R package sae: Methodology

Isabel Molina*, Yolanda Marhuenda

March, 2015

Contents

1 Introduction **3**

2 Function direct **3**
2.1 Sampling without replacement 3
2.2 Sampling with replacement 5

3 Function pssynt **5**

4 Function ssd **6**

5 Function eblupFH **7**
5.1 FH fitting method 9
5.2 ML fitting method 10
5.3 REML fitting method 11

6 Function mseFH **12**
6.1 FH fitting method 13
6.2 ML fitting method 14
6.3 REML fitting method 14

7 Function eblupSFH **14**
7.1 ML fitting method 16
7.2 REML fitting method 17

8 Function mseSFH **18**
8.1 REML fitting method 19
8.2 ML fitting method 19

*Department of Statistics, Universidad Carlos III de Madrid. Address: C/Madrid 126, 28903 Getafe (Madrid), Spain, Tf: +34 916249887, Fax: +34 916249849, E-mail: isabel.molina@uc3m.es

9 Function pbmseSFH **20**

10 Function npbmseSFH **21**

11 Function eblupSTFH **23**
11.1 REML fitting method 26

12 Function pbmseSTFH **27**

13 Function eblupBHF **28**

14 Function pbmseBHF **30**

15 Function ebBHF **31**

16 Function pbmseebBHF **32**

References **33**

1 Introduction

The R package `sae` estimates characteristics of the domains of a population, when a sample is available from the whole population, but not all the target domains have a sufficiently large sample size. This document describes the methodology behind the functions of the package, giving the exact formulas and procedures implemented in each function.

The following notation will be used through all this document. U denotes the target population, assumed to be partitioned into D subsets $U_1, \ldots, U_D$ called indistinctly domains or areas, of respective sizes $N_1, \ldots, N_D$. The measurement of the variable of interest for j-th individual within d-th area is denoted Y_{dj} and $\mathbf{y}_d = (Y_{d1}, \ldots, Y_{dN_d})'$ is the vector with the measurements for d-th area. The target domain parameters have the form $\delta_d = h_d(\mathbf{y}_d)$, $d = 1, \ldots, D$, for known real measurable functions $h_d()$. Particular cases of δ_d are the domain means

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

These parameters are estimated using the information coming from a random sample s of size n drawn from U. Here, s_d is the subsample of size n_d from domain U_d, $d = 1, \ldots, D$, where $n = \sum_{d=1}^{D} n_d$, and $r_d = U_d - s_d$ is the sample complement from domain d, of size $N_d - n_d$, $d = 1, \ldots, D$.

2 Function direct

Function `direct` estimates the area means $\bar{Y}_d$, $d = 1, \ldots, D$, where, for each area d, the estimator of $\bar{Y}_d$ is calculated using only the sample data s_d from that same domain d. The obtained estimators are unbiased with respect to the sampling design (design-unbiased). The call to the function is

```
direct(y, dom, sweight, domsize, data, replace = FALSE)
```

The particular estimator calculated by the function depends on the specified arguments `replace` and `sweight`, related to the sampling design. If `replace` is `TRUE`, the sampling design is assumed to be with replacement and otherwise without replacement. The sampling weights should be introduced through the argument `sweight`, but when this argument is dropped, the function assumes simple random sampling (SRS). Under SRS, the sizes of the domains N_d, $d = 1, \ldots, D$ (`domsize`) are unnecessary.

2.1 Sampling without replacement

Consider that the sample s_d is drawn without replacement within domain U_d, $d = 1, \ldots, D$. Let π_{dj} be the inclusion probability of j-th unit from d-th domain in the corresponding domain sample s_d and let $w_{dj} = \pi_{dj}^{-1}$ be the corresponding sampling

weight. The unbiased estimator of $\bar{Y}_d$ is the Horvitz-Thompson estimator, given by

$$\hat{\bar{Y}}_d^{DIR} = N_d^{-1} \sum_{j \in s_d} \frac{Y_{dj}}{\pi_{dj}} = N_d^{-1} \sum_{j \in s_d} w_{dj} Y_{dj}.$$

Now let $\pi_{d,jk}$ be the inclusion probability of the pair of units j and k from d-th domain in the sample s_d. The sampling variance is given by

$$V_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{N_d^2} \left\{ \sum_{j=1}^{N_d} \frac{Y_{dj}^2}{\pi_{dj}} (1 - \pi_{dj}) + 2 \sum_{j=1}^{N_d} \sum_{\substack{k=1 \\ k>j}}^{N_d} \frac{Y_{dj}}{\pi_{dj}} \frac{Y_{dk}}{\pi_{dk}} (\pi_{d,jk} - \pi_{dj}\pi_{dk}) \right\}.$$

If $\pi_{d,jk} > 0$, $\forall (j,k)$, an unbiased estimator of this variance is given by

$$\hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{N_d^2} \left\{ \sum_{j \in s_d} \frac{Y_{dj}^2}{\pi_{dj}^2} (1 - \pi_{dj}) + 2 \sum_{j \in s_d} \sum_{\substack{k \in s_d \\ k>j}} \frac{Y_{dj}}{\pi_{dj}} \frac{Y_{dk}}{\pi_{dk}} \frac{(\pi_{d,jk} - \pi_{dj}\pi_{dk})}{\pi_{d,jk}} \right\}. \quad (1)$$

This estimator requires the second order inclusion probabilities, but many times these are not available. Then, it is common to find an approximation of (1). A simple approximation is obtained by considering $\pi_{d,jk} \approx \pi_{dj}\pi_{dk}$, which holds exactly in the case of Poisson sampling. This approximation makes the second sum in (1) equal to zero and leads to the estimator

$$\hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{N_d^2} \sum_{j \in s_d} \frac{1 - \pi_{dj}}{\pi_{dj}^2} Y_{dj}^2.$$

Writing the approximate unbiased estimator of the variance in terms of the sampling weights w_{dj} (`sweight`), we get the expression provided by function `direct` when the argument `sweight` is specified,

$$\hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{N_d^2} \sum_{j \in s_d} w_{dj}(w_{dj} - 1) Y_{dj}^2.$$

Under SRS without replacement, the result of the previous estimator does not coincide with the usual unbiased estimator. Thus, when the argument `sweight` is dropped, the function `direct` assumes SRS without replacement and returns the usual unbiased estimators under this design

$$\hat{\bar{Y}}_d = \bar{y}_d = n_d^{-1} \sum_{j \in s_d} Y_{dj}, \quad \hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = (1 - f_d) \frac{S_d^2}{n_d},$$

where $f_d = n_d/N_d$ is the domain sampling fraction and S_d^2 is the domain quasi-variance $S_d^2 = (n_d - 1)^{-1} \sum_{j \in s_d} (Y_{dj} - \bar{y}_d)^2$.

2.2 Sampling with replacement

Now consider the case in which s_d is drawn with replacement within domain U_d, $d = 1, \ldots, D$, with initial selection probabilities P_{dj}, $j = 1, \ldots, N_d$. In this case, the unbiased estimator of the domain mean is given by

$$\hat{\bar{Y}}_d^{DIR} = N_d^{-1} \sum_{j \in s_d} \frac{Y_{dj}}{n_d P_{dj}}.$$

To obtain this estimator, the argument `sweight` must contain the vector of weights calculated as $w_{dj} = (n_d P_{dj})^{-1}$, $j \in s_d$, $d = 1, \ldots, D$. Using these weights, the unbiased estimator of $\bar{Y}_d$ calculated by the function `direct` with `replace=TRUE` has exactly the same shape as that one in Section 2.1, i.e.

$$\hat{\bar{Y}}_d^{DIR} = N_d^{-1} \sum_{j \in s_d} w_{dj} Y_{dj}.$$

The sampling variance of this estimator is given by

$$V_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{n_d} \sum_{j=1}^{N_d} \left(\frac{Y_{dj}}{N_d P_{dj}} - \bar{Y}_d \right)^2 P_{dj}.$$

and using again $w_{dj} = (n_d P_{dj})^{-1}$, we obtain the unbiased variance estimator calculated by the function `direct`,

$$\hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{n_d} \sum_{j \in s_d} \left(\frac{Y_{dj}}{N_d P_{dj}} - \hat{\bar{Y}}_d^{DIR} \right)^2 = \frac{1}{n_d} \sum_{j \in s_d} \left(f_d w_{dj} Y_{dj} - \hat{\bar{Y}}_d \right)^2.$$

Under SRS with replacement, the population sizes N_d (`domsize`) are not needed. Thus, when the argument `domsize` is dropped, the function assumes SRS and calculates the classical unbiased estimators

$$\hat{\bar{Y}}_d^{DIR} = \bar{y}_d = n_d^{-1} \sum_{j \in s_d} Y_{dj}, \quad \hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = n_d^{-1} S_d^2.$$

