head Count Ratio, HCR lower and upper limits, mean household equivalent income with s.e.

Areas	N_d	n_d	HCR	HCR lower limit	HCR upper limit	Mean income	Mean s.e.
Massa-Carrara / Toscana	80810	96	0.26	0.21	0.32	14842.84	676.21
Lucca / Toscana	146117	131	0.21	0.16	0.26	16690.43	799.59
Pistoia / Toscana	104466	129	0.14	0.11	0.18	19255.50	1134.13
Firenze Province / Toscana	216531	326	0.11	0.09	0.13	18644.50	454.15
Livorno / Toscana	133729	115	0.15	0.11	0.20	18738.69	940.46
Pisa / Toscana	150259	136	0.15	0.10	0.19	19167.14	878.97
Arezzo / Toscana	123880	145	0.13	0.10	0.16	19414.74	1056.74
Siena / Toscana	101399	116	0.12	0.09	0.16	20928.79	1078.15
Grosseto / Toscana	87720	59	0.17	0.12	0.23	17874.16	1138.76
Prato / Toscana	83617	117	0.12	0.08	0.16	18097.09	691.09
Firenze Municipality / Toscana	159724	119	0.12	0.09	0.16	22203.33	1320.28
Varese / Lombardia	320899	253	0.10	0.08	0.13	21928.52	1297.09
Como / Lombardia	205963	153	0.16	0.13	0.22	19361.44	1130.66
Sondrio / Lombardia	69817	41	0.19	0.12	0.28	16894.17	1625.11
Milano Province / Lombardia	957305	543	0.12	0.10	0.14	20265.20	510.59
Bergamo / Lombardia	375778	219	0.17	0.13	0.20	19212.59	829.83
Brescia / Lombardia	437706	216	0.18	0.14	0.21	16921.99	572.86
Pavia / Lombardia	211786	60	0.15	0.10	0.23	22053.49	4053.48
Cremona / Lombardia	135321	75	0.16	0.10	0.21	17222.00	882.97
Mantova/ Lombardia	146249	234	0.13	0.10	0.16	18546.50	656.01
Lecco / Lombardia	121321	103	0.11	0.07	0.16	20281.30	1127.10
Lodi / Lombardia	77978	62	0.11	0.07	0.17	17986.44	995.64
Milano Municipality / Lombardia	588197	255	0.12	0.09	0.14	23876.47	1113.60
Caserta / Campania	279684	155	0.38	0.33	0.43	12056.16	611.14
Benevento / Campania	102441	70	0.43	0.35	0.52	12109.72	1003.61
Napoli Province / Campania	631523	596	0.40	0.38	0.43	12104.09	351.01
Avellino / Campania	152340	84	0.36	0.29	0.43	13609.70	990.96
Salerno/ Campania	359080	191	0.32	0.28	0.37	13534.64	522.29
Napoli Municipality / Campania	337787	221	0.26	0.22	0.30	16399.39	626.67

4.7.2 Second case study

In the second application our target is again the estimation of the mean household equivalised income, as well as the estimation of the quantiles of the income distribution and of the HCR for the Provinces of the Toscana Region. In this case however data come from a different year respect to the previous application, the EU-SILC survey 2006. The preliminary analysis suggested the

Table 4.6: Small areas estimates using the M-quantile CD model: quantiles of the household equivalised income.

