Chapter 13
Random Regression Coefficient Models

13.1 Introduction

Coefficients of auxiliary variables in the standard nested error regression (NER) model are not allowed to vary across sampling units or domains. This assumption is too rigid in many practical situations. In some small area estimation (SAE) problems, we can intuitively expect that the slope parameters of some explanatory variable are not constant and therefore they should take different values in different domains. The random regression coefficient (RRC) model gives a practical solution to this problem by assuming that the beta parameters are random and therefore they give a more flexible way of modeling.

This section describes a modification of the nested error regression (NER) model having random regression coefficients. In the framework of SAE, Prasad and Rao (1990) derived empirical best linear unbiased predictors (EBLUP) of domain linear parameters under a unit-level RRC model. They also derived a second order approximation of the mean squared error (MSE) of the EBLUP and gave an estimator of the MSE approximation. They considered a special case of the RRC model proposed by Dempster et al. (1981), with a single concomitant variable x and a null intercept parameter (regression through origin).

Moura and Hold (1999) used a class of models allowing for variation between areas because of: (1) differences in the distribution of unit-level or area-level variables between areas, and (2) area-specific components of variance which cannot be explained by covariates. Their family of models contains RRC models as particular cases. These authors derived EBLUPs of linear parameters, gave an approximation to the MSE of the EBLUP, and proposed MSE estimators.

Hobza and Morales (2013) applied a flexible class of RRC models to the prediction of domain linear parameters. They gave a Fisher-scoring algorithm to calculate the residual maximum likelihood estimators of the model parameters and they derived EBLUPs and MSEs estimators. They applied the introduced

D. Morales et al., *A Course on Small Area Estimation and Mixed Models*, Statistics for Social and Behavioral Sciences, https://doi.org/10.1007/978-3-030-63757-6_13

methodology to the estimation of household normalized net annual incomes in the Spanish Living Conditions Survey.

This chapter extends the results of Hobza and Morales (2013) by considering a model where the set of random effects has a multivariate normal distribution that includes all variances and covariances as unknown parameters. It also studies the more simple model without covariances and gives some R codes for the last model.

13.2 The RRC Model with Covariance Parameters

We start with a description of the more general model.

13.2.1 The Model

Let us consider the model

$$y_{dj} = \sum_{k=1}^{p} \beta_k x_{kdj} + \sum_{k=1}^{p} u_{kd} x_{kdj} + e_{dj}, \quad d = 1, \ldots, D, \ j = 1, \ldots, n_d, \tag{13.1}$$

where

- $\boldsymbol{u}_d^* = (u_{1d}, \ldots, u_{pd})' \overset{iid}{\sim} N_p(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{u}_d^*})$ and $e_{dj} \overset{ind}{\sim} N(0, w_{dj}^{-1}\sigma_e^2)$ are independent, $d = 1, \ldots, D$, $j = 1 \ldots, n_d$,
- $\boldsymbol{V}_{\boldsymbol{u}_d^*}$ is a covariance matrix with components $\text{cov}(u_{k_1 d}, u_{k_2, d}) = \sigma_{k_1 k_2}$, $k_1, k_2 = 1, \ldots, p$.

In models with intercept, the first auxiliary variable is equal to one. The model (13.1) has p regression parameters and $1 + p + \frac{1}{2}p(p-1) = 1 + \frac{1}{2}p(p+1)$ variance component parameters. They are β_k, σ_e^2, $\sigma_{k_1 k_2}$, with $\sigma_{k_1 k_2} = \sigma_{k_2 k_1}$, $k, k_1, k_2 = 1, \ldots, p$.

An example of model (13.1) with intercept, $D = 2$, $n_1 = n_2 = 2$, $n = 4$, and $p = 2$ is

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \hline y_{21} \\ y_{22} \end{pmatrix} = \begin{pmatrix} 1 & x_{211} \\ 1 & x_{212} \\ \hline 1 & x_{221} \\ 1 & x_{222} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix} + \begin{pmatrix} x_{211} & 0 \\ x_{212} & 0 \\ \hline 0 & x_{221} \\ 0 & x_{222} \end{pmatrix} \begin{pmatrix} u_{21} \\ u_{22} \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ \hline e_{21} \\ e_{22} \end{pmatrix}.$$

In matrix notation model (13.1) is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{k=1}^{p} \boldsymbol{Z}_k \boldsymbol{u}_k + \boldsymbol{e}, \tag{13.2}$$

where $n = \sum_{d=1}^{D} n_d$, $\boldsymbol{\beta} = \boldsymbol{\beta}_{p\times 1}$, $\boldsymbol{y} = \underset{1\le d\le D}{\text{col}}(\boldsymbol{y}_d)$, $\boldsymbol{y}_d = \underset{1\le j\le n_d}{\text{col}}(y_{dj})$, $\boldsymbol{e} = \underset{1\le d\le D}{\text{col}}(\boldsymbol{e}_d)$, $\boldsymbol{e}_d = \underset{1\le j\le n_d}{\text{col}}(e_{dj})$, $\boldsymbol{u}_k = \underset{1\le d\le D}{\text{col}}(u_{kd})$, $\boldsymbol{X} = \underset{1\le d\le D}{\text{col}}(\boldsymbol{X}_d)$, $\boldsymbol{X}_d = \underset{1\le k\le p}{\text{col}'}(\boldsymbol{x}_{k,n_d}) = \underset{1\le j\le n_d}{\text{col}}(\boldsymbol{x}_{dj})$, $\boldsymbol{x}_{k,n_d} = \underset{1\le j\le n_d}{\text{col}}(x_{kdj})$, $\boldsymbol{x}_{dj} = \underset{1\le k\le p}{\text{col}'}(x_{kdj})$, $\boldsymbol{Z}_k = \underset{1\le d\le D}{\text{diag}}(\boldsymbol{x}_{k,n_d})$, $\boldsymbol{I}_a = \underset{1\le j\le a}{\text{diag}}(1)$, $\boldsymbol{W} = \underset{1\le d\le D}{\text{diag}}(\boldsymbol{W}_d)$, $\boldsymbol{W}_d = \underset{1\le j\le n_d}{\text{diag}}(w_{dj})$, with $w_{dj} > 0$ known, $d = 1,\dots,D$, $j = 1,\dots,n_d$. Covariance matrices are $\boldsymbol{V}_e = \text{var}(\boldsymbol{e}) = \sigma_e^2 \boldsymbol{W}^{-1}$, $\boldsymbol{V}_{k_1k_2} = \text{cov}(\boldsymbol{u}_{k_1}, \boldsymbol{u}_{k_2}) = \sigma_{k_1k_2}\boldsymbol{I}_D$, $k_1, k_2 = 1,\dots,p$, and

$$\boldsymbol{V} = \text{var}(\boldsymbol{y}) = \boldsymbol{V}_e + \sum_{k_1=1}^{p}\sum_{k_2=1}^{p} \boldsymbol{Z}_{k_1}\boldsymbol{V}_{k_1k_2}\boldsymbol{Z}'_{k_2} = \underset{1\le d\le D}{\text{diag}}(\boldsymbol{V}_d),$$

where

$$\boldsymbol{V}_d = \sigma_e^2 \boldsymbol{W}_d^{-1} + \sum_{k_1=1}^{p}\sum_{k_2=1}^{p} \sigma_{k_1k_2}\boldsymbol{x}_{k_1,n_d}\boldsymbol{x}'_{k_2,n_d}, \quad d = 1,\dots,D.$$

Let $\boldsymbol{u} = \underset{1\le k\le p}{\text{col}}(\boldsymbol{u}_k)$ and $\boldsymbol{Z} = \underset{1\le k\le p}{\text{col}'}(\boldsymbol{Z}_k)$. Under this notation, the variance of $\boldsymbol{u}$ is

$$\boldsymbol{V}_u = \text{var}(\boldsymbol{u}) = \big(\boldsymbol{V}_{k_1k_2}\big)_{k_1k_2=1,\dots,p}$$

and the model (13.2) can be written in the general form

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

If the variance components are known, then the BLUE of $\boldsymbol{\beta} = (\beta_1,\dots,\beta_p)'$ is (cf. (6.12))

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} = \Big(\sum_{d=1}^{D}\boldsymbol{X}'_d\boldsymbol{V}_d^{-1}\boldsymbol{X}_d\Big)^{-1}\Big(\sum_{d=1}^{D}\boldsymbol{X}'_d\boldsymbol{V}_d^{-1}\boldsymbol{y}_d\Big)$$

and the BLUP of $\boldsymbol{u}$ is $\tilde{\boldsymbol{u}} = \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})$.