3 Function pssynt

Indirect estimators “borrow strength" from other domains by making assumptions establishing some homogeneity relationship among domains. The post-stratified synthetic estimator is a basic indirect estimator. It assumes that data are distributed into K groups (called post-strata) that cut across the domains, and such that the within group mean is constant across domains. The groups are assumed to have sufficient sample sizes to allow obtaining accurate direct estimates of the group means. These assumptions allow us to estimate a domain mean using a weighted combination of the (reliable) estimates of the group means. The function `pssynt` calculates post-stratified synthetic estimates of the domain means $\bar{Y}_d$, $d = 1, \ldots, D$. The call to this function is

```
pssynt(y, sweight, ps, domsizebyps, data)
```

More specifically, the post-stratified synthetic estimator considers that there is a grouping variable (`ps`) which divides the data into K post-strata. The population count in the crossing between post-stratum k and domain d, N_{dk} (`domsizebyps`), is assumed to be known for all k and d with $N_d = \sum_{k=1}^{K} N_{dk}$. Then, the mean of domain d can be calculated as a weighted combination of the means in the crossings of domain d with each post-strata $\bar{Y}_{dk}$, $k = 1, \ldots, K$, as follows

$$\bar{Y}_d = \frac{1}{N_d} \sum_{k=1}^{K} N_{dk} \bar{Y}_{dk}. \tag{2}$$

Under the assumption of constant mean across domains within each post-stratum,

$$\bar{Y}_{dk} = \bar{Y}_{+k}, \quad k = 1, \ldots, K,$$

where $\bar{Y}_{+k}$ denotes the mean of post-stratum k, an estimator of $\bar{Y}_d$ can be obtained by replacing $\bar{Y}_{+k} = \bar{Y}_{dk}$ in (2) by some direct estimate of $\bar{Y}_{+k}$, $k = 1, \ldots, K$. Thus, we estimate the domain mean $\bar{Y}_d$ using all the observations from the post-strata that cut across that domain. To estimate $\bar{Y}_{+k}$, we consider the ratio HT estimator, given by

$$\hat{\bar{Y}}_{+k} = \frac{\hat{Y}_{+k}^{DIR}}{\hat{N}_{+k}^{DIR}}, \tag{3}$$

where $\hat{Y}_{+k}^{DIR}$ is the direct estimator of the total Y_{+k}^{DIR} in post-stratum k and $\hat{N}_{+k}^{DIR}$ is the direct estimator of the population count in the same post-stratum, N_{+k}, calculated using the sampling weights w_{dj} (`sweight`) of the units in that post-stratum. Replacing (3) for $\bar{Y}_{dk}$ in (2), we obtain the post-stratified synthetic estimate returned by the function `pssynt`,

$$\hat{\bar{Y}}_d^{SYN} = \frac{1}{N_d} \sum_{k=1}^{K} N_{dk} \hat{\bar{Y}}_{+k}.$$

4 Function ssd

The direct estimator of $\bar{Y}_d$ is inefficient for a domain d with a small sample size. On the other hand, the post-stratified synthetic estimator is biased when the means across domains within a post-stratum are not constant, which is likely to occur in practice. To balance the bias of a synthetic estimator and the instability of a direct estimator, Drew, Singh & Choudhry (1982) proposed to take a weighted combination (or composition) of the two, with weight depending on the sample size of the domain. Thus, the function `ssd` estimates the domain means $\bar{Y}_d$, $d = 1, \ldots, D$ by a kind of composite estimators called sample-size dependent estimators. The call to this function is

```
ssd(dom, sweight, domsize, direct, synthetic, delta = 1, data)
```

As mentioned above, the sample-size dependent estimator is obtained by composition of a direct estimator (`direct`) and a synthetic estimator (`synthetic`), both specified by the user, that is,

$$\hat{\bar{Y}}_d^C = \phi_d \hat{\bar{Y}}_d^{DIR} + (1 - \phi_d)\hat{\bar{Y}}_d^{SYN}.$$

The composition weight ϕ_d is obtained as

$$\phi_d = \begin{cases} 1, & \hat{N}_d^{DIR} \geq \delta N_d; \\ \hat{N}_d^{DIR}/(\delta N_d), & \hat{N}_d^{DIR} < \delta N_d, \end{cases}$$

for N_d known (`domsize`) and for a given constant $\delta > 0$ (`delta`). Thus, for a domain with sample size large enough so that the estimated count $\hat{N}_d^{DIR}$ is greater than δN_d, the sample size dependent estimator becomes the direct estimator $\hat{\bar{Y}}_d^{DIR}$. Otherwise, it becomes the composition of the direct and the synthetic estimator $\hat{\bar{Y}}_d^{SYN}$. The constant δ (`delta`) controls how much strength to borrow from other domains by attaching more or less weight to the synthetic estimator, with a large value of δ meaning to borrow more strength.

5 Function eblupFH

Fay-Herriot (FH) models were introduced by Fay & Herriot (1979) to obtain small area estimators of median income in small places in the U.S. These models are well known in the literature of small area estimation (SAE) and are the basic tool when only aggregated auxiliary data at the area level are available. The function `eblupFH` estimates domain characteristics $\delta_d = h_d(\mathbf{y}_d)$, $d = 1, \ldots, D$, based on the mentioned FH model, and the call to this function is

```
eblupFH(formula, vardir, method = "REML", MAXITER = 100,
    PRECISION= 0.0001, data)
```

The basic FH model is defined in two stages. First, since true values δ_d are not observable, our data will be the direct estimates $\hat{\delta}_d^{DIR}$ (left hand side of `formula`). These estimates have an error and this error might be different for each area because samples sizes in the areas are generally different. Thus, in the first stage, we assume the following model representing the error of direct estimates,

$$\hat{\delta}_d^{DIR} = \delta_d + e_d, \quad e_d \overset{ind}{\sim} N(0, \psi_d), \quad d = 1, \ldots, D, \tag{4}$$

and referred to as sampling model, where ψ_d is the sampling variance (`vardir`) of direct estimator $\hat{\delta}_d^{DIR}$ given δ_d, assumed to be known for all d.

In a second stage, true values δ_d are assumed to be linearly related with a vector of auxiliary variables (right hand side of `formula`),

$$\delta_d = \mathbf{x}_d'\boldsymbol{\beta} + v_d, \quad v_d \overset{iid}{\sim} N(0, A), \quad d = 1, \ldots, D, \tag{5}$$

where v_d is independent of e_d, $d = 1, \ldots, D$. This is called the linking model because it links all the areas through the common model parameter $\boldsymbol{\beta}$. Replacing (5) in (4), we obtain

$$\hat{\delta}_d^{DIR} = \mathbf{x}_d'\boldsymbol{\beta} + v_d + e_d, \quad v_d \overset{iid}{\sim} N(0, A), \quad e_d \overset{ind}{\sim} N(0, \psi_d), \quad d = 1, \ldots, D,$$

or equivalently,

$$\hat{\delta}_d^{DIR} \overset{ind}{\sim} N(\mathbf{x}_d'\boldsymbol{\beta}, \psi_d + A), \quad d = 1, \ldots, D. \tag{6}$$

In matrix notation, (6) may be written as $\mathbf{y} \sim N\{\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(A)\}$, where $\mathbf{y} = (\hat{\delta}_1^{DIR}, \ldots, \hat{\delta}_D^{DIR})'$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_D)'$ and $\boldsymbol{\Sigma}(A) = \text{diag}(A + \psi_1, \ldots, A + \psi_D)$. The best linear unbiased predictor (BLUP) of δ_d is given by

$$\begin{aligned} \tilde{\delta}_d &= \hat{\delta}_d^{DIR} - \frac{\psi_d}{A + \psi_d}\left\{\hat{\delta}_d^{DIR} - \mathbf{x}_d'\tilde{\boldsymbol{\beta}}(A)\right\} \\ &= \{1 - B_d(A)\}\hat{\delta}_d^{DIR} + B_d(A)\,\mathbf{x}_d'\tilde{\boldsymbol{\beta}}(A), \end{aligned} \tag{7}$$

where $B_d(A) = \psi_d/(A+\psi_d)$ and $\tilde{\boldsymbol{\beta}}(A)$ is the maximum likelihood (ML) estimator of $\boldsymbol{\beta}$ obtained from (6) and also the weighted least squares (WLS) estimator of $\boldsymbol{\beta}$ without normality assumption, given by

$$\begin{aligned} \tilde{\boldsymbol{\beta}}(A) &= \{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(A)\mathbf{X}\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(A)\mathbf{y} \\ &= \left\{\sum_{d=1}^{D}(A + \psi_d)^{-1}\mathbf{x}_d\mathbf{x}_d'\right\}^{-1}\sum_{d=1}^{D}(A + \psi_d)^{-1}\mathbf{x}_d\hat{\delta}_d^{DIR}. \end{aligned} \tag{8}$$

Expression (7) shows that $\tilde{\delta}_d$ is obtained by composition of the direct estimator $\hat{\delta}_d^{DIR}$ and the regression synthetic estimator $\mathbf{x}_d'\tilde{\boldsymbol{\beta}}(A)$, with more weight attached to the direct estimator when ψ_d is small relative to the total variance $A + \psi_d$, which means that the direct estimator is reliable, and more weight attached to the regression synthetic estimator $\mathbf{x}_d'\boldsymbol{\beta}$ when the direct estimator is not reliable enough and then more strength is required to borrow from the other domains.