Areas	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Massa-Carrara / Toscana	5927.56	9302.67	13897.29	19656.76	25352.02
Lucca / Toscana	6698.79	10341.86	15518.95	21849.49	28174.46
Pistoia / Toscana	8100.64	11771.77	16466.57	22676.13	35314.37
Firenze Province / Toscana	9081.97	12815.30	17265.50	23021.86	30129.86
Livorno / Toscana	7637.14	11932.00	17169.57	24035.98	31825.37
Pisa / Toscana	7715.23	12310.66	17651.00	25020.65	31989.20
Arezzo / Toscana	8526.24	12627.70	17321.35	22980.46	31214.38
Siena / Toscana	8616.74	13122.44	19234.35	26341.03	34390.34
Grosseto / Toscana	7369.55	11342.94	16950.41	24261.85	30086.36
Prato / Toscana	8735.63	12888.64	17351.54	22823.14	28960.56
Firenze Municipality / Toscana	8685.82	13104.39	19497.05	27816.45	40169.88
Varese / Lombardia	9386.46	13148.25	17646.15	24105.93	33694.71
Como / Lombardia	7823.42	11592.31	16501.40	22992.27	33327.84
Sondrio / Lombardia	7466.26	10656.90	15113.02	20461.93	26047.79
Milano Province / Lombardia	8731.20	13192.02	18269.21	24382.96	33250.05
Bergamo / Lombardia	7407.39	11490.67	16682.27	23879.87	35312.04
Brescia / Lombardia	7463.28	10973.61	15567.05	21559.37	28221.11
Pavia / Lombardia	7727.44	11662.38	16815.47	24158.36	30846.34
Cremona / Lombardia	8014.57	11612.81	15944.49	21565.67	29013.89
Mantova/ Lombardia	8703.20	12167.10	16950.97	22972.99	29916.88
Lecco / Lombardia	8926.71	13201.13	18592.04	25331.02	34951.54
Lodi / Lombardia	8851.35	12909.56	17385.32	22686.72	28789.72
Milano Municipality / Lombardia	8921.27	13658.16	20336.33	28928.33	43009.34
Caserta / Campania	3955.83	6966.30	11865.18	17680.72	24114.55
Benevento / Campania	4047.04	7013.11	10658.07	15280.13	21469.55
Napoli Province / Campania	3764.81	6709.65	11190.69	16569.87	22664.62
Avellino / Campania	4163.12	7469.51	12163.71	18480.82	24591.62
Salerno/ Campania	4650.65	8145.19	13018.57	18407.38	24263.83
Napoli Municipality / Campania	5360.07	9301.79	15256.73	22367.41	30363.27

selection of the following covariates: the household size (integer value), the ownership of the dwelling (owner/tenant), the age of the head of the household (integer value), the years of education of the head of the household (integer value) and the working position of the head of the household (employed/unemployed in the previous week). Furthemore, a more in-depth analysis of the relationship of the household equivalised income with the continuous selected covariates showed a non linearity of the relationship with the age of the head of the household (Figure 4.8). Thus, the small area model we specify to estimate the mean and the quantiles of the income is in this case a nonparametric CD M-quantile model of the form (4.11), with a nonparametric part for one covariate as in (4.9).

The small areas of interest are the 10 Provinces of Toscana, plus the Municipality of Firenze, considered as a stand alone area with 125 units out of 457 in the Province (see Table 4.7).

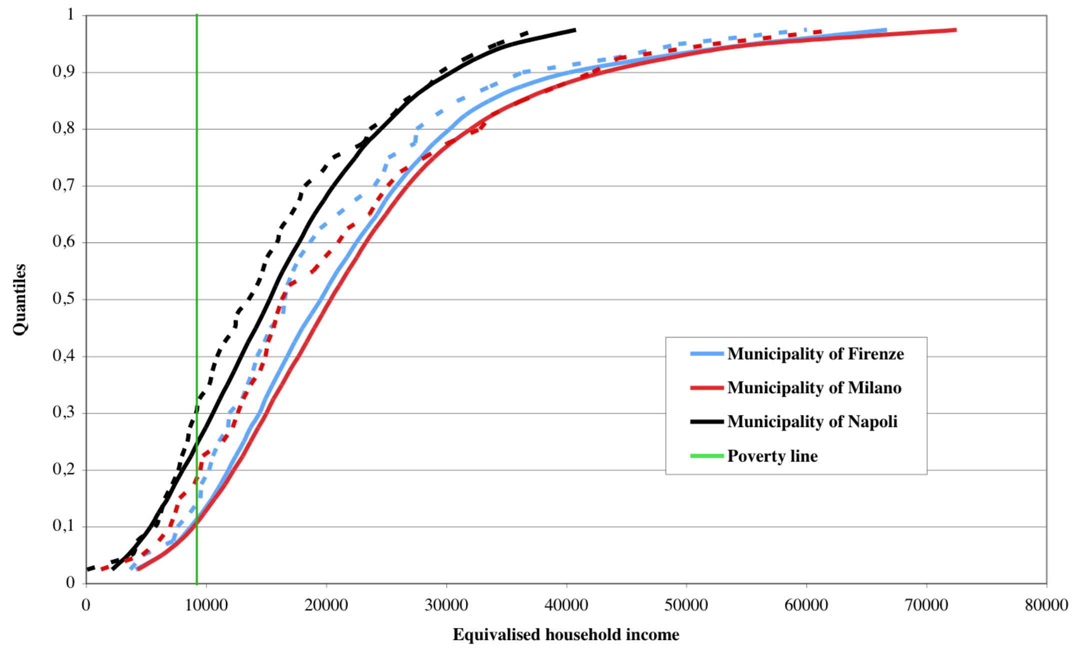

Figure 4.7: Estimated cumulative distribution functions of household equivalised income - M-quantile CD Estimator, Municipalities of Milano, Firenze and Napoli.