13.2.2 REML Estimators

In order to derive formulas for calculating the REML estimates of the unknown variance parameters we consider the alternative parametrization $\sigma^2 = \sigma_e^2$, $\varphi_{k_1k_2} = \sigma_{k_1k_2}/\sigma_e^2$, $k_1, k_2 = 1\dots,p$, and we define $\boldsymbol{\sigma} = (\sigma^2, \varphi_{k_1k_2}, k_1, k_2 = 1\dots,p)$ and $\boldsymbol{\Sigma} = \underset{1\le d\le D}{\text{diag}}(\boldsymbol{\Sigma}_d)$ where $\boldsymbol{\Sigma}_d = \sigma^{-2}\boldsymbol{V}_d$.

The REML log-likelihood is (cf. (6.32))

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

where

$$\boldsymbol{P} = \boldsymbol{K}(\boldsymbol{K}'\boldsymbol{\Sigma}\boldsymbol{K})^{-1}\boldsymbol{K}' = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1},$$
$$\boldsymbol{K} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}$$

are such that $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{0}$ and $\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P} = \boldsymbol{P}$. The matrix $\boldsymbol{\Sigma}$ can be written in the form

$$\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\varphi_{k_1k_2}\boldsymbol{A}_{k_1k_2},$$

where $\boldsymbol{A}_{k_1k_2} = \boldsymbol{Z}_{k_1}\boldsymbol{Z}'_{k_2} = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{x}_{k_1,n_d}\boldsymbol{x}'_{k_2,n_d})$, $k_1, k_2 = 1, \ldots, p$. In the same way as in (6.43) we obtain $\frac{\partial \boldsymbol{P}}{\partial \varphi_{k_1k_2}} = -\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\boldsymbol{P}$. Thus, by taking partial derivatives of the REML log-likelihood with respect to σ^2 and $\varphi_{k_1k_2}$, $k_1, k_2 = 1, \ldots, p$, one gets the scores

$$S_{\sigma^2} = -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4}\,\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}, \quad S_{\varphi_{k_1k_2}} = -\frac{1}{2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\} + \frac{1}{2\sigma^2}\,\boldsymbol{y}'\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\boldsymbol{P}\boldsymbol{y},$$

and the second partial derivatives

$$H_{\sigma^2\sigma^2} = \frac{n-p}{2\sigma^4} - \frac{1}{\sigma^6}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}, \quad H_{\sigma^2\varphi_{k_1k_2}} = -\frac{1}{2\sigma^4}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\boldsymbol{P}\boldsymbol{y},$$
$$H_{\varphi_{k_1k_2}\varphi_{i_1i_2}} = \frac{1}{2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\boldsymbol{P}\boldsymbol{A}_{i_1i_2}\} - \frac{1}{2\sigma^2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\boldsymbol{P}\boldsymbol{A}_{i_1i_2}\boldsymbol{P}\boldsymbol{y}$$
$$- \frac{1}{2\sigma^2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{A}_{i_1i_2}\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\boldsymbol{P}\boldsymbol{y},$$

where $k_1, k_2, i_1, i_2 = 1, \ldots, p$. By taking expectations and multiplying by -1, we obtain the components of the Fisher information matrix. For $k_1, k_2, i_1, i_2 = 1, \ldots, p$, we have

$$F_{\sigma^2\sigma^2} = \frac{n-p}{2\sigma^4},\ F_{\sigma^2\varphi_{k_1k_2}} = \frac{1}{2\sigma^2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\},\ F_{\varphi_{k_1k_2}\varphi_{i_1i_2}} = \frac{1}{2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_{k_1k_2}\boldsymbol{P}\boldsymbol{A}_{i_1i_2}\}.$$

To calculate the REML estimates, the Fisher-scoring updating formula, at iteration i, is

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

where $\boldsymbol{S}(\boldsymbol{\sigma})$ and $\boldsymbol{F}(\boldsymbol{\sigma})$ are the vector of scores and the Fisher information matrix evaluated at $\boldsymbol{\sigma}$. The following seeds can be used as starting values in the Fisher-scoring algorithm:

$$\sigma^{2(0)} = S^2/(p+2), \quad \varphi_{k_1k_2} = \delta_{k_1k_2}, \quad k_1, k_2 = 1, \ldots, p,$$

where $\delta_{k_1,k_2} = 1$ if $k_1 = k_2$, $\delta_{k_1,k_2} = 0$ if $k_1 \neq k_2$, $S^2 = \frac{1}{n-p}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_0)'\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_0)$ and $\hat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}$.

13.2.3 EBLUP of the Domain Mean

Let us now consider a finite population U partitioned into D domains U_d, i.e. $U = \cup_{d=1}^D U_d$. Let N and N_d be the sizes of U and U_d, so that $N = \sum_{d=1}^D N_d$. We assume that the population target vector $\boldsymbol{y} = \boldsymbol{y}_{N\times 1}$ follows the RRC model (13.2) with the obvious size changes, i.e. with N and N_d in the place of n and n_d, respectively.

Let $s \subset U$ be a sample of $n \leq N$ units and let $r = U - s$ be the set of non-sampled units. The domain and subdomain subsets of s and r are denoted by s_d and r_d, respectively. The subindexes s and r in vectors or matrices are used to denote their sampled and the non-sampled parts. Without loss of generality, we renumber the population units and we write

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_s \\ \boldsymbol{y}_r \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_s \\ \boldsymbol{X}_r \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_s \\ \boldsymbol{e}_r \end{pmatrix}, \quad \boldsymbol{Z}_k = \begin{pmatrix} \boldsymbol{Z}_{sk} \\ \boldsymbol{Z}_{rk} \end{pmatrix}, \ k = 1, \ldots, p,$$

and

$$\boldsymbol{V} = \text{var}(\boldsymbol{y}) = \begin{pmatrix} \boldsymbol{V}_s & \boldsymbol{V}_{sr} \\ \boldsymbol{V}_{rs} & \boldsymbol{V}_r \end{pmatrix}.$$

Using Theorem 4.1, the EBLUP of the linear parameter $\eta = \boldsymbol{a}'\boldsymbol{y} = \boldsymbol{a}_s'\boldsymbol{y}_s + \boldsymbol{a}_r'\boldsymbol{y}_r$ is

$$\hat{\eta} = \boldsymbol{a}_s'\boldsymbol{y}_s + \boldsymbol{a}_r'\left[\boldsymbol{X}_r\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{V}}_{rs}\hat{\boldsymbol{V}}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}})\right],$$

where

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

As $\boldsymbol{V}_{ers} = \boldsymbol{0}$, $\boldsymbol{V}_{rs} = \boldsymbol{Z}_r \boldsymbol{V}_u \boldsymbol{Z}_s' + \boldsymbol{V}_{ers} = \boldsymbol{Z}_r \boldsymbol{V}_u \boldsymbol{Z}_s'$ and $\hat{\boldsymbol{u}} = \hat{\boldsymbol{V}}_u \boldsymbol{Z}_s' \hat{\boldsymbol{V}}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s \hat{\boldsymbol{\beta}})$, then

$$\hat{\eta} = \boldsymbol{a}_s' \boldsymbol{y}_s + \boldsymbol{a}_r' \left[\boldsymbol{X}_r \hat{\boldsymbol{\beta}} + \boldsymbol{Z}_r \hat{\boldsymbol{V}}_u \boldsymbol{Z}_s' \hat{\boldsymbol{V}}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s \hat{\boldsymbol{\beta}})\right] = \boldsymbol{a}_s' \boldsymbol{y}_s + \boldsymbol{a}_r' \left[\boldsymbol{X}_r \hat{\boldsymbol{\beta}} + \boldsymbol{Z}_r \hat{\boldsymbol{u}}\right]$$
$$= \boldsymbol{a}' \left[\boldsymbol{X} \hat{\boldsymbol{\beta}} + \sum_{k=1}^{p} \boldsymbol{Z}_k \hat{\boldsymbol{u}}_k\right] + \boldsymbol{a}_s' \left[\boldsymbol{y}_s - \boldsymbol{X}_s \hat{\boldsymbol{\beta}} - \sum_{k=1}^{p} \boldsymbol{Z}_{sk} \hat{\boldsymbol{u}}_k\right].$$

The domain mean is $\overline{Y}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj} = \mu_d + \overline{e}_d$, where $\overline{e}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} e_{dj}$ and

$$\mu_d = \sum_{k=1}^{p} \overline{X}_{kd} \beta_k + \sum_{k=1}^{p} \overline{X}_{kd} u_{kd}, \qquad \overline{X}_{kd} = \frac{1}{N_d} \sum_{j=1}^{N_d} x_{kdj}.$$

The domain mean $\overline{Y}_d$ can be written in the form $\eta = \boldsymbol{a}' \boldsymbol{y}$, where

$$\boldsymbol{a}' = \frac{1}{N_d} (\boldsymbol{0}_{N_1}', \ldots, \boldsymbol{0}_{N_{d-1}}', \boldsymbol{1}_{N_d}', \boldsymbol{0}_{N_{d+1}}', \ldots, \boldsymbol{0}_{N_D}') = \frac{1}{N_d} \underset{1 \le \ell \le D}{\text{col}'} \{\delta_{d\ell} \boldsymbol{1}_{N_\ell}'\},$$