Since A is unknown, in practice it is replaced by a consistent estimator $\hat{A}$. Several estimation methods (`method`) for A are considered including a moment estimator called Fay-Herriot (FH) method, maximum likelihood (ML) and restricted (or residual) ML (REML), see the next subsections. Substituting the obtained estimator $\hat{A}$ for A in the BLUP (7), we get the final empirical BLUP (EBLUP) returned by `eblupFH`, and given by

$$\hat{\delta}_d = \tilde{\delta}_d(\hat{A}) = \{1 - B_d(\hat{A})\}\hat{\delta}_d^{DIR} + B_d(\hat{A})\mathbf{x}_d'\hat{\boldsymbol{\beta}}, \tag{9}$$

where $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{A})$.

Function `eblupFH` delivers, together with the estimated model coefficients, i.e. the components of $\hat{\boldsymbol{\beta}}$, their asymptotic standard errors given by the diagonal

elements of the Fisher information depending on the specified estimation method (`method`), the Z statistics obtained by dividing the estimates by their standard errors, and the p-values of the significance tests. Since for large D, the three possible estimators satisfy

$$\hat{\boldsymbol{\beta}} \sim N\left\{\boldsymbol{\beta}, \mathcal{I}^{-1}(\boldsymbol{\beta})\right\},$$

where $\mathcal{I}(\boldsymbol{\beta})$ is the Fisher information, then the Z statistic for a coefficient β_j is

$$Z_j = \hat{\beta}_j / \sqrt{v_{jj}}, \quad j = 1, \ldots, p,$$

where v_{jj} is the estimated asymptotic variance of $\hat{\beta}_j$, given by the j-th element in the diagonal of $\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})$. Finally, for the test

$$H_0 : \beta_j = 0 \quad \text{versus} \quad H_1 : \beta_j \neq 0,$$

p-values are obtained as

$$\text{p-value} = 2\, P(Z > |Z_j|),$$

where Z is a standard normal random variable.

Three different goodness of fit measures are also delivered by function `eblupFH`. The first one is the estimated log-likelihood $\ell(\hat{A}, \hat{\boldsymbol{\beta}}; \mathbf{y})$, obtained by replacing the obtained estimates $\hat{A}$ and $\hat{\boldsymbol{\beta}}$ in (12). The second criteria is the Akaike Information Criteria (AIC), given in this case by

$$\text{AIC} = -2\, \ell(\hat{A}, \hat{\boldsymbol{\beta}}; \mathbf{y}) + 2(p+1).$$

Finally, the Bayesian Information Criteria (BIC) is obtained as

$$\text{BIC} = -2\, \ell(\hat{A}, \hat{\boldsymbol{\beta}}; \mathbf{y}) + (p+1)\log(D).$$

5.1 FH fitting method

FH method gives an estimator of A based on a moments method originally proposed by Fay & Herriot (1979). The FH estimator is given by $\hat{A}_{FH} = \max(0, A^*_{FH})$ with A^*_{FH} obtained iteratively as the solution of the following non-linear equation in A,

$$\sum_{d=1}^{D} (A + \psi_d)^{-1} \{\hat{\delta}_d^{DIR} - \mathbf{x}_d' \tilde{\boldsymbol{\beta}}(A)\}^2 = D - p. \tag{10}$$

This equation is solved using an iterative method such as the Fisher-scoring algorithm. For this, let us define

$$S_{FH}(A) = \sum_{d=1}^{D} (A + \psi_d)^{-1} \{\hat{\delta}_d^{DIR} - \mathbf{x}_d' \tilde{\boldsymbol{\beta}}(A)\}^2 - D - p.$$

By a first order Taylor expansion of $S_{FH}(A_{FH})$ around the true A, we get

$$0 = S_{FH}(A_{FH}) \approx S_{FH}(A) + \frac{\partial S_{FH}(A)}{\partial A}(A^*_{FH} - A). \tag{11}$$

The Fisher-scoring algorithm replaces in this equation, the derivative $\partial S_{FH}(A)/\partial A$ by its expectation $E\{-\partial S_{FH}(A)/\partial A\}$, which in this case is equal to

$$E\left\{-\frac{\partial S_{FH}(A)}{\partial A}\right\} = \sum_{d=1}^{D}(A + \psi_d)^{-1}.$$

Then, solving for A_{FH} in (11), we get

$$A_{FH} = A + \left[E\left\{-\frac{\partial S_{FH}(A)}{\partial A}\right\}\right]^{-1} S_{FH}(A).$$

This algorithm is started taking $A = A^{(0)}_{FH}$, and then in each iteration it updates the estimate of A with the updating equation

$$A^{(k+1)}_{FH} = A^{(k)}_{FH} + \left[E\left\{-\frac{\partial S_{FH}(A)}{\partial A}\right\}\Big|_{A=A^{(k)}_{FH}}\right]^{-1} S_{FH}(A^{(k)}_{FH}).$$

In the function `eblupFH`, the starting value is set to $A^{(0)}_{FH} = \text{median}(\psi_d)$. It stops either when the number of iterations $k >$ MAXITER where MAXITER can be chosen by the user, or when

$$\left|\frac{A^{(k+1)}_{FH} - A^{(k)}_{FH}}{A^{(k)}_{FH}}\right| < \text{PRECISION}.$$

Convergence of the iteration is generally rapid.

5.2 ML fitting method

The model parameters A and $\boldsymbol{\beta}$ can be estimated by ML or REML procedures based on the normal likelihood (6). In fact, under regularity conditions, the estimators derived from these two methods (and using the Normal likelihood) remain consistent at order $O_p(D^{-1/2})$ even without the Normality assumption, for details see Jiang (1996). ML estimators of A and $\boldsymbol{\beta}$ are obtained by maximizing the log-likelihood, given by

$$\ell(A, \boldsymbol{\beta}; \mathbf{y}) = c - \frac{1}{2}\log|\boldsymbol{\Sigma}(A)| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(A)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{12}$$

where c denotes a constant. Taking derivative of ℓ with respect to $\boldsymbol{\beta}$ and equating to zero, we obtain the equation that gives the ML (or WLS) estimator (8). The

ML equation for A is obtained taking derivative of ℓ with respect to A and equating to zero, and is given by

$$\sum_{d=1}^{D}(A+\psi_d)^{-2}\{\hat{\delta}_d^{DIR}-\mathbf{x}_d'\tilde{\boldsymbol{\beta}}(A)\}^2=\sum_{d=1}^{D}(A+\psi_d)^{-1}. \tag{13}$$

Again, Fisher-scoring algorithm is used to solve this equation. The score is defined as $S_{ML}(A)=\partial\ell(A,\boldsymbol{\beta};\mathbf{y})/\partial A$ and is given by

$$S_{ML}(A)=\sum_{d=1}^{D}(A+\psi_d)^{-2}\{\hat{\delta}_d^{DIR}-\mathbf{x}_d'\tilde{\boldsymbol{\beta}}(A)\}^2-\sum_{d=1}^{D}(A+\psi_d)^{-1}.$$

The Fisher information for A is obtained by taking expectation of the negative derivative of $S_{ML}(A)$, and is given by

$$\mathcal{I}_{ML}(A)=E\left\{-\frac{\partial S_{ML}(A)}{\partial A}\right\}=\frac{1}{2}\sum_{d=1}^{D}(A+\psi_d)^{-2}. \tag{14}$$

Finally, the updating equation for the ML estimator of A is

$$A_{ML}^{(k+1)}=A_{ML}^{(k)}+\left\{\mathcal{I}_{ML}(A_{ML}^{(k)})\right\}^{-1}S_{ML}(A_{ML}^{(k)}).$$

Starting value $A_{ML}^{(0)}$ and stopping criterion are the same as in the FH method described above. If A_{ML}^{*} is the estimate obtained in the last iteration of the algorithm, then the final ML estimate returned by `eblupFH` is $\hat{A}_{ML}=\max(0,A_{ML}^{*})$.