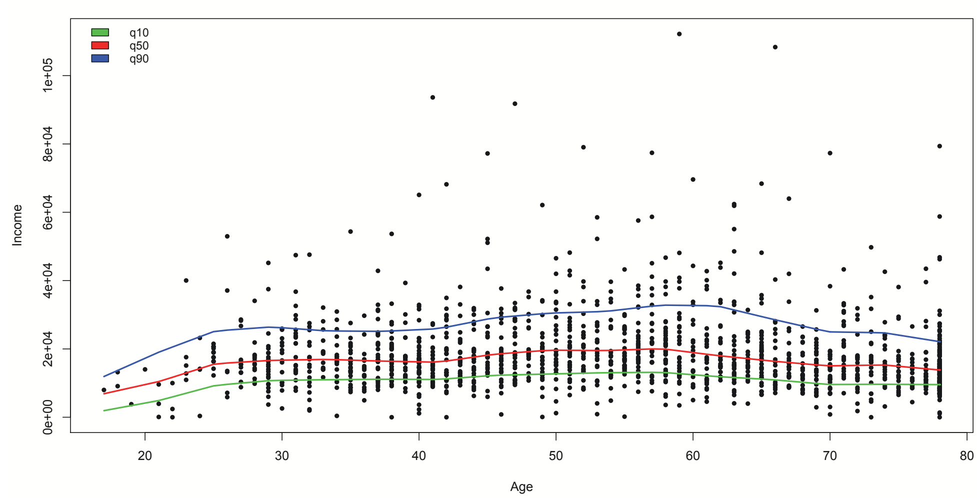

Figure 4.8: Estimation of the relationship between the household equivalised income and the age of the head of the household, for some quantiles of the income.

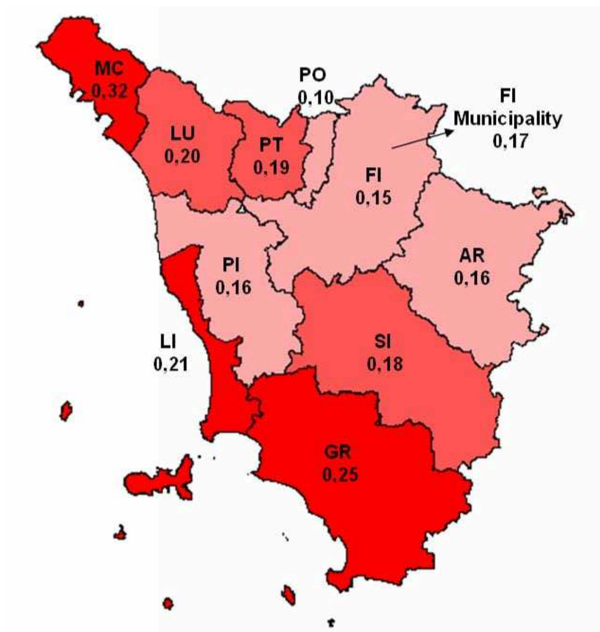

Figure 4.9: Estimate of the Head Count Ratio (% of individuals below the poverty line) - Nonparametric M-quantile CD Estimator, Toscana Region.

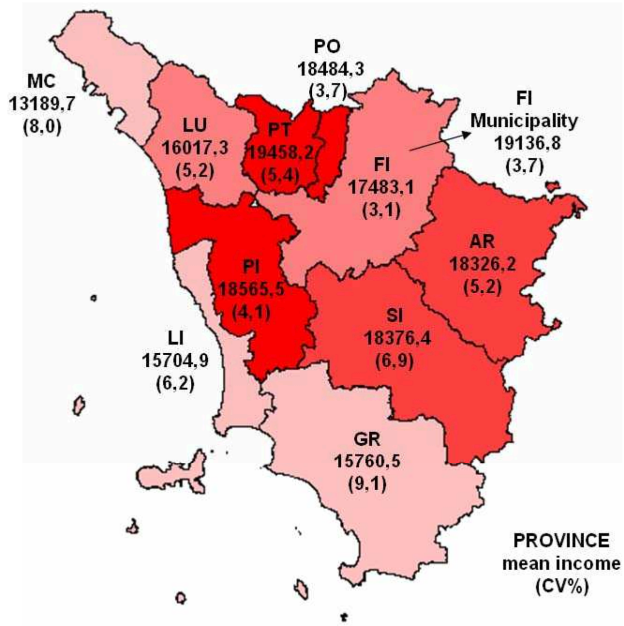

Figure 4.10: Estimate of the mean of the household equivalised income - Nonparametric M-quantile CD Estimator, Toscana Region.