$\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ if $a \neq b$. It holds that $\boldsymbol{a}' \boldsymbol{X} = \overline{\boldsymbol{X}}_d = (\overline{X}_{1d}, \ldots, \overline{X}_{pd})$,

$$\boldsymbol{a}' \boldsymbol{Z}_k \hat{\boldsymbol{u}}_k = \frac{1}{N_d} \underset{1 \le \ell \le D}{\text{col}'} \{\delta_{d\ell} \boldsymbol{1}_{N_\ell}'\} \underset{1 \le \ell \le D}{\text{diag}} (\boldsymbol{x}_{k,N_\ell}) \hat{\boldsymbol{u}}_k = \underset{1 \le \ell \le D}{\text{col}'} \{\delta_{d\ell} \overline{X}_{k\ell}\} \hat{\boldsymbol{u}}_k = \overline{X}_{kd} \hat{u}_{kd}.$$

If $n_d > 0$, then the EBLUP of $\overline{Y}_d$ is

$$\widehat{\overline{Y}}_d^{eblup} = \sum_{k=1}^{p} \overline{X}_{kd} \hat{\beta}_k + \sum_{k=1}^{p} \overline{X}_{kd} \hat{u}_{kd} + f_d \left[\overline{y}_{s,d} - \sum_{k=1}^{p} \overline{X}_{s,kd} \hat{\beta}_k - \sum_{k=1}^{p} \overline{X}_{s,kd} \hat{u}_{kd}\right],$$

where $\overline{y}_{s,d} = \frac{1}{n_d} \sum_{j=1}^{n_d} y_{dj}$, $\overline{X}_{s,kd} = \frac{1}{n_d} \sum_{j=1}^{n_d} x_{kdj}$ and $f_d = \frac{n_d}{N_d}$. If $f_d \approx 0$, then the EBLUP of $\overline{Y}_d$ is approximately equal to

$$\hat{\mu}_d^{eblup} = \sum_{k=1}^{p} \overline{X}_{kd} \hat{\beta}_k + \sum_{k=1}^{p} \overline{X}_{kd} \hat{u}_{kd}.$$

The MSE of the EBLUP can be estimated by adapting the steps 1–6 of the parametric bootstrap procedure described in Sect. 8.5.

13.3 The RRC Model Without Covariance Parameters

For the ease of exposition, we consider now a slightly simpler model under which we derive more detailed formulas for the REML estimators and the formulas for the analytic approximation of the MSE of EBLUPs.

13.3.1 The Model

Let us consider the RRC model

$$y_{dj} = \sum_{k=1}^{p} \beta_k x_{kdj} + \sum_{k=1}^{p} u_{kd} x_{kdj} + e_{dj}, \quad d = 1, \ldots, D, \ j = 1, \ldots, n_d, \tag{13.3}$$

where $u_{kd} \overset{iid}{\sim} N(0, \sigma_k^2)$ and $e_{dj} \overset{iid}{\sim} N(0, w_{dj}^{-1}\sigma_e^2)$ are independent, $d = 1, \ldots, D$, $j = 1 \ldots, n_d$, $k = 1, \ldots, p$. The model variance and covariance parameters are σ_e^2, σ_k^2, $k = 1, \ldots, p$, ($p + 1$ parameters). In matrix notation model (13.3) is

$$\boldsymbol{y} = \boldsymbol{X\beta} + \sum_{k=1}^{p} \boldsymbol{Z}_k \boldsymbol{u}_k + \boldsymbol{e}, \tag{13.4}$$

where the meaning of the used symbols is exactly the same as the one given below formula (13.2). The difference with respect to the model (13.2) is only in the variance matrices which are now simpler and have the form $\boldsymbol{V}_e = \text{var}(\boldsymbol{e}) = \sigma_e^2 \boldsymbol{W}^{-1}$, $\boldsymbol{V}_{\boldsymbol{u}_k} = \text{var}(\boldsymbol{u}_k) = \sigma_k^2 \boldsymbol{I}_D$, $k = 1, \ldots, p$, and

$$\boldsymbol{V} = \text{var}(\boldsymbol{y}) = \boldsymbol{V}_e + \sum_{k=1}^{p} \boldsymbol{Z}_k \boldsymbol{V}_{\boldsymbol{u}_k} \boldsymbol{Z}_k' = \underset{1 \le d \le D}{\text{diag}} (\boldsymbol{V}_d),$$

where

$$\boldsymbol{V}_d = \sigma_e^2 \boldsymbol{W}_d^{-1} + \sum_{k=1}^{p} \sigma_k^2 \boldsymbol{x}_{k,n_d} \boldsymbol{x}_{k,n_d}', \ \ d = 1, \ldots, D.$$

We consider the alternative parameters $\sigma^2 = \sigma_e^2$, $\varphi_k = \sigma_k^2/\sigma_e^2$, $k = 1, \ldots, p$, in such a way that $\boldsymbol{V} = \sigma^2 \boldsymbol{\Sigma}$ and $\boldsymbol{V}_d = \sigma^2 \boldsymbol{\Sigma}_d$, where

$$\boldsymbol{\Sigma}_d = \boldsymbol{W}_d^{-1} + \sum_{k=1}^{p} \varphi_k \boldsymbol{x}_{k,n_d} \boldsymbol{x}_{k,n_d}', \ \ d = 1, \ldots, D.$$

Let $\boldsymbol{\theta} = (\sigma^2, \varphi_1, \ldots, \varphi_p)'$ be the vector of variance components, with $\sigma^2 > 0$, $\varphi_1 > 0, \ldots, \varphi_p > 0$. Let $\boldsymbol{u} = \underset{1\le k\le p}{\text{col}}(\boldsymbol{u}_k)$ and $\boldsymbol{Z} = \underset{1\le k\le p}{\text{col}'}(\boldsymbol{Z}_k)$. The variance of $\boldsymbol{u}$ is

$$\boldsymbol{V_u} = \text{var}(\boldsymbol{u}) = \underset{1\le k\le p}{\text{diag}}(\boldsymbol{V}_{\boldsymbol{u}_k}).$$

Using this notation, the model (13.4) can be written in the general form

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

If $\boldsymbol{\theta}$ is known, then the BLUE of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is (cf. (6.12))

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} = \left(\sum_{d=1}^{D}\boldsymbol{X}_d'\boldsymbol{\Sigma}_d^{-1}\boldsymbol{X}_d\right)^{-1}\left(\sum_{d=1}^{D}\boldsymbol{X}_d'\boldsymbol{\Sigma}_d^{-1}\boldsymbol{y}_d\right)$$

and the BLUP of $\boldsymbol{u}$ is $\tilde{\boldsymbol{u}} = \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})$, i.e.

$$\tilde{\boldsymbol{u}} = \underset{1\le k\le p}{\text{diag}}(\boldsymbol{V}_{\boldsymbol{u}_k})\underset{1\le k\le p}{\text{col}}(\boldsymbol{Z}_k')\underset{1\le d\le D}{\text{diag}}(\boldsymbol{V}_d^{-1})\underset{1\le d\le D}{\text{col}}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\boldsymbol{\beta}})$$

$$= \begin{pmatrix} \varphi_1 \underset{1\le d\le D}{\text{diag}}(\boldsymbol{x}_{1,n_d}')\underset{1\le d\le D}{\text{col}}\left(\boldsymbol{\Sigma}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\boldsymbol{\beta}})\right) \\ \varphi_2 \underset{1\le d\le D}{\text{diag}}(\boldsymbol{x}_{2,n_d}')\underset{1\le d\le D}{\text{col}}\left(\boldsymbol{\Sigma}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\boldsymbol{\beta}})\right) \\ \vdots \\ \varphi_p \underset{1\le d\le D}{\text{diag}}(\boldsymbol{x}_{p,n_d}')\underset{1\le d\le D}{\text{col}}\left(\boldsymbol{\Sigma}_d^{-1}(\boldsymbol{y}_d - \boldsymbol{X}_d\tilde{\boldsymbol{\beta}})\right) \end{pmatrix}.$$

13.3.2 REML Estimators

In this section we follow the same steps as in the Sect. 13.2.2 and we derive the REML estimators under the model (13.4) with more details concerning the matrix calculations. The REML log-likelihood is