5.3 REML fitting method

The REML estimator of A is obtained by maximizing the so called restricted likelihood, which is the joint p.d.f. of a vector of $D-p$ independent contrasts $\mathbf{F}'\mathbf{y}$ of the data $\mathbf{y}$, where $\mathbf{F}$ is an $D\times(D-p)$ full column rank matrix satisfying $\mathbf{F}'\mathbf{F}=\mathbf{I}_{D-p}$ and $\mathbf{F}'\mathbf{X}=\mathbf{0}_{(D-p)\times p}$. The restricted likelihood of $\mathbf{F}'\mathbf{y}$ does not depend on $\boldsymbol{\beta}$, and the log-restricted likelihood is

$$\ell_R(A;\mathbf{y})=c-\frac{1}{2}\log|\mathbf{F}'\boldsymbol{\Sigma}(A)\mathbf{F}|-\frac{1}{2}\mathbf{y}'\mathbf{F}\left\{\mathbf{F}'\boldsymbol{\Sigma}(A)\mathbf{F}\right\}^{-1}\mathbf{F}'\mathbf{y}.$$

It holds that

$$\mathbf{F}\left\{\mathbf{F}'\boldsymbol{\Sigma}(A)\mathbf{F}\right\}^{-1}\mathbf{F}'=\mathbf{P}(A),$$

where

$$\mathbf{P}(A)=\boldsymbol{\Sigma}^{-1}(A)-\boldsymbol{\Sigma}^{-1}(A)\mathbf{X}\left\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(A)\mathbf{X}\right\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(A).$$

Using this relation, we obtain

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

The score is defined as $S_R(A) = \partial \ell_R(A; \mathbf{y})/\partial A$, and is given by

$$\begin{aligned} S_R(A) &= -\frac{1}{2}\text{trace}\{\mathbf{P}(A)\} + \frac{1}{2}\mathbf{y}'\mathbf{P}^2(A)\mathbf{y} \\ &= -\sum_{d=1}^{D}(A+\psi_d)^{-1} - \text{trace}\left[\left\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(A)\mathbf{X}\right\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-2}(A)\mathbf{X}\right] \\ &\quad +\frac{1}{2}\{\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}(A)\}'\boldsymbol{\Sigma}^{-2}(A)\{\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}(A)\} \\ &= \sum_{d=1}^{D}(A+\psi_d)^{-1} - \sum_{d=1}^{D}\frac{\mathbf{x}_d'\left\{\sum_{k=1}^{D}(A+\psi_k)^{-1}\mathbf{x}_k\mathbf{x}_k'\right\}^{-1}\mathbf{x}_d}{(A+\psi_d)^2} \\ &\quad -\sum_{d=1}^{D}\frac{\{\hat{\delta}_d^{DIR}-\mathbf{x}_d'\tilde{\boldsymbol{\beta}}(A)\}^2}{(A+\psi_d)^2}. \end{aligned}$$

The REML estimator of A is obtained by solving the non-linear equation $S_R(A) = 0$. Again, application of Fisher-scoring algorithm requires also the Fisher information for A associated with $\ell_R(A; \mathbf{y})$, which is given by

$$\begin{aligned} \mathcal{I}_R(A) &= E\left\{-\frac{\partial S_R(A)}{\partial A}\right\} = \frac{1}{2}\text{trace}\{\mathbf{P}^2(A)\} \\ &= \frac{1}{2}\text{trace}\{\boldsymbol{\Sigma}(A)^{-2}\} - \text{trace}\left[\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(A)\mathbf{X}\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-3}(A)\mathbf{X}\right] \\ &\quad +\frac{1}{2}\text{trace}\left(\left[\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(A)\mathbf{X}\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-3}(A)\mathbf{X}\right]^2\right). \end{aligned}$$

Finally, the updating equation is

$$A_{REML}^{(k+1)} = A_{REML}^{(k)} + \left\{\mathcal{I}_R(A_{REML}^{(k)})\right\}^{-1} S_R(A_{REML}^{(k)}).$$

Starting value $A_{REML}^{(0)}$ and stopping criterion are the same as in FH and ML methods. Again, if A_{REML}^* is the value obtained in the last iteration, then the REML estimate is finally $\hat{A}_{REML} = \max(0, A_{REML}^*)$.

6 Function mseFH

The accuracy of an EBLUP $\hat{\delta}_d$ is usually assessed by the estimated MSE. Function `mseFH` accompanies the EBLUPs with their corresponding estimated MSEs. The call to this function is

```
mseFH(formula, vardir, method = "REML", MAXITER = 100,
    PRECISION = 0.0001, data)
```

where the arguments are exactly those of function `eblupFH`.

Under model (4)–(5), the MSE of the BLUP for A known is given by

$$\text{MSE}(\tilde{\delta}_d) = E(\tilde{\delta}_d - \delta_d)^2 = g_{1d}(A) + g_{2d}(A),$$

where

$$g_{1d}(A) = \psi_d \{1 - B_d(A)\}, \tag{15}$$

$$g_{2d}(A) = \{B_d(A)\}^2 \mathbf{x}_d' \left\{ \sum_{d=1}^{D} (A + \psi_d)^{-1} \mathbf{x}_d \mathbf{x}_d' \right\}^{-1} \mathbf{x}_d, \tag{16}$$

where $g_{1d}(A)$ is due to the prediction of the random effect v_d and is $O(1)$ for large D, and $g_{2d}(A)$ is due to the estimation of $\boldsymbol{\beta}$ and is $O(D^{-1})$. This means that, for large D, a large reduction in MSE over $MSE(\hat{\delta}_d^{DIR}) = \psi_d$ can be obtained when $1 - B_d(A) = A/(A + \psi_d)$ is small.

Under normality of random effects and errors, the MSE of the EBLUP satisfies

$$\begin{array}{rcccl} \text{MSE}(\hat{\delta}_d) = & \text{MSE}(\hat{\delta}_d) & + & E(\hat{\delta}_d - \tilde{\delta}_d)^2 \\ = & [g_{1d}(A) + g_{2d}(A)] & + & g_{3d}(A) + o(D^{-1}), \end{array} \tag{17}$$

where $g_{3d}(A)$ is the uncertainty arising from the estimation of A, given by

$$g_{3d}(A) = \{B_d(A)\}^2 (A + \psi_d)^{-1} \bar{V}(\hat{A}), \tag{18}$$

where $\bar{V}(\hat{A})$ is the asymptotic variance (as $D \to \infty$) of the estimator $\hat{A}$ of A. Thus, $g_{3d}(A)$ depends on the choice of estimator $\hat{A}$ but for the three available fitting methods FH, ML and REML, this term is $O(D^{-1})$ (Prasad & Rao, 1990).

The MSE of the EBLUP depends on the true variance A, which is unknown. If we want to have an unbiased estimator of the MSE up to $o(D^{-1})$ terms (or second order unbiased), the MSE estimator depends on the method used to find $\hat{A}$ in the EBLUP (`method`). The following subsections describe the MSE estimates returned by `mseFH` for each selected fitting method.

6.1 FH fitting method

When using the FH estimator $\hat{A}_{FH}$, a second order unbiased estimator of $\text{MSE}(\hat{\delta}_d)$ using $\hat{A}_{FH}$ is given by

$$mse_{FH}(\hat{\delta}_d) = g_{1d}(\hat{A}_{FH}) + g_{2d}(\hat{A}_{FH}) + 2\, g_{3d}(\hat{A}_{FH}) - b_{FH}(\hat{A}_{FH})\{B_d(\hat{A}_{FH})\}^2, \tag{19}$$

where, in $g_{3d}(\hat{A}_{FH})$, the asymptotic variance is

$$\bar{V}(\hat{A}_{FH}) = \frac{2D}{\left\{ \sum_{d=1}^{D} (A + \psi_d)^{-1} \right\}^2} \tag{20}$$

and

$$b_{FH}(A) = \frac{2 \left[D \sum_{d=1}^{D} (A + \psi_d)^{-2} - \left\{ \sum_{d=1}^{D} (A + \psi_d)^{-1} \right\}^2 \right]}{\left\{ \sum_{d=1}^{D} (A + \psi_d)^{-1} \right\}^3}, \tag{21}$$

see Datta, Rao & Smith (2005).

6.2 ML fitting method

When using the ML estimator $\hat{A}_{ML}$ of A, a second order unbiased estimator of MSE($\hat{\delta}_d$) was obtained by Datta & Lahiri (2000) and is given by

$$mse_{ML}(\hat{\delta}_d) = g_{1d}(\hat{A}_{ML}) + g_{2d}(\hat{A}_{ML}) + 2g_{3d}(\hat{A}_{ML}) - b_{ML}(\hat{A}_{ML}) \bigtriangledown g_{1d}(\hat{A}_{ML}), \quad (22)$$

where the asymptotic variance involved in $g_{3d}(A)$ is equal to the inverse of the Fisher information given in (14),

$$\bar{V}(\hat{A}_{ML}) = \mathcal{I}_{ML}^{-1}(A) = \frac{2}{\sum_{d=1}^{D}(A+\psi_d)^{-2}}, \quad (23)$$

$$b_{ML}(A) = -\frac{\text{trace}\left[\left\{\sum_{\ell=1}^{D}(A+\psi_\ell)^{-1}\mathbf{x}_\ell\mathbf{x}_\ell'\right\}^{-1}\left\{\sum_{\ell=1}^{D}(A+\psi_\ell)^{-2}\mathbf{x}_\ell\mathbf{x}_\ell'\right\}\right]}{\sum_{\ell=1}^{D}(A+\psi_\ell)^{-2}}$$

and

$$\bigtriangledown g_{1d}(A) = \frac{\partial g_{1d}(A)}{\partial A} = \left\{\frac{\psi_d}{A+\psi_d}\right\}^2.$$

6.3 REML fitting method

When using the REML estimator $\hat{A}_{REML}$, a second order unbiased estimator of MSE($\hat{\delta}_d$) is given by

$$mse_{REML}(\hat{\delta}_d) = g_{1d}(\hat{A}_{REML}) + g_{2d}(\hat{A}_{REML}) + 2g_{3d}(\hat{A}_{REML}), \quad (24)$$

where $\bar{V}(\hat{A}_{REML}) = \bar{V}(\hat{A}_{ML})$ is given in (23), see Datta & Lahiri (2000).