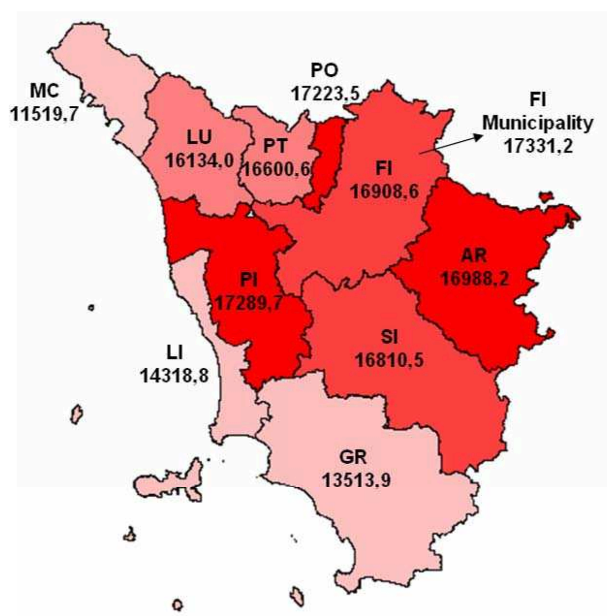

Figure 4.11: Estimate of the median of the household equivalised income - Nonparametric M-quantile CD Estimator, Toscana Region.

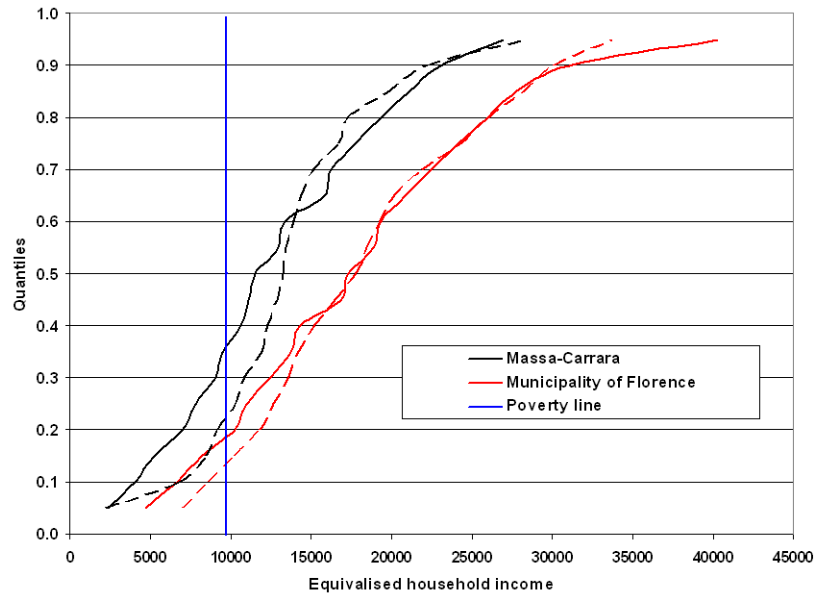

Figure 4.12: Estimated cumulative distribution functions of household equivalised income - Non-parametric M-quantile CD Estimator, Province of Massa-Carrara and Municipality of Florence.

Figure 4.9 report the results for the estimation of the HCR using a poverty line equal to 9667 Euros (60% of the median of household equivalised income): the darkest Provinces in Figure 4.9 are characterized by an higher HCR and thus by an higher incidence of poverty. Thus, we see that the Massa-Carrara Province (MC) has the highest percentage of poor individuals (more than 30%), while the Provinces of Firenze (FI), Arezzo (AR), Pisa (PI) and Prato (PO) have the lowest percentage of poor individuals (15-16%). The mean income estimates (Figure 4.10) indicate that the richest areas are the Provinces of Pistoia (PT), Prato (PO), Pisa (PI) and the Municipality of Firenze (FI); the poorest are the Provinces of Massa-Carrara (MC), Livorno (LI) and Grosseto (GR). Note that in this case the darkest is the color in Figure 4.10, the richest is the Province. Similar results emerge from the estimates of the median income on the small areas (Figure 4.11). These results are consistent with estimates referring to income in 2003, computed using data form the EU-SILC survey 2004 and a similar small area model (Giusti *et al.*, 2009a and 2009b).