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

where $\boldsymbol{P} = \boldsymbol{K}(\boldsymbol{K}'\boldsymbol{\Sigma}\boldsymbol{K})^{-1}\boldsymbol{K}' = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{K} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}$ are such that $\boldsymbol{PX} = \boldsymbol{0}$ and $\boldsymbol{P\Sigma P} = \boldsymbol{P}$. The matrix $\boldsymbol{\Sigma}$ can be written in the form

$$\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \sum_{k=1}^{p} \varphi_k \boldsymbol{A}_k,$$

where $\boldsymbol{A}_k = \boldsymbol{Z}_k \boldsymbol{Z}_k' = \underset{1 \le d \le D}{\text{diag}} (\boldsymbol{x}_{k,n_d} \boldsymbol{x}_{k,n_d}')$, $k = 1, \ldots, p$. As $\frac{\partial \boldsymbol{P}}{\partial \varphi_k} = -\boldsymbol{P}\boldsymbol{A}_k \boldsymbol{P}$, by taking partial derivatives with respect to σ^2 and φ_k, $k = 1, \ldots, p$, one gets

$$S_{\sigma^2} = -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4}\, \boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}, \quad S_{\varphi_k} = -\frac{1}{2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_k\} + \frac{1}{2\sigma^2}\, \boldsymbol{y}'\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{P}\boldsymbol{y}, \quad k = 1, \ldots, p.$$

The second partial derivatives are

$$H_{\sigma^2\sigma^2} = \frac{n-p}{2\sigma^4} - \frac{1}{\sigma^6}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}, \quad H_{\sigma^2\varphi_k} = -\frac{1}{2\sigma^4}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{P}\boldsymbol{y},$$
$$H_{\varphi_k\varphi_i} = \frac{1}{2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{P}\boldsymbol{A}_i\} - \frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{P}\boldsymbol{A}_i\boldsymbol{P}\boldsymbol{y}, \quad k, i = 1, \ldots, p.$$

By taking expectations and multiplying by -1, we obtain the components of the Fisher information matrix

$$F_{\sigma^2\sigma^2} = \frac{n-p}{2\sigma^4},\ F_{\sigma^2\varphi_k} = \frac{1}{2\sigma^2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_k\},\ F_{\varphi_k\varphi_i} = \frac{1}{2}\,\text{tr}\{\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{P}\boldsymbol{A}_i\},\ k, i = 1, \ldots, p.$$

To calculate the REML estimates, the Fisher-scoring updating formula, at iteration i, is

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

The following seeds can be used as starting values in the Fisher-scoring algorithm:

$$\sigma^{2(0)} = \varphi_1^{(0)} = \ldots = \varphi_p^{(0)} = S^2/(p+2),$$

where $S^2 = \frac{1}{n-p}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0)'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0)$ and $\hat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}$.

13.3.2.1 Matrix Calculations for the RRC Model

In this section we show how to do the matrix calculations in the Fisher-scoring algorithm. We define

$$\boldsymbol{\Sigma} = \underset{1 \le d \le D}{\text{diag}} (\boldsymbol{\Sigma}_d),\ \boldsymbol{R} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} = \left(\sum_{d=1}^{D} \boldsymbol{X}_d'\boldsymbol{\Sigma}_d^{-1}\boldsymbol{X}_d\right)^{-1}$$

such that

$$P = \Sigma^{-1} - \Sigma^{-1} X R X' \Sigma^{-1} = \underset{1 \le d \le D}{\text{diag}} (\Sigma_d^{-1}) - \underset{1 \le d \le D}{\text{col}} (\Sigma_d^{-1} X_d) R \underset{1 \le d \le D}{\text{col}'} (X_d' \Sigma_d^{-1}).$$

For $k = 1, \ldots, p$, the components of the vector of scores are

$$S_{\sigma^2} = -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{d=1}^{D} y_d^t \Sigma_d^{-1} y_d - \frac{1}{2\sigma^4} \left(\sum_{d=1}^{D} y_d^t \Sigma_d^{-1} X_d \right) R \left(\sum_{d=1}^{D} X_d^t \Sigma_d^{-1} y_d \right),$$

$$\begin{aligned}
S_{\varphi_k} &= -\frac{1}{2} \operatorname{tr}\{Z_k' P Z_k\} + \frac{1}{2\sigma^2} y' P Z_k Z_k' P y \\
&= -\frac{1}{2} \sum_{d=1}^{D} x_{k,n_d}' [\Sigma_d^{-1} - \Sigma_d^{-1} X_d R X_d' \Sigma_d^{-1}] x_{k,n_d} \\
&+ \frac{1}{2\sigma^2} \sum_{d=1}^{D} y_d' \Sigma_d^{-1} x_{k,n_d} x_{k,n_d}' \Sigma_d^{-1} y_d \\
&- \frac{1}{\sigma^2} \left(\sum_{d=1}^{D} y_d' \Sigma_d^{-1} x_{k,n_d} x_{k,n_d}' \Sigma_d^{-1} X_d \right) R \left(\sum_{d=1}^{D} X_d' \Sigma_d^{-1} y_d \right) \\
&+ \frac{1}{2\sigma^2} \left(\sum_{d=1}^{D} y_d' \Sigma_d^{-1} X_d \right) R \left(\sum_{d=1}^{D} X_d' \Sigma_d^{-1} x_{k,n_d} x_{k,n_d}' \Sigma_d^{-1} X_d \right) R \left(\sum_{d=1}^{D} X_d' \Sigma_d^{-1} y_d \right).
\end{aligned}$$

For $k, k_1, k_2 = 1, \ldots, p$, the components of the REML Fisher information matrix are

$$\begin{aligned}
F_{\sigma^2\sigma^2} &= \frac{n-p}{2\sigma^4}, \\
F_{\sigma^2\varphi_k} &= \frac{1}{2\sigma^2} \operatorname{tr}\{Z_k' P Z_k\} = \frac{1}{2\sigma^2} \sum_{d=1}^{D} x_{k,n_d}' \left[\Sigma_d^{-1} - \Sigma_d^{-1} X_d R X_d' \Sigma_d^{-1} \right] x_{k,n_d},
\end{aligned}$$

$$\begin{aligned}
F_{\varphi_{k_1}\varphi_{k_2}} &= \frac{1}{2} \operatorname{tr}\{Z_{k_2}' P Z_{k_1} Z_{k_1}' P Z_{k_2}\} = \frac{1}{2} \sum_{d=1}^{D} (x_{k_2,n_d}' \Sigma_d^{-1} x_{k_1,n_d})^2 \\
&- \sum_{d=1}^{D} x_{k_2,n_d}' \Sigma_d^{-1} x_{k_1,n_d} x_{k_1,n_d}' \Sigma_d^{-1} X_d R X_d' \Sigma_d^{-1} x_{k_2,n_d} \\
&+ \frac{1}{2} \sum_{d=1}^{D} x_{k_2,n_d}' \Sigma_d^{-1} X_d R \left(\sum_{d=1}^{D} X_d' \Sigma_d^{-1} x_{k_1,n_d} x_{k_1,n_d}' \Sigma_d^{-1} X_d \right) R X_d' \Sigma_d^{-1} x_{k_2,n_d}.
\end{aligned}$$

The inverse of matrix $\boldsymbol{\Sigma}_d$ can be calculated by applying iteratively the formula

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

STEP 1: Define $\boldsymbol{L}_{1d} = \boldsymbol{W}_d^{-1} + \varphi_1 \boldsymbol{x}_{1,n_d}\boldsymbol{x}'_{1,n_d}$. Take $A = \boldsymbol{W}_d^{-1}$, $C = \varphi_1 \boldsymbol{x}_{1,n_d}$, $B = \boldsymbol{I}_1$, and $D = \boldsymbol{x}'_{1,n_d}$, so that

$$\boldsymbol{L}_{1d}^{-1} = \boldsymbol{W}_d - \boldsymbol{W}_d\varphi_1\boldsymbol{x}_{1,n_d}(1+\boldsymbol{x}'_{1,n_d}\boldsymbol{W}_d\varphi_1\boldsymbol{x}_{1,n_d})^{-1}\boldsymbol{x}'_{1,n_d}\boldsymbol{W}_d.$$

STEP 2: Define $\boldsymbol{L}_{2d} = \boldsymbol{L}_{1d} + \varphi_2 \boldsymbol{x}_{2,n_d}\boldsymbol{x}'_{2,n_d}$. Take $A = \boldsymbol{L}_{1d}$, $C = \varphi_2 \boldsymbol{x}_{2,n_d}$, $B = \boldsymbol{I}_1$ and $D = \boldsymbol{x}'_{2,n_d}$, so that

$$\boldsymbol{L}_{2d}^{-1} = \boldsymbol{L}_{1d}^{-1} - \boldsymbol{L}_{1d}^{-1}\varphi_2\boldsymbol{x}_{2,n_d}(1+\boldsymbol{x}'_{2,n_d}\boldsymbol{L}_{1d}^{-1}\varphi_2\boldsymbol{x}_{2,n_d})^{-1}\boldsymbol{x}'_{2,n_d}\boldsymbol{L}_{1d}^{-1}.$$

...