7 Function eblupSFH

Function `eblupSFH` estimates domain parameters $\delta_d = h_d(\mathbf{y}_d)$, $d = 1, \ldots, D$, based on a FH model with spatially correlated area effects. The call to the function is

```
eblupSFH(formula, vardir, proxmat, method = "REML", MAXITER = 100,
        PRECISION = 0.0001, data)
```

The model considered by this function is, in matrix notation,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} + \mathbf{e}, \quad (25)$$

where $\mathbf{y} = (\hat{\delta}_1^{DIR}, \ldots, \hat{\delta}_D^{DIR})'$ is the vector of direct estimates for the D small areas (left hand side of `formula`), $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_D)'$ is a $D \times p$ matrix containing in its columns the values of p explanatory variables for the D areas (right hand side of `formula`), $\mathbf{v} = (v_1, \ldots, v_D)'$ is the vector of area effects and $\mathbf{e} = (e_1, \ldots, e_D)'$ is the vector of independent sampling errors, independent of $\mathbf{v}$, with $\mathbf{e} \sim N(\mathbf{0}_D, \boldsymbol{\Psi})$,

where the covariance matrix $\boldsymbol{\Psi} = \text{diag}(\psi_1, \ldots, \psi_D)$ is known (`vardir` contains the diagonal elements). The vector $\boldsymbol{\delta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} = (\delta_1, \ldots, \delta_D)'$ collects the target domain parameters.

The vector $\mathbf{v}$ follows an simultaneously autoregressive (SAR) process with unknown autoregression parameter $\rho \in (-1, 1)$ and proximity matrix $\mathbf{W}$, i.e.,

$$\mathbf{v} = \rho \mathbf{W}\mathbf{v} + \mathbf{u}, \tag{26}$$

see Anselin (1988) and Cressie (1993). Model (25)–(26) will be called hereafter spatial FH (SFH) model. We assume that the matrix $(\mathbf{I}_D - \rho\mathbf{W})$ is non-singular, where $\mathbf{I}_D$ denotes the $D \times D$ identity matrix. Then $\mathbf{v}$ can be expressed as

$$\mathbf{v} = (\mathbf{I}_D - \rho\mathbf{W})^{-1}\mathbf{u}, \tag{27}$$

where $\mathbf{u} = (u_1, \ldots, u_D)'$ satisfies $\mathbf{u} \sim N(\mathbf{0}_D, A\,\mathbf{I}_D)$ for A unknown.

The matrix $\mathbf{W}$ (`proxmat`) is obtained from an original proximity matrix $\mathbf{W}^0$, whose diagonal elements area equal to zero and the remaining entries are equal to 1 when the two areas corresponding to the row and the column indices are considered as neighbor and zero otherwise. Then $\mathbf{W}$ is obtained by row-standardization of $\mathbf{W}^0$, obtained by dividing each entry of $\mathbf{W}^0$ by the sum of elements in the same row, see Anselin (1988), Cressie (1993) and Banerjee, Carlin & Gelfand (2004) for more details on the SAR(1) process with the above parametrization. When $\mathbf{W}$ is defined in this fashion, ρ is called spatial autocorrelation parameter (Banerjee, Carlin & Gelfand, 2004). Hereafter, the vector of variance components will be denoted $\boldsymbol{\theta} = (\theta_1, \theta_2)' = (A, \rho)'$. Equation (27) implies that $\mathbf{v}$ has mean vector $\mathbf{0}$ and covariance matrix equal to

$$\mathbf{G}(\boldsymbol{\theta}) = A\,\{(\mathbf{I}_D - \rho\mathbf{W})'(\mathbf{I}_D - \rho\mathbf{W})\}^{-1}. \tag{28}$$

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

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta}) + \boldsymbol{\Psi}.$$

Combining (25) and (27), the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_D - \rho\mathbf{W})^{-1}\mathbf{u} + \mathbf{e} \tag{29}$$

The BLUP of $\delta_d = \mathbf{x}_d'\boldsymbol{\beta} + v_d$ under model (29) is called Spatial BLUP (Petrucci & Salvati, 2006) and is given by

$$\tilde{\delta}_d(\boldsymbol{\theta}) = \mathbf{x}_d'\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \mathbf{b}_d'\mathbf{G}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\{\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})\}, \tag{30}$$

where $\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{y}$ is the WLS estimator of $\boldsymbol{\beta}$ and $\mathbf{b}_d'$ is the $1 \times d$ vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the d-th position. The Spatial BLUPs $\tilde{\delta}_d(\boldsymbol{\theta})$, $d = 1, \ldots, D$, depend on the unknown vector of variance components $\boldsymbol{\theta} = (A, \rho)'$. Replacing a consistent estimator $\hat{\boldsymbol{\theta}} = (\hat{A}, \hat{\rho})'$ for $\boldsymbol{\theta}$ in (30), we obtain the Spatial EBLUPs $\hat{\delta}_d = \tilde{\delta}_d(\hat{\boldsymbol{\theta}})$, $d = 1, \ldots, D$, returned by function `eblupSFH`. The following subsections describe the two fitting methods (`method`) for the SFH model (25)–(26) supported by `eblupSFH`.

7.1 ML fitting method

The ML estimator of $\boldsymbol{\theta} = (A, \rho)'$ is obtained by maximizing the log-likelihood of $\boldsymbol{\theta}$ given the data vector $\mathbf{y}$,

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

where c denotes a constant. The Fisher-scoring iterative algorithm is applied to maximize this log-likelihood. Let $\mathbf{S}(\boldsymbol{\theta}) = (S_A(\boldsymbol{\theta}), S_\rho(\boldsymbol{\theta}))'$ be the scores or derivatives of the log-likelihood with respect to A and ρ, and let $\mathcal{I}(\boldsymbol{\theta})$ be the Fisher information matrix obtained from $\ell(\boldsymbol{\theta}; \mathbf{y})$, with elements

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{A,A}(\boldsymbol{\theta}) & \mathcal{I}_{A,\rho}(\boldsymbol{\theta}) \\ \mathcal{I}_{\rho,A}(\boldsymbol{\theta}) & \mathcal{I}_{\rho,\rho}(\boldsymbol{\theta}) \end{pmatrix}. \tag{31}$$

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

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

The ML equation for $\boldsymbol{\beta}$ yields

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left\{ \mathbf{X}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{X} \right\}^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{y}. \tag{32}$$

Let us denote

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

and

$$\mathbf{P}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{X} \left\{ \mathbf{X}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{X} \right\}^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}). \tag{33}$$

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

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

and the derivatives of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ with respect to A and ρ are respectively given by

$$\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial A} = \mathbf{C}^{-1}(\rho), \quad \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \rho} = -A\, \mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} \mathbf{C}^{-1}(\rho) \triangleq \mathbf{R}(\boldsymbol{\theta}).$$

The scores associated to A and ρ, after replacing (32), are given by

$$\begin{aligned} S_A(\boldsymbol{\theta}) &= -\frac{1}{2} \operatorname{trace} \left\{ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{C}^{-1}(\rho) \right\} + \frac{1}{2} \mathbf{y}' \mathbf{P}(\boldsymbol{\theta}) \mathbf{C}^{-1}(\rho) \mathbf{P}(\boldsymbol{\theta}) \mathbf{y}, \\ S_\rho(\boldsymbol{\theta}) &= -\frac{1}{2} \operatorname{trace} \left\{ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{R}^{-1}(\boldsymbol{\theta}) \right\} + \frac{1}{2} \mathbf{y}' \mathbf{P}(\boldsymbol{\theta}) \mathbf{R}(\boldsymbol{\theta}) \mathbf{P}(\boldsymbol{\theta}) \mathbf{y}. \end{aligned}$$

The elements of the Fisher information matrix are

$$\mathcal{I}_{A,A}(\boldsymbol{\theta}) = \frac{1}{2}\,\text{trace}\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\right\}, \tag{34}$$

$$\mathcal{I}_{A,\rho}(\boldsymbol{\theta}) = \mathcal{I}_{\rho,A} = \frac{1}{2}\,\text{trace}\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\right\}, \tag{35}$$

$$\mathcal{I}_{\rho,\rho}(\boldsymbol{\theta}) = \frac{1}{2}\,\text{trace}\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\right\}. \tag{36}$$