We can can get a more complete picture looking at the estimated cumulative distribution function in the richest (Municipality of Firenze) and poorest (Province of Massa-Carrara) areas (Figure 4.12 and Table 4.8). In particular, the cumulative distribution function of Massa-Carrara rapidly approaches the value 1, and it is steeper than the cumulative distribution function of the Municipality of Firenze. The amount of people immediately below the poverty line (blue vertical line) is bigger under the black line than under the red one. The direct estimates (dashed lines) computed through the Horvitz-Thompson estimator are not always consistent with the model based ones, especially for the first quantiles. This is reasonably due to the small number of observations in these areas: 110 in the Province of Massa-Carrara, 125 in the Municipality of

Firenze.

Note that in this application a measure of variability has been computed only for the mean estimates, using (4.18). The estimation of the variability also for the estimated HCRs and income quantiles will be considered in the next steps of the analysis, using nonparametric bootstrap.

Table 4.7: Small areas estimates using the Nonparametric M-quantile CD model: population and small area sizes, Head Count Ratio, mean household equivalent income with s.e., median household equivalent income.

Area	N_d	n_d	HCR	Mean Income	Mean s.e.	Median Income
Massa-Carrara	80810	110	0.32	13189.69	1059.54	11519.7
Lucca	146117	109	0.20	16017.27	833.54	16134.0
Pistoia	104466	124	0.19	19458.15	1052.04	16600.6
Province of Firenze	216531	332	0.15	17483.12	536.02	16908.6
Livorno	133729	105	0.21	15704.88	977.93	14318.8
Pisa	150259	143	0.16	18565.54	767.93	17289.7
Arezzo	123880	159	0.16	18326.21	953.89	16988.2
Siena	101399	119	0.18	18376.39	1273.40	16810.5
Grosseto	87720	71	0.25	15760.54	1441.02	13513.9
Prato	83617	128	0.16	18484.34	682.11	17223.5
Municipality of Firenze	159724	125	0.17	19136.84	709.90	17331.2

The applications presented in this section suggest that small area methods play a crucial role in providing poverty measures at local level. In particular, M-quantile small area models avoid unduly restrictive constraints on the distribution of the error terms and of the area effects, and allows to handle outlying observations, a common feature when dealing with income data. Moreover, nonparametric M-quantile models avoid unduly restrictive constraints on the shape of the relationship between the response variable and the covariates.

These applications are only a starting point for further developments. Next applications will focus on obtaining the estimates for all the Municipalities in the three Italian Regions using M-quantile, nonparametric M-quantile and M-quantile GWR models, including also non-monetary measures of poverty (Cheli and Lemmi, 1995). Moreover, a measure of variability will be provided for all the estimates of interest, including those deriving from a nonparametric small area model.

Table 4.8: Small areas estimates using the Nonparametric M-quantile CD model: quantiles of the household equivalised income.

Quantile	Province of Massa-Carrara	Municipality of Firenze
0.05	2411.9	4731.4
0.10	4047.8	6664.4
0.15	5232.6	8203.0
0.20	6999.5	10216.1
0.25	7803.2	10970.4
0.30	9057.4	12442.7
0.35	9466.6	13751.6
0.40	10649.0	14320.1
0.45	11106.6	16805.6
0.50	11519.7	17331.2
0.55	12944.4	18847.3
0.60	13349.2	19344.4
0.65	15774.1	20859.4
0.70	16227.9	22448.4
0.75	17722.6	24169.3
0.80	19309.6	25950.0
0.85	21074.6	27732.7
0.90	23086.7	31053.8
0.95	26920.8	40282.2

Chapter 5

References

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.

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.

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.

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].

Cheli, B. and Lemmi, A. (1995). A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty. *Economic Notes*, 24, 115-134.

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.

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.

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.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2008). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, **78**, 443–462.

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.* [Paper submitted for publication. A copy is available upon request].

Hall, P. and Maiti, T. (2006). On Parametric Bootstrap Methods for Small Area Prediction. *Journal Royal Statistical Society, Series B*, **68**, 221–238.

Haslett, S. and Jones, G. (2005). Small area estimation using surveys and some practical and statistical issues. *Statistics in Transition*, **7**, 541–555.

Huber, P.J. (1981). *Robust Statistics.* John Wiley & Sons.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1-96.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Kokic, P., Chambers, R.L., Breckling, J., Beare, S. (1997). A measure of production performance. *Journal of Business and Economic Statistics*, **15**, 445-451.

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. and Rao, J.N.K. (2009) Small Area Estimation of Poverty Indicators. [Paper submitted for publication. A copy is available upon request].

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/personas/imolina/MiDocencia/SmallAreaEstimation/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.

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.

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.

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].

Särndal, C.E., Swensson, B. and Wretman J. (1992) *Model assisted survey sampling*, Springer-Verlag.

Tarozzi, A. and Deaton, A. (forthcoming). Using census and survey data to estimate poverty and inequality 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.

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.