STEP p: Finally, $\boldsymbol{L}_{pd} = \boldsymbol{L}_{p-1d} + \varphi_p \boldsymbol{x}_{p,n_d}\boldsymbol{x}'_{p,n_d}$. Take $A = \boldsymbol{L}_{p-1d}$, $C = \varphi_p \boldsymbol{x}_{p,n_d}$, $B = \boldsymbol{I}_1$, and $D = \boldsymbol{x}'_{p,n_d}$, so that

$$\boldsymbol{L}_{pd}^{-1} = \boldsymbol{L}_{p-1d}^{-1} - \boldsymbol{L}_{p-1d}^{-1}\varphi_p\boldsymbol{x}_{p,n_d}(1+\boldsymbol{x}'_{p,n_d}\boldsymbol{L}_{p-1d}^{-1}\varphi_p\boldsymbol{x}_{p,n_d})^{-1}\boldsymbol{x}'_{p,n_d}\boldsymbol{L}_{p-1d}^{-1}.$$

13.3.3 EBLUP of a Domain Mean

The formulas for the EBLUP of the linear parameter $\overline{Y}_d$ under model (13.4) have exactly the same form as the ones derived under model (13.2) in Sect. 13.2.3. Namely, if $n_d > 0$, then the EBLUP of $\overline{Y}_d$ is

$$\widehat{\overline{Y}}_d^{eblup} = \sum_{k=1}^{p}\overline{X}_{kd}\hat{\beta}_k + \sum_{k=1}^{p}\overline{X}_{kd}\hat{u}_{kd} + f_d\left[\overline{y}_{s,d} - \sum_{k=1}^{p}\overline{X}_{s,kd}\hat{\beta}_k - \sum_{k=1}^{p}\overline{X}_{s,kd}\hat{u}_{kd}\right],$$

where $\overline{X}_{kd} = \frac{1}{N_d}\sum_{j=1}^{N_d} x_{kdj}$, $\overline{y}_{s,d} = \frac{1}{n_d}\sum_{j=1}^{n_d} y_{dj}$, $\overline{X}_{s,kd} = \frac{1}{n_d}\sum_{j=1}^{n_d} x_{kdj}$ and $f_d = \frac{n_d}{N_d}$. If $n_d = 0$, then the EBLUP of $\overline{Y}_d$ is the synthetic part

$$\hat{\mu}_d^{eblup} = \sum_{k=1}^{p}\overline{X}_{kd}\hat{\beta}_k + \sum_{k=1}^{p}\overline{X}_{kd}\hat{u}_{kd}.$$

13.3.4 MSE of the EBLUP

Let $\boldsymbol{\theta} = (\sigma^2, \varphi_1, \ldots, \varphi_p)'$ be the vector of variance components and $\hat{\boldsymbol{\theta}}$ be the corresponding REML estimate. The MSEs of the EBLUPs of $\overline{Y}_d$ and μ_d are (cf. pp. 220 and 217, respectively)

$$MSE(\widehat{\overline{Y}}_d^{eblup}) = g_1(\boldsymbol{\theta})+g_2(\boldsymbol{\theta})+g_3(\boldsymbol{\theta})+g_4(\boldsymbol{\theta}), \quad MSE(\hat{\mu}_d^{eblup}) = g_1(\boldsymbol{\theta})+g_2(\boldsymbol{\theta})+g_3(\boldsymbol{\theta}),$$

where

$$g_1(\boldsymbol{\theta}) = \boldsymbol{a}_r'\boldsymbol{Z}_r\boldsymbol{T}_s\boldsymbol{Z}_r'\boldsymbol{a}_r,$$
$$g_2(\boldsymbol{\theta}) = [\boldsymbol{a}_r'\boldsymbol{X}_r - \boldsymbol{a}_r'\boldsymbol{Z}_r\boldsymbol{T}_s\boldsymbol{Z}_s'\boldsymbol{V}_{es}^{-1}\boldsymbol{X}_s]\boldsymbol{Q}_s[\boldsymbol{X}_r'\boldsymbol{a}_r - \boldsymbol{X}_s'\boldsymbol{V}_{es}^{-1}\boldsymbol{Z}_s\boldsymbol{T}_s\boldsymbol{Z}_r'\boldsymbol{a}_r],$$
$$g_3(\boldsymbol{\theta}) \approx \operatorname{tr}\left\{(\nabla\boldsymbol{b}')\boldsymbol{V}_s(\nabla\boldsymbol{b}')'E\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right]\right\},$$
$$g_4(\boldsymbol{\theta}) = \boldsymbol{a}_r'\boldsymbol{V}_{er}\boldsymbol{a}_r$$

and definitions of the symbols $\boldsymbol{T}_s$, $\boldsymbol{Q}_s$, and $\boldsymbol{b}'$ are given in Sect. 9.2 and will be revised on the following pages. The Prasad-Rao (PR) estimator of $MSE(\widehat{\overline{Y}}_d^{eblup})$ is

$$mse_d = mse(\widehat{\overline{Y}}_d^{eblup}) = g_1(\hat{\boldsymbol{\theta}}) + g_2(\hat{\boldsymbol{\theta}}) + 2g_3(\hat{\boldsymbol{\theta}}) + g_4(\hat{\boldsymbol{\theta}}).$$

Now, we present a detailed description of calculation of the terms $g_i(\boldsymbol{\theta})$, $i = 1, \ldots, 4$, under the present model.

Calculation of $g_1(\theta)$

To calculate $g_1(\boldsymbol{\theta}) = \boldsymbol{a}_r'\boldsymbol{Z}_r\boldsymbol{T}_s\boldsymbol{Z}_r'\boldsymbol{a}_r$, basic elements are

$$\boldsymbol{a}_r' = \frac{1}{N_d}\underset{1\le\ell\le D}{\operatorname{col}'}(\delta_{d\ell}\boldsymbol{1}_{N_\ell-n_\ell}'), \quad \boldsymbol{V}_u = \sigma^2\underset{1\le k\le p}{\operatorname{diag}}(\varphi_k\boldsymbol{I}_D),$$

$$\boldsymbol{Z}_r = \underset{1\le k\le p}{\operatorname{col}'}(\boldsymbol{Z}_{rk}), \quad \boldsymbol{Z}_{rk} = \underset{1\le d\le D}{\operatorname{diag}}(\boldsymbol{x}_{k,N_d-n_d}),$$

$$\boldsymbol{T}_s = \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}_s'\boldsymbol{V}_s^{-1}\boldsymbol{Z}_s\boldsymbol{V}_u = \sigma^2\underset{1\le k\le p}{\operatorname{diag}}(\varphi_k\boldsymbol{I}_D)$$
$$-\sigma^2\underset{1\le k\le p}{\operatorname{col}}(\varphi_k\boldsymbol{Z}_{sk}')\underset{1\le\ell\le D}{\operatorname{diag}}(\boldsymbol{\Sigma}_{s\ell}^{-1})\underset{1\le k\le p}{\operatorname{col}'}(\varphi_k\boldsymbol{Z}_{sk}) = (\boldsymbol{T}_{k_1k_2})_{k_1,k_2=1,\ldots,p},$$

where

$$\boldsymbol{T}_{k_1k_2} = \sigma^2\varphi_{k_1}\delta_{k_1k_2}\boldsymbol{I}_D - \sigma^2\varphi_{k_1}\varphi_{k_2}\boldsymbol{Z}'_{sk_1}\underset{1\le\ell\le D}{\text{diag}}(\boldsymbol{\Sigma}_{s\ell}^{-1})\boldsymbol{Z}_{sk_2}$$
$$= \sigma^2\varphi_{k_1}\delta_{k_1k_2}\boldsymbol{I}_D - \sigma^2\varphi_{k_1}\varphi_{k_2}\underset{1\le\ell\le D}{\text{diag}}(\boldsymbol{x}'_{k_1,n_\ell}\boldsymbol{\Sigma}_{s\ell}^{-1}\boldsymbol{x}_{k_2,n_\ell})$$

and $\delta_{k_1k_2} = 0$ if $k_1 \neq k_2$, $\delta_{k_1k_2} = 1$ if $k_1 = k_2$. Therefore