In the function `eblupSFH`, the starting value of A is set to $A^{(0)} = \text{median}(\psi_d)$. For ρ, we take $\rho^{(0)} = 0.5$. The algorithm stops either when the number of iterations $k >$ MAXITER where MAXITER can be chosen by the user, or when

$$\max\left\{\left|\frac{\sigma_u^{2(k+1)} - \sigma_u^{2(k)}}{\sigma_u^{2(k)}}\right|, \left|\frac{\rho^{(k+1)} - \rho^{(k)}}{\rho^{(k)}}\right|\right\} < \text{PRECISION}.$$

7.2 REML fitting method

The REML estimator of $\boldsymbol{\theta}$ is obtained by maximizing the restricted likelihood defined as in Section 5.3. Under the SFH model, the restricted log-likelihood is given by

$$\ell_R(\boldsymbol{\theta};\mathbf{y}) = c - \frac{1}{2}\log|\mathbf{F}'\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{F}| - \frac{1}{2}\mathbf{y}'\mathbf{P}(\boldsymbol{\theta})\mathbf{y},$$

where $\mathbf{P}(\boldsymbol{\theta})$ is defined in (33). Using the following properties of the matrix $\mathbf{P}(\boldsymbol{\theta})$,

$$\mathbf{P}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta}) = \mathbf{P}(\boldsymbol{\theta}), \quad \frac{\partial\mathbf{P}(\boldsymbol{\theta})}{\partial\theta_r} = -\mathbf{P}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_r}\mathbf{P}(\boldsymbol{\theta}),$$

we obtain the scores or derivatives of $\ell_R(\boldsymbol{\theta};\mathbf{y})$ with respect to each element of $\boldsymbol{\theta}$,

$$S_A^R(\boldsymbol{\theta}) = -\frac{1}{2}\,\text{trace}\left\{\mathbf{P}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\right\} + \frac{1}{2}\mathbf{y}'\mathbf{P}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\mathbf{P}(\boldsymbol{\theta})\mathbf{y},$$
$$S_\rho^R(\boldsymbol{\theta}) = -\frac{1}{2}\,\text{trace}\left\{\mathbf{P}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\right\} + \frac{1}{2}\mathbf{y}'\mathbf{P}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{y},$$

Finally, the elements of the Fisher information obtained from $\ell_R(\boldsymbol{\theta};\mathbf{y})$ are

$$\mathcal{I}_{A,A}^R(\boldsymbol{\theta}) = \frac{1}{2}\,\text{tr}\{\mathbf{P}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\mathbf{P}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\},$$
$$\mathcal{I}_{A,\rho}^R(\boldsymbol{\theta}) = \mathcal{I}_{\rho,A}^R = \frac{1}{2}\,\text{tr}\{\mathbf{P}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\},$$
$$\mathcal{I}_{\rho,\rho}^R(\boldsymbol{\theta}) = \frac{1}{2}\,\text{tr}\{\mathbf{P}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\}.$$

The updating equation of the Fisher-scoring algorithm is then

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \left\{\mathcal{I}^R(\boldsymbol{\theta}^{(k)})\right\}^{-1}\mathbf{S}^R(\boldsymbol{\theta}^{(k)}),$$

where $\mathbf{S}^R(\boldsymbol{\theta}) = (S_A^R(\boldsymbol{\theta}), S_\rho^R(\boldsymbol{\theta}))'$ is the scores vector and

$$\mathcal{I}^R(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}^R_{A,A}(\boldsymbol{\theta}) & \mathcal{I}^R_{A,\rho}(\boldsymbol{\theta}) \\ \mathcal{I}^R_{\rho,A}(\boldsymbol{\theta}) & \mathcal{I}^R_{\rho,\rho}(\boldsymbol{\theta}) \end{pmatrix}. \tag{37}$$

is the Fisher information matrix. Starting values and stopping criterion are set the same as in the case of ML estimates.

8 Function mseSFH

Function `mseSFH` gives estimated MSEs of the Spatial EBLUPs under the SFH model (25)–(26). The call to the function is

```
mseSFH(formula, vardir, proxmat, method = "REML", MAXITER = 100,
      PRECISION = 0.0001, data)
```

which has the same arguments as function `eblupSFH`.

Similarly as in Section 6, under normality of random effects and errors, the MSE of the Spatial EBLUP can be decomposed as

$$\begin{array}{rcl} \mathrm{MSE}\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}})\} = & \mathrm{MSE}\{\tilde{\delta}_d(\boldsymbol{\theta})\} & + \quad E\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}}) - \tilde{\delta}_d(\boldsymbol{\theta})\}^2 \\ = & [g_{1d}(\boldsymbol{\theta}) + g_{2d}(\boldsymbol{\theta})] & + \quad g_{3d}(\boldsymbol{\theta}), \end{array} \tag{38}$$

where the first two terms on the right hand side are easily calculated due to the linearity of the Spatial BLUP $\tilde{\delta}_d(\boldsymbol{\theta})$ in the data vector $\mathbf{y}$. They are given by

$$g_{1d}(\boldsymbol{\theta}) = \mathbf{b}_d' \left\{ \mathbf{G}(\boldsymbol{\theta}) - \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta}) \right\} \mathbf{b}_d, \tag{39}$$

$$\begin{aligned} g_{2d}(\boldsymbol{\theta}) = {} & \mathbf{b}_d' \left\{ \mathbf{I}_D - \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \right\} \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X})^{-1}\mathbf{X}' \\ & \times \left\{ \mathbf{I}_D - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta}) \right\} \mathbf{b}_d. \end{aligned} \tag{40}$$

However, for the last term $g_{3d}(\boldsymbol{\theta}) = E\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}}) - \tilde{\delta}_d(\boldsymbol{\theta})\}^2$, an exact analytical expression does not exist due to the non-linearity of the EBLUP $\tilde{\delta}_d(\hat{\boldsymbol{\theta}})$ in $\mathbf{y}$. Under the basic FH model (4)-(5) with independent random effects v_d (diagonal covariance matrix $\boldsymbol{\Sigma}$), Prasad & Rao (1990) obtained an approximation up to $o(D^{-1})$ terms of $g_{3d}(\boldsymbol{\theta})$ through Taylor linearization, see Section 6. Their formula can be taken as a naive approximation of the true $g_{3d}(\boldsymbol{\theta})$ under the SFH model (25)–(26). Straightforward application of this formula to model (25)–(26) yields

$$g_{3d}^{PR}(\boldsymbol{\theta}) = \mathrm{trace}\left\{ \mathbf{L}_d(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{L}_d'(\boldsymbol{\theta})\mathcal{I}^{-1}(\boldsymbol{\theta}) \right\}, \tag{41}$$

where $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information defined in (31) with elements (34)–(36) and

$$\mathbf{L}_d(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{b}_d' \{ \mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - A\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \} \\ \mathbf{b}_d' \{ \mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - A\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \} \end{pmatrix}.$$

Then the full MSE can be approximated by

$$\mathrm{MSE}^{PR}\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}})\} = g_{1d}(\boldsymbol{\theta}) + g_{2d}(\boldsymbol{\theta}) + g_{3d}^{PR}(\boldsymbol{\theta}). \tag{42}$$

Singh, Shukla & Kundu (2005) arrived to the same formula (42) for the true MSE under a SFH model. However, this formula is not accounting for the extra uncertainty due to estimation of the spatial autocorrelation parameter ρ. Next subsections describe the MSE estimates returned by function `mseSFH`, depending on the specified fitting method.

8.1 REML fitting method

Let $\hat{\boldsymbol{\theta}}_{REML}$ be the estimator of $\hat{\boldsymbol{\theta}}$ when REML fitting method is specified in `mseSFH`. In that case, the function `mseSFH` returns the MSE estimator derived by Singh, Shukla & Kundu (2005) and given by

$$\text{mse}^{SSK}[\tilde{\delta}_d(\hat{\boldsymbol{\theta}}_{REML})] = g_{1d}(\hat{\boldsymbol{\theta}}_{REML}) + g_{2d}(\hat{\boldsymbol{\theta}}_{REML}) + 2g_{3d}^{PR}(\hat{\boldsymbol{\theta}}_{REML}) - g_{4d}(\hat{\boldsymbol{\theta}}_{REML}), \tag{43}$$

where g_{1d}, g_{2d} and g_{3d}^{PR} are given respectively in (39), (40) and (41), and $g_{4d}(\boldsymbol{\theta})$ reads

$$g_{4d}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{r=1}^{2}\sum_{s=1}^{2} \mathbf{b}_d' \boldsymbol{\Psi}\, \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_r \partial\theta_s} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Psi}\, v_{rs}(\boldsymbol{\theta})\, \mathbf{b}_d,$$

where $v_{rs}(\boldsymbol{\theta})$ denotes the element (r,s) of $\mathcal{I}(\boldsymbol{\theta})^{-1}$, for the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$ defined in (31) with elements (34)–(36). The second order derivatives of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are given by

$$\begin{aligned}
&\frac{\partial^2 \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial A^2} = \mathbf{0}_{D\times D},\\
&\frac{\partial^2 \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial A \partial\rho} = \frac{\partial^2 \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial A \partial\rho} = -\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\rho)}{\partial\rho} \mathbf{C}^{-1}(\boldsymbol{\theta}),\\
&\frac{\partial^2 \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \partial\rho^2} = 2A\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\rho)}{\partial\rho} \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\rho)}{\partial\rho} \mathbf{C}^{-1}(\boldsymbol{\theta}) - 2A\mathbf{C}^{-1}(\boldsymbol{\theta}) \mathbf{W}'\mathbf{W} \mathbf{C}^{-1}(\boldsymbol{\theta}).
\end{aligned}$$