$$\begin{aligned}
g_1(\theta) &= \boldsymbol{a}'_r\boldsymbol{Z}_r\boldsymbol{T}_s\boldsymbol{Z}'_r\boldsymbol{a}_r \\
&= \frac{1}{N_d^2}\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\boldsymbol{1}'_{N_\ell-n_\ell})\underset{1\le k\le p}{\text{col}'}(\boldsymbol{Z}_{rk})\boldsymbol{T}_s\underset{1\le k\le p}{\text{col}}(\boldsymbol{Z}'_{rk})\underset{1\le\ell\le D}{\text{col}}(\delta_{d\ell}\boldsymbol{1}_{N_\ell-n_\ell}) \\
&= \frac{1}{N_d^2}\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\boldsymbol{1}'_{N_\ell-n_\ell})\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\boldsymbol{Z}_{rk_1}\boldsymbol{T}_{k_1k_2}\boldsymbol{Z}'_{rk_2}\underset{1\le\ell\le D}{\text{col}}(\delta_{d\ell}\boldsymbol{1}_{N_\ell-n_\ell}) \\
&= \frac{1}{N_d^2}\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\boldsymbol{1}'_{N_\ell-n_\ell}\boldsymbol{x}_{k_1,N_\ell-n_\ell})\boldsymbol{T}_{k_1k_2}\underset{1\le\ell\le D}{\text{col}}(\delta_{d\ell}\boldsymbol{x}'_{k_2,N_\ell-n_\ell}\boldsymbol{1}_{N_\ell-n_\ell}) \\
&= (1-f_d)^2\sigma^2\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\underset{1\le\ell\le D}{\text{col}'}(\delta_{d\ell}\overline{X}^*_{k_1\ell}) \\
&\cdot\left[\varphi_{k_1}\delta_{k_1k_2}\boldsymbol{I}_D - \varphi_{k_1}\varphi_{k_2}\underset{1\le\ell\le D}{\text{diag}}(\boldsymbol{x}'_{k_1,n_\ell}\boldsymbol{\Sigma}_{s\ell}^{-1}\boldsymbol{x}_{k_2,n_\ell})\right]\underset{1\le\ell\le D}{\text{col}}(\delta_{d\ell}\overline{X}^*_{k_2\ell}) \\
&= (1-f_d)^2\sigma^2\left\{\sum_{k=1}^{p}\varphi_k\overline{X}^{*2}_{kd} - \sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\varphi_{k_1}\varphi_{k_2}\overline{X}^*_{k_1d}\boldsymbol{x}'_{k_1,n_d}\boldsymbol{\Sigma}_{sd}^{-1}\boldsymbol{x}_{k_2,n_d}\overline{X}^*_{k_2d}\right\},
\end{aligned}$$

where $f_d = n_d/N_d$ and $\overline{X}^*_{kd} = \frac{1}{N_d-n_d}\sum_{j\in r_d} x_{kdj} = (1-f_d)^{-1}(\overline{X}_{kd} - f_d\overline{x}_{kd})$.

Calculation of $g_2(\theta)$

We recall that

$$\begin{aligned}
g_2(\boldsymbol{\theta}) &= [\boldsymbol{a}'_r\boldsymbol{X}_r - \boldsymbol{a}'_r\boldsymbol{Z}_r\boldsymbol{T}_s\boldsymbol{Z}'_s\boldsymbol{V}_{es}^{-1}\boldsymbol{X}_s]\boldsymbol{Q}_s[\boldsymbol{X}'_r\boldsymbol{a}_r - \boldsymbol{X}'_s\boldsymbol{V}_{es}^{-1}\boldsymbol{Z}_s\boldsymbol{T}_s\boldsymbol{Z}'_r\boldsymbol{a}_r] \\
&= [\boldsymbol{a}'_1 - \boldsymbol{a}'_2]\boldsymbol{Q}_s[\boldsymbol{a}_1 - \boldsymbol{a}_2],
\end{aligned}$$

where $\boldsymbol{Q}_s = (\boldsymbol{X}_s'\boldsymbol{V}^{-1}\boldsymbol{X}_s)^{-1} = \sigma^2\left(\sum_{d=1}^{D}\boldsymbol{X}_{sd}'\boldsymbol{\Sigma}_{sd}^{-1}\boldsymbol{X}_{sd}\right)^{-1}$ and $\boldsymbol{V}_{es}^{-1} = \sigma^{-2}\boldsymbol{W}_s$. The second vector is

$$\boldsymbol{a}_2' = \boldsymbol{a}_r' \underset{1\le k\le p}{\mathrm{col}'}(\boldsymbol{Z}_{rk})\boldsymbol{T}_s \underset{1\le k\le p}{\mathrm{col}}(\boldsymbol{Z}_{sk}')\sigma^{-2}\boldsymbol{W}_s\boldsymbol{X}_s = \sigma^{-2}\boldsymbol{a}_r' \sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\boldsymbol{Z}_{rk_1}\boldsymbol{T}_{k_1k_2}\boldsymbol{Z}_{sk_2}'\boldsymbol{W}_s\boldsymbol{X}_s$$

$$= \frac{1}{N_d}\underset{1\le\ell\le D}{\mathrm{col}'}(\delta_{d\ell}\boldsymbol{1}_{N_\ell-n_\ell}')\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\underset{1\le\ell\le D}{\mathrm{diag}}(\boldsymbol{x}_{k_1,N_\ell-n_\ell})$$

$$\cdot\left[\varphi_{k_1}\delta_{k_1k_2}\boldsymbol{I}_D - \varphi_{k_1}\varphi_{k_2}\underset{1\le\ell\le D}{\mathrm{diag}}(\boldsymbol{x}_{k_1,n_\ell}'\boldsymbol{\Sigma}_{s\ell}^{-1}\boldsymbol{x}_{k_2,n_\ell})\right]$$

$$\cdot\underset{1\le\ell\le D}{\mathrm{diag}}(\boldsymbol{x}_{k_2,n_\ell}')\underset{1\le\ell\le D}{\mathrm{diag}}(\boldsymbol{W}_{s\ell})\underset{1\le\ell\le D}{\mathrm{col}}(\boldsymbol{X}_{s\ell})$$

$$= \frac{1}{N_d}\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\underset{1\le\ell\le D}{\mathrm{col}'}(\delta_{d\ell}\boldsymbol{1}_{N_\ell-n_\ell}'\boldsymbol{x}_{k_1,N_\ell-n_\ell})$$

$$\cdot\left[\varphi_{k_1}\delta_{k_1k_2}\boldsymbol{I}_D - \varphi_{k_1}\varphi_{k_2}\underset{1\le\ell\le D}{\mathrm{diag}}(\boldsymbol{x}_{k_1,n_\ell}'\boldsymbol{\Sigma}_{s\ell}^{-1}\boldsymbol{x}_{k_2,n_\ell})\right]$$

$$\cdot\underset{1\le\ell\le D}{\mathrm{diag}}(\boldsymbol{x}_{k_2,n_\ell}')\underset{1\le\ell\le D}{\mathrm{col}}(\boldsymbol{W}_{s\ell}\boldsymbol{X}_{s\ell})$$

$$=(1-f_d)\left\{\sum_{k=1}^{p}\varphi_k\overline{X}_{kd}^*\boldsymbol{x}_{k,n_d}' - \sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\varphi_{k_1}\varphi_{k_2}\overline{X}_{k_1d}^*\boldsymbol{x}_{k_1,n_d}'\boldsymbol{\Sigma}_{sd}^{-1}\boldsymbol{x}_{k_2,n_d}\boldsymbol{x}_{k_2,n_d}'\right\}\boldsymbol{W}_{sd}\boldsymbol{X}_{sd}.$$

The first vector is

$$\boldsymbol{a}_1' = \boldsymbol{a}_r'\boldsymbol{X}_r = \frac{1}{N_d}\boldsymbol{1}_{N_d-n_d}'\boldsymbol{X}_{rd} = \frac{1}{N_d}\sum_{j\in r_d}\boldsymbol{x}_{dj} = (1-f_d)\overline{\boldsymbol{X}}_d^*, \quad \overline{\boldsymbol{X}}_d^* = (\overline{X}_{1d}^*,\ldots,\overline{X}_{pd}^*).$$

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

We recall that

$$g_3(\boldsymbol{\theta}) \approx \mathrm{tr}\left\{(\nabla\boldsymbol{b}')\boldsymbol{V}_s(\nabla\boldsymbol{b}')'E\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right]\right\},$$

where

$$\boldsymbol{b}' = \boldsymbol{a}_r'\boldsymbol{Z}_r\boldsymbol{V}_u\boldsymbol{Z}_s'\boldsymbol{V}_s^{-1} = \boldsymbol{a}_r'\underset{1\le k\le p}{\mathrm{col}'}(\boldsymbol{Z}_{rk})\underset{1\le k\le p}{\mathrm{diag}}(\varphi_k\boldsymbol{I}_D)\underset{1\le k\le p}{\mathrm{col}}(\boldsymbol{Z}_{sk}')\boldsymbol{\Sigma}_s^{-1}$$

$$= \boldsymbol{a}_r'\sum_{k=1}^{p}\varphi_k\boldsymbol{Z}_{rk}\boldsymbol{Z}_{sk}'\underset{1\le\ell\le D}{\mathrm{diag}}(\boldsymbol{\Sigma}_{s\ell}^{-1}).$$

The first derivative is $\frac{\partial \boldsymbol{b}'}{\partial \sigma^2} = 0$. As $\frac{\partial \boldsymbol{\Sigma}_{s\ell}}{\partial \varphi_k} = \boldsymbol{x}_{k,n_\ell} \boldsymbol{x}'_{k,n_\ell}$, the remaining derivatives are

$$\frac{\partial \boldsymbol{b}'}{\partial \varphi_k} = \boldsymbol{a}'_r \boldsymbol{Z}_{rk} \boldsymbol{Z}'_{sk} \underset{1\le\ell\le D}{\text{diag}} (\boldsymbol{\Sigma}_{s\ell}^{-1}) - \boldsymbol{a}'_r \left(\sum_{i=1}^{p} \varphi_i \boldsymbol{Z}_{ri} \boldsymbol{Z}'_{si} \right) \underset{1\le\ell\le D}{\text{diag}} (\boldsymbol{\Sigma}_{s\ell}^{-1} \boldsymbol{x}_{k,n_\ell} \boldsymbol{x}'_{k,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1}),$$

$k = 1, \ldots, p$, where the formula for derivative of an inverse matrix given in Appendix A was used. As $\boldsymbol{Z}_{rk} = \underset{1\le\ell\le D}{\text{diag}} (\boldsymbol{x}_{k,N_\ell - n_\ell})$, then

$$\begin{aligned}
\frac{\partial \boldsymbol{b}'}{\partial \varphi_k} &= \frac{1}{N_d} \underset{1\le\ell\le D}{\text{col}'} (\delta_{d\ell} \boldsymbol{1}'_{N_\ell - n_\ell}) \Bigg[\underset{1\le\ell\le D}{\text{diag}} (\boldsymbol{x}_{k,N_\ell - n_\ell} \boldsymbol{x}'_{k,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1}) \\
&\quad - \sum_{i=1}^{p} \varphi_i \underset{1\le\ell\le D}{\text{diag}} (\boldsymbol{x}_{i,N_\ell - n_\ell} \boldsymbol{x}'_{i,n_\ell}) \underset{1\le\ell\le D}{\text{diag}} (\boldsymbol{\Sigma}_{s\ell}^{-1} \boldsymbol{x}_{k,n_\ell} \boldsymbol{x}'_{k,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1}) \Bigg] \\
&= \frac{1}{N_d} \underset{1\le\ell\le D}{\text{col}'} (\delta_{d\ell} \boldsymbol{1}'_{N_\ell - n_\ell} \boldsymbol{x}_{k,N_\ell - n_\ell} \boldsymbol{x}'_{k,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1}) \\
&\quad - \frac{1}{N_d} \underset{1\le\ell\le D}{\text{col}'} \left(\delta_{d\ell} \left(\sum_{i=1}^{p} \varphi_i \boldsymbol{1}'_{N_\ell - n_\ell} \boldsymbol{x}_{i,N_\ell - n_\ell} \boldsymbol{x}'_{i,n_\ell} \right) \boldsymbol{\Sigma}_{s\ell}^{-1} \boldsymbol{x}_{k,n_\ell} \boldsymbol{x}'_{k,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1} \right) \\
&= (1 - f_d) \Bigg[\underset{1\le\ell\le D}{\text{col}'} (\delta_{d\ell} \overline{X}^*_{k\ell} \boldsymbol{x}'_{k,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1}) \\
&\quad - \underset{1\le\ell\le D}{\text{col}'} \left(\delta_{d\ell} \left(\sum_{i=1}^{p} \varphi_i \overline{X}^*_{i\ell} \boldsymbol{x}'_{i,n_\ell} \right) \boldsymbol{\Sigma}_{s\ell}^{-1} \boldsymbol{x}_{k,n_\ell} \boldsymbol{x}'_{k,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1} \right) \Bigg], \quad k = 1, \ldots, p.
\end{aligned}$$

Let us define $\boldsymbol{Q}(\boldsymbol{\theta}) = (q_{k_1,k_2})_{k_1,k_2=0,1,\ldots,p}$, where $q_{0,k} = q_{k,0} = 0$, $k = 0, 1, \ldots, p$ and

$$\begin{aligned}
q_{k_1,k_2} &= \frac{\partial \boldsymbol{b}'}{\partial \varphi_{k_1}} \boldsymbol{V}_s \left(\frac{\partial \boldsymbol{b}'}{\partial \varphi_{k_2}} \right)' = \sigma^2 (1 - f_d)^2 \Bigg[\underset{1\le\ell\le D}{\text{col}'} (\delta_{d\ell} \overline{X}^*_{k_1\ell} \boldsymbol{x}'_{k_1,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1}) \\
&\quad - \underset{1\le\ell\le D}{\text{col}'} \left[\delta_{d\ell} \left(\sum_{i=1}^{p} \varphi_i \overline{X}^*_{i\ell} \boldsymbol{x}'_{i,n_\ell} \right) \boldsymbol{\Sigma}_{s\ell}^{-1} \boldsymbol{x}_{k_1,n_\ell} \boldsymbol{x}'_{k_1,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1} \right] \Bigg] \\
&\quad \cdot \underset{1\le\ell\le D}{\text{diag}} (\boldsymbol{\Sigma}_{s\ell}) \Bigg[\underset{1\le\ell\le D}{\text{col}} (\delta_{d\ell} \boldsymbol{\Sigma}_{s\ell}^{-1} \boldsymbol{x}_{k_2,n_\ell} \overline{X}^*_{k_2\ell}) \\
&\quad - \underset{1\le\ell\le D}{\text{col}} \left[\delta_{d\ell} \boldsymbol{\Sigma}_{s\ell}^{-1} \boldsymbol{x}_{k_2,n_\ell} \boldsymbol{x}'_{k_2,n_\ell} \boldsymbol{\Sigma}_{s\ell}^{-1} \left(\sum_{i=1}^{p} \varphi_i \boldsymbol{x}_{i,n_\ell} \overline{X}^*_{i\ell} \right) \right] \Bigg]
\end{aligned}$$

for any $k_1, k_2 = 1, \ldots, p$. After further simplifications we obtain

$$
\begin{aligned}
q_{k_1,k_2} &= \sigma^2(1-f_d)^2\, \overline{X}^*_{k_1 d} \boldsymbol{x}'_{k_1,n_d} \boldsymbol{\Sigma}^{-1}_{sd} \boldsymbol{x}_{k_2,n_d} \overline{X}^*_{k_2 d} \\
&- \sigma^2(1-f_d)^2\, \overline{X}^*_{k_1 d} \boldsymbol{x}'_{k_1,n_d} \boldsymbol{\Sigma}^{-1}_{sd} \boldsymbol{x}_{k_2,n_d} \boldsymbol{x}'_{k_2,n_d} \boldsymbol{\Sigma}^{-1}_{sd} \left(\sum_{i=1}^{p} \varphi_i \boldsymbol{x}_{i,n_d} \overline{X}^*_{id} \right) \\
&- \sigma^2(1-f_d)^2 \left(\sum_{i=1}^{p} \varphi_i \overline{X}^*_{id} \boldsymbol{x}'_{i,n_d} \right) \boldsymbol{\Sigma}^{-1}_{sd} \boldsymbol{x}_{k_1,n_d} \boldsymbol{x}'_{k_1,n_d} \boldsymbol{\Sigma}^{-1}_{sd} \boldsymbol{x}_{k_2,n_d} \overline{X}^*_{k_2 d} \\
&+ \sigma^2(1-f_d)^2 \left(\sum_{i=1}^{p} \varphi_i \overline{X}^*_{id} \boldsymbol{x}'_{i,n_d} \right) \boldsymbol{\Sigma}^{-1}_{sd}\, \boldsymbol{x}_{k_1,n_d} \boldsymbol{x}'_{k_1,n_d} \boldsymbol{\Sigma}^{-1}_{sd} \boldsymbol{x}_{k_2,n_d} \boldsymbol{x}'_{k_2,n_d} \boldsymbol{\Sigma}^{-1}_{sd} \\
&\cdot \left(\sum_{i=1}^{p} \varphi_i \boldsymbol{x}_{i,n_d} \overline{X}^*_{id} \right), \quad k_1, k_2 = 1, \ldots, p.
\end{aligned}
$$

Then

$$
g_3(\boldsymbol{\theta}) \approx \operatorname{tr} \left\{ \boldsymbol{Q}(\boldsymbol{\theta}) \boldsymbol{F}^{-1}(\boldsymbol{\theta}) \right\},
$$

where $\boldsymbol{F}(\boldsymbol{\theta})$ is the REML Fisher information matrix.