8.2 ML fitting method

Let $\hat{\boldsymbol{\theta}}_{ML}$ be the estimator of $\hat{\boldsymbol{\theta}}$ when ML fitting method is specified. In that case, the estimate returned by `mseSFH` reads

$$\begin{aligned}
\text{mse}_{ML}^{SSK}\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}})\} \;=\; & g_{1d}(\hat{\boldsymbol{\theta}}_{ML}) + g_{2d}(\hat{\boldsymbol{\theta}}_{ML}) + 2g_{3d}^{PR}(\hat{\boldsymbol{\theta}}_{ML}) - g_{4d}(\hat{\boldsymbol{\theta}}_{ML})\\
& -\mathbf{b}_{ML}'(\hat{\boldsymbol{\theta}}_{ML}) \nabla g_{1d}(\hat{\boldsymbol{\theta}}_{ML}).
\end{aligned}$$

In this expression, g_{1d}, g_{2d} and g_{3d}^{PR} are defined respectively in (39), (40) and (41), $\nabla g_{1d}(\boldsymbol{\theta}) = (\partial g_{1d}(\boldsymbol{\theta})/\partial A, \partial g_{1d}(\boldsymbol{\theta})/\partial\rho)'$ is the gradient of $g_{1d}(\boldsymbol{\theta})$ with derivatives

$$\begin{aligned}
\frac{\partial g_{1d}(\boldsymbol{\theta})}{\partial A} \;=\; & \mathbf{b}_d' \big\{ \mathbf{C}^{-1}(\rho) - 2\,A\,\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\\
& + A^2\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho) \big\}\, \mathbf{b}_d,\\
\frac{\partial g_{1d}(\boldsymbol{\theta})}{\partial \rho} \;=\; & \mathbf{b}_d' \big\{ \mathbf{R}(\boldsymbol{\theta}) - 2\,A\,\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\\
& A^2\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho) \big\}\, \mathbf{b}_d,
\end{aligned}$$

and finally, $\mathbf{b}_{ML}(\hat{\boldsymbol{\theta}}_{ML})$ is the bias of the ML estimator $\hat{\boldsymbol{\theta}}_{ML}$ up to $o(D^{-1})$ terms, given by $\mathbf{b}_{ML}(\hat{\boldsymbol{\theta}}) = \mathcal{I}^{-1}(\hat{\boldsymbol{\theta}}_{ML})\mathbf{h}(\hat{\boldsymbol{\theta}}_{ML})/2$ with $\mathbf{h}(\hat{\boldsymbol{\theta}}_{ML}) = (h_1(\hat{\boldsymbol{\theta}}_{ML}), h_2(\hat{\boldsymbol{\theta}}_{ML}))'$ and

$$h_1(\boldsymbol{\theta}) = -\text{trace}\left[\left\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{C}^{-1}(\rho)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right],$$
$$h_2(\boldsymbol{\theta}) = -\text{trace}\left[\left\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right].$$

9 Function pbmseSFH

Function `pbmseSFH` gives parametric bootstrap estimates of the MSEs of the Spatial EBLUPs under the SFH model (25)–(26) using an extension of González-Manteiga et al. (2008b). The MSE estimators obtained by this procedure are expected to be consistent if the model parameter estimates are consistent. The call to the function is

```
pbmseSFH(formula, vardir, proxmat, B = 100, method = "REML",
    MAXITER = 100, PRECISION = 0.0001, data)
```

where the arguments are those of function `eblupSFH`, and additionally the number of bootstrap replicates `B`. The bootstrap procedure proceeds as follows:

1) Fit the SFH model (25)–(26) to the initial data $\mathbf{y} = (\hat{\delta}_1^{DIR}, \ldots, \hat{\delta}_D^{DIR})'$, obtaining model parameter estimates $\hat{\boldsymbol{\theta}} = (\hat{A}, \hat{\rho})'$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$.

2) Generate a vector $\mathbf{t}_1^*$ whose elements are D independent copies of a $N(0,1)$. Construct bootstrap vectors $\mathbf{u}^* = \hat{A}^{1/2}\,\mathbf{t}_1^*$ and $\mathbf{v}^* = (\mathbf{I}_D - \hat{\rho}\mathbf{W})^{-1}\mathbf{u}^*$, and calculate the bootstrap quantity of interest $\boldsymbol{\delta}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{v}^*$, by regarding $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ as the true values of the model parameters.

3) Generate a vector $\mathbf{t}_2^*$ with D independent copies of a $N(0,1)$, independently of the generation of $\mathbf{t}_1^*$, and construct the vector of bootstrap random errors $\mathbf{e}^* = \boldsymbol{\Psi}^{1/2}\,\mathbf{t}_2^*$.

4) Obtain bootstrap data applying the model $\mathbf{y}^* = \boldsymbol{\delta}^* + \mathbf{e}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{v}^* + \mathbf{e}^*$.

5) Regarding $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ as the true values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, fit the SFH model (25)–(26) to bootstrap data $\mathbf{y}^*$, obtaining estimates of the "true" $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ based on $\mathbf{y}^*$. For this, first calculate the estimator of $\hat{\boldsymbol{\beta}}$ evaluated at the "true" value $\hat{\boldsymbol{\theta}}$,

$$\tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\theta}}) = \left\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{X}\right\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{y}^*;$$

next, obtain the estimator $\hat{\boldsymbol{\theta}}^*$ of $\hat{\boldsymbol{\theta}}$ based on $\mathbf{y}^*$ and, finally, the estimator of $\hat{\boldsymbol{\beta}}$ evaluated at $\hat{\boldsymbol{\theta}}^*$, that is, $\tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\theta}}^*)$.

6) Calculate the bootstrap Spatial BLUP from bootstrap data $\mathbf{y}^*$ and regarding $\hat{\boldsymbol{\theta}}$ as the true value of $\boldsymbol{\theta}$,

$$\tilde{\delta}_d^*(\hat{\boldsymbol{\theta}}) = \mathbf{x}_d'\tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\theta}}) + \mathbf{b}_d'\mathbf{G}(\hat{\boldsymbol{\theta}})\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}\left\{\mathbf{y}^* - \mathbf{X}\tilde{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\theta}})\right\}.$$

Calculate also the bootstrap Spatial EBLUP replacing $\hat{\boldsymbol{\theta}}^*$ for the "true" $\hat{\boldsymbol{\theta}}$,

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

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

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

$$g_{3d}^{PB}(\hat{\boldsymbol{\theta}}) = B^{-1}\sum_{b=1}^{B}\left\{\tilde{\delta}_d^{*(b)}(\hat{\boldsymbol{\theta}}^{*(b)}) - \tilde{\delta}_d^{*(b)}(\hat{\boldsymbol{\theta}})\right\}^2.$$

Function `pbmseSFH` returns the naive parametric bootstrap estimator of the full MSE, given by

$$mse^{naPB}\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}})\} = B^{-1}\sum_{b=1}^{B}\left\{\tilde{\delta}_d^{*(b)}(\hat{\boldsymbol{\theta}}^{*(b)}) - \delta_d^{*(b)}\right\}^2. \tag{44}$$

Function `pbmseSFH` also returns a bias-corrected MSE estimate obtained as in Pfeffermann and Tiller (2005), and given by

$$\begin{aligned} mse^{bcPB}\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}})\} &= 2\left\{g_{1d}(\hat{\boldsymbol{\theta}}) + g_{2d}(\hat{\boldsymbol{\theta}})\right\} + g_{3d}^{PB}(\hat{\boldsymbol{\theta}}) \\ &\quad - B^{-1}\sum_{b=1}^{B}\left\{g_{1d}(\hat{\boldsymbol{\theta}}^{*(b)}) + g_{2d}(\hat{\boldsymbol{\theta}}^{*(b)})\right\}. \end{aligned} \tag{45}$$

10 Function npbmseSFH

Function `npbmseSFH` gives MSE estimates for the Spatial EBLUPs under the SFH model (25)–(26), using the nonparametric bootstrap approach of Molina, Salvati & Pratesi (2009). The call to the function is

```
npbmseSFH(formula, vardir, proxmat, B = 100, method = "REML",
   MAXITER = 100,PRECISION = 0.0001, data)
```

where the arguments are the same as in `pbmseSFH`. The function resamples random effects $\{u_1^*, \ldots, u_D^*\}$ and errors $\{e_1^*, \ldots, e_D^*\}$ from the respective empirical distribution of predicted random effects $\{\hat{u}_1, \ldots, \hat{u}_D\}$ and residuals $\{\hat{r}_1, \ldots, \hat{r}_D\}$, where $r_d = \hat{\delta}_d^{DIR} - \tilde{\delta}_d(\hat{\boldsymbol{\theta}})$, $d = 1, \ldots, D$, all previously standardized. This method avoids the need for distributional assumptions of u_d and e_d; therefore, it is expected to be more robust to non-normality of the random model components.