Calculation of $g_4(\boldsymbol{\theta})$

We recall that $g_4(\boldsymbol{\theta}) = \boldsymbol{a}'_r \boldsymbol{V}_{er} \boldsymbol{a}_r$, where

$$
\boldsymbol{a}'_r = \frac{1}{N_d} \underset{1 \le \ell \le D}{\operatorname{col}'} (\delta_{d\ell} \mathbf{1}'_{N_\ell - n_\ell}), \quad \boldsymbol{V}^{-1}_{er} = \sigma^{-2} \boldsymbol{W}_r = \sigma^{-2} \underset{1 \le d \le D}{\operatorname{diag}} \{\boldsymbol{W}_{rd}\}.
$$

Therefore

$$
\begin{aligned}
g_4(\boldsymbol{\theta}) &= \frac{1}{N_d} \underset{1 \le \ell \le D}{\operatorname{col}'} (\delta_{d\ell} \mathbf{1}'_{N_\ell - n_\ell}) \sigma^2 \underset{1 \le \ell \le D}{\operatorname{diag}} \{\boldsymbol{W}^{-1}_{r\ell}\} \frac{1}{N_d} \underset{1 \le \ell \le D}{\operatorname{col}} (\delta_{d\ell} \mathbf{1}_{N_\ell - n_\ell}) \\
&= \frac{\sigma^2}{N_d^2} \mathbf{1}'_{N_d - n_d} \underset{j \in r_d}{\operatorname{diag}} \{w^{-1}_{dj}\} \mathbf{1}_{N_d - n_d} = \frac{\sigma^2}{N_d^2} \sum_{j \in r_d} \frac{1}{w_{dj}}.
\end{aligned}
$$

13.4 R Codes for EBLUPs

This section gives R codes for fitting the RRC model to the survey data file `LFS20.txt`. The target variable y is the variable INCOME. As auxiliary variables we take REGISTERED and EDUCATION. The function `dir2` is employed for calculating direct estimators. The domains are defined by the variable AREA crossed by SEX. The parameters of interest are the income means by domains.

We install and/or load some R packages: `Matrix`, `lme4`, and `sae`.

```
if(!require(Matrix)){
  install.packages("Matrix")
  library(Matrix)
}
if(!require(lme4)){
  install.packages("lme4")
  library(lme4)
}
if(!require(sae)){
  install.packages("sae")
  library(sae)
}
```

The following code reads the data files and calculate some variables:

```
# Read unit-level data
dat <- read.table("LFS20.txt", header=TRUE, sep = "\t", dec = ".")
# Education level 2
edu2 <- as.numeric(dat$EDUCATION==2)
# Education level 3
edu3 <- as.numeric(dat$EDUCATION==3)
# Read domain-level data
aux <- read.table("Nds20.txt", header=TRUE, sep = "\t", dec = ".")
# Prop. of registered people
aux$mreg <- aux$reg/aux$N
# Proportion of edu2 people
aux$medu2 <- aux$edu2/aux$N
# Proportion of edu3 people
aux$medu3 <- aux$edu3/aux$N
```

We calculate direct estimators of domain average incomes and the population sizes by domain, by using `dir2` function described in Sect. 2.8.4. We also define some new variables.

```
income.dir <- dir2(data=dat$INCOME, w=dat$WEIGHT, domain=list(sex=dat$SEX,
                   area=dat$AREA))
diry <- income.dir$mean        # Direct estimates of domain means
hatNd <- income.dir$Nd         # Direct estimates of population sizes
nd <- income.dir$nd            # Sample sizes
fd <- nd/aux$N                 # Sample fractions
```

The following code calculates sample means by domains:

```
dat2 <- data.frame(income=dat$INCOME, edu2, edu3, reg=dat$REGISTERED)
smeans <- aggregate(dat2, by=list(sex=dat$SEX,area=dat$AREA), mean)
meany <- smeans$income           # Sample means of income
meanedu2 <- smeans$edu2          # Sample means of edu2
meanedu3 <- smeans$edu3          # Sample means of edu3
meanreg <- smeans$reg            # Sample means of registered
```

We fit a random regression coefficient model with INCOME as dependent variable and REGISTERED and EDUCATION as explanatory variables. The fitted model has a

Table 13.1 Estimated regression parameters of RRC model

Parameter	Estimate	Std. error	t-value	p-value
Intercept	40,187.0	485.6	82.75	0.00
Registered	−11,742.2	1124.9	−10.44	0.00
edu2	9704.1	674.8	14.38	0.00
edu3	20,013.6	957.0	20.91	0.00

random intercept and random slopes on the coefficients of the categories edu2 and edu3 of the variable EDUCATION. The employed R code is

```
dat$EDUCATION <- as.factor(dat$EDUCATION)
rrc <- lmer(formula=INCOME ~ REGISTERED + EDUCATION + (EDUCATION|AREA:SEX),
            data=dat, REML=FALSE)
summary(rrc)                                  # Summary of the fitting procedure
anova(rrc)                                    # Analysis of Variance Table
beta <- fixef(rrc); beta                      # Regression parameters
var <- as.data.frame(VarCorr(rrc))            # Variance parameters
ref <- ranef(rrc)[[1]]                        # Modes of the random effects
head(fitted(rrc))                             # Predicted values
residuals <- resid(rrc)                       # Residuals
p.values <- 2*pnorm(abs(coef(summary(rrc))[,3]), low=F)
p.values                                      # p values
```

Table 13.1 presents the estimated regression parameters and p-values.
We calculate the EBLUPs of income means by domain.

```
Xbeta <- beta[1] + beta[2]*aux$mreg + beta[3]*aux$medu2 + beta[4]*aux$medu3
Xubeta <- ref[,1] + aux$medu2*ref[,2] + aux$medu3*ref[,3]
mu <- Xbeta + Xubeta                # Projective estimates of income means
xbeta <- beta[1] + beta[2]*meanreg + beta[3]*meanedu2 + beta[4]*meanedu3
xubeta <- ref[,1] + meanedu2*ref[,2] + meanedu3*ref[,3]
mu.s <- meany - xbeta - xubeta
eb <- mu + fd*mu.s                  # EBLUPs of income means
```

Summary of results

```
output <- data.frame(Nd=aux$N[c(T,F)], hatNd=hatNd[c(T,F)] , nd=nd[c(T,F)],
           meany=round(meany[c(T,F)],0), dir=round(diry[c(T,F)],0),
           hatmu=round(mu[c(T,F)],0), eblup=round(eb[c(T,F)],0))
head(output, 10)
```

Figure 13.1 (left) plots the RRC model residuals $\hat{e}_d = y_d - \hat{y}_d$. The residuals are situated symmetrically around 0. Figure 13.1 (right) plots the EBLUPs and direct estimates of men income means by areas. The EBLUPs behave more smoothly than the direct estimators.

For the ten first areas, Table 13.2 gives a summary of results for men. The population sizes, the estimated population sizes, and the sample sizes are denoted by N_d, $\hat{N}_d$, and n_d, respectively. The columns meany and dir contain the sample means and the direct estimates of the population mean of the variable income. The projective predictors and the EBLUPs of the population means are labelled by hatmu and eblup, respectively.

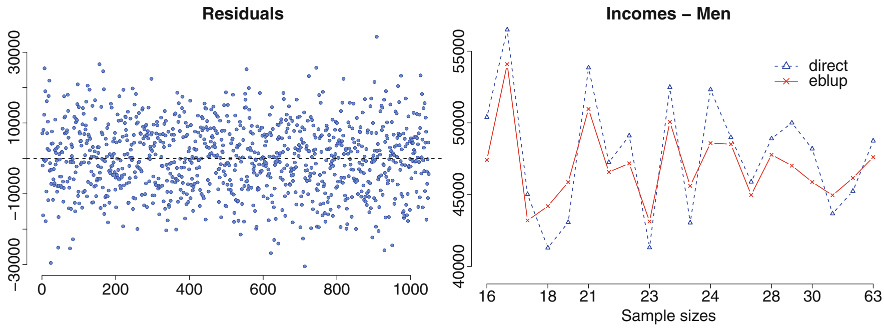

Fig. 13.1 Plots of residuals (left) and estimated men income means (right)

Table 13.2 Estimates of domain mean incomes for men

Area	N_d	$\hat{N}_d$	n_d	meany	dir	hatmu	eblup
1	8020	7950	29	49,729	50,005	47,013	47,021
2	3576	3522	18	43,847	45,019	43,191	43,199
3	4446	4359	24	51,968	52,482	50,053	50,065
4	4807	4740	27	45,748	45,889	44,974	44,977
5	3252	3375	22	46,247	47,250	46,575	46,574
6	5461	5316	30	48,562	48,193	45,863	45,879
7	4023	3984	23	48,082	49,103	47,167	47,177
8	3816	3796	23	42,005	41,310	43,140	43,130
9	10,785	10,873	63	48,202	48,746	47,607	47,612
10	3256	3195	19	43,684	43,068	45,878	45,865

References

Dempster, A.P., Rubin, D.B., Tsutakawa, R.K.: Estimation in covariance component models. J. Am. Stat. Assoc. **76**, 341–353 (1981)

Hobza, T., Morales, D.: Small area estimation under random regression coefficient models. J. Stat. Comput. Simul. **83**(11), 2160–2177 (2013)

Moura, F.A.S., Holt, D.: Small area estimation using multilevel models. Surv. Methodol. **25**(1), 73–80 (1999)

Prasad, N.G.N., Rao, J.N.K.: The estimation of the mean squared error of small-area estimators. J. Am. Stat. Assoc. **85**, 163–171 (1990)