Under model (25)–(26), the BLUPs of $\mathbf{u}$ and $\mathbf{v}$ are respectively given by

$$\tilde{\mathbf{v}}(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\{\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})\}, \quad \tilde{\mathbf{u}}(\boldsymbol{\theta}) = (\mathbf{I} - \rho\mathbf{W})\tilde{\mathbf{v}}(\boldsymbol{\theta}),$$

and the covariance matrix of $\tilde{\mathbf{u}}(\boldsymbol{\theta})$ is

$$\boldsymbol{\Sigma}_{\mathbf{u}}(\boldsymbol{\theta}) = (\mathbf{I} - \rho\mathbf{W})\mathbf{G}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta})(\mathbf{I} - \rho\mathbf{W}').$$

Let us define the vector of residuals obtained from the BLUP

$$\tilde{\mathbf{r}}(\boldsymbol{\theta}) = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \tilde{\mathbf{v}}(\boldsymbol{\theta}) = (\hat{\delta}_1^{DIR} - \tilde{\delta}_1(\boldsymbol{\theta}), \ldots, \hat{\delta}_D^{DIR} - \tilde{\delta}_d(\boldsymbol{\theta}))'.$$

It is easy to see that the covariance matrix of $\tilde{\mathbf{r}}(\boldsymbol{\theta})$ is

$$\boldsymbol{\Sigma}_{\mathbf{r}}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\,\mathbf{P}(\boldsymbol{\theta})\boldsymbol{\Psi},$$

for $\mathbf{P}(\boldsymbol{\theta})$ defined in (33). The covariance matrices $\boldsymbol{\Sigma}_{\mathbf{u}}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_{\mathbf{r}}(\boldsymbol{\theta})$ are not diagonal; hence, the elements of the vectors $\tilde{\mathbf{u}}(\boldsymbol{\theta})$ and $\tilde{\mathbf{r}}(\boldsymbol{\theta})$ are correlated. Indeed, both $\tilde{\mathbf{u}}(\boldsymbol{\theta})$ and $\tilde{\mathbf{r}}(\boldsymbol{\theta})$ lie in a subspace of dimension $D-p$. Since the methods that resample from the empirical distribution work well under an ideally iid setup, before resampling a previous standardization step is crucial. Here $\hat{\mathbf{u}} = \tilde{\mathbf{u}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{r}} = \tilde{\mathbf{r}}(\hat{\boldsymbol{\theta}})$ are transformed to achieve vectors that are as close as possible to be uncorrelated and with unit variance elements. We describe the standardization method only for $\hat{\mathbf{u}}$, since for $\hat{\mathbf{r}}$ the process is analogous. Let us consider the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}} = \boldsymbol{\Sigma}_{\mathbf{u}}(\hat{\boldsymbol{\theta}})$. The spectral decomposition of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$ is

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{u}} = \mathbf{Q}_{\mathbf{u}}\boldsymbol{\Delta}_{\mathbf{u}}\mathbf{Q}_{\mathbf{u}}',$$

where $\boldsymbol{\Delta}_{\mathbf{u}}$ is a diagonal matrix with the $D-p$ non-zero eigenvalues of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$ and $\mathbf{Q}_{\mathbf{u}}$ is the matrix with the corresponding eigenvectors in the columns. Take the square root matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1/2} = \mathbf{Q}_{\mathbf{u}}\boldsymbol{\Delta}_{\mathbf{u}}^{-1/2}\mathbf{Q}_{\mathbf{u}}'$. Squaring this matrix gives a generalized inverse of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$. With the obtained square root, we transform $\hat{\mathbf{u}}$ as

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

The covariance matrix of $\hat{\mathbf{u}}^S$ is then $V(\hat{\mathbf{u}}^S) = \mathbf{Q}_{\mathbf{u}}\mathbf{Q}_{\mathbf{u}}'$, which is close to an identity matrix. Observe that in the transformation

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

the vector $\mathbf{Q}_{\mathbf{u}}'\hat{\mathbf{u}}$ contains the coordinates of $\hat{\mathbf{u}}$ in its principal components, which are uncorrelated and with covariance matrix $\boldsymbol{\Delta}_{\mathbf{u}}$. Then multiplying by $\boldsymbol{\Delta}_{\mathbf{u}}^{-1/2}$, these coordinates are standardized to have unit variance. Finally, this standardized vector in the space of the principal components is returned to the original space by multiplying by $\mathbf{Q}_{\mathbf{u}}$. Thus, the transformed vector $\hat{\mathbf{u}}^S$ contains the coordinates of the vector $\boldsymbol{\Delta}_{\mathbf{u}}^{-1/2}\mathbf{Q}_{\mathbf{u}}'\hat{\mathbf{u}}$, with standard elements, in the original space. The eigenvalues, which are the variances of the uncorrelated principal components, collect better the variability than the diagonals of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$. Indeed, simulations were indicated that taking simply $\hat{u}_d^S = \hat{u}_d/\sqrt{v_{dd}}$, where v_{dd} is the d-th diagonal element of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$, does not work well.

The final nonparametric bootstrap procedure is obtained by replacing steps 2) and 3) in the parametric bootstrap 1)–8) of Section 9 by the new steps 2') and 3') given below:

2') With the estimates $\hat{\boldsymbol{\theta}} = (\hat{A}, \hat{\rho})'$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ obtained in step 1), calculate predictors of $\mathbf{v}$ and $\mathbf{u}$ as follows

$$\hat{\mathbf{v}} = \mathbf{G}(\hat{\boldsymbol{\theta}})\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad \hat{\mathbf{u}} = (\mathbf{I} - \hat{\rho}\mathbf{W})\hat{\mathbf{v}} = (\hat{u}_1, \ldots, \hat{u}_m)'.$$

Then take $\hat{\mathbf{u}}^S = \hat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1/2}\hat{\mathbf{u}} = (\hat{u}_1^S, \ldots, \hat{u}_D^S)'$, where $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{1/2}$ is the square root of the generalized inverse of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$ obtained by the spectral decomposition. It is convenient to re-scale the elements $\hat{u}_d^S$ so that they have sample mean exactly equal to zero and sample variance $\hat{A}$. This is achieved by the transformation

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

Construct the vector $\mathbf{u}^* = (u_1^*, \ldots, u_D^*)'$, whose elements are obtained by extracting a simple random sample with replacement of size D from the set $\{\hat{u}_1^{SS}, \ldots, \hat{u}_D^{SS}\}$. Then obtain $\mathbf{v}^* = (\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{u}^*$ and calculate the vector of bootstrap target parameters $\boldsymbol{\delta}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{v}^* = (\delta_1^*, \ldots, \delta_d^*)'$

3') Compute the vector of residuals $\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\mathbf{v}} = (\hat{r}_1, \ldots, \hat{r}_D)'$. Standardize these residuals as $\hat{\mathbf{r}}^S = \hat{\boldsymbol{\Sigma}}_{\mathbf{r}}^{-1/2}\hat{\mathbf{r}} = (\hat{r}_1^S, \ldots, \hat{r}_D^S)'$, where $\hat{\boldsymbol{\Sigma}}_{\mathbf{r}} = \boldsymbol{\Psi}\,\mathbf{P}(\hat{\boldsymbol{\theta}})\boldsymbol{\Psi}$ is the estimated covariance matrix and $\hat{\boldsymbol{\Sigma}}_{\mathbf{r}}^{-1/2}$ is a square root of the generalized inverse derived from the spectral decomposition of $\hat{\boldsymbol{\Sigma}}_{\mathbf{r}}$. Again, re-standardize these values

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

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

The function `npbmseSFH` yields naive and bias-corrected nonparametric bootstrap estimators $\text{mse}^{naNPB}\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}})\}$ and $\text{mse}^{bcNPB}\{\tilde{\delta}_d(\hat{\boldsymbol{\theta}})\}$ analogous to (44) and (45) respectively.

11 Function eblupSTFH

Function `eblupSTFH` gives small area estimators of $\delta_d = h_d(\mathbf{y}_d)$, $d = 1, \ldots, D$, under an extension of the FH model that takes into account the spatial correlation between neighbor areas and also incorporates historical data (Marhuenda, Molina & Morales, 2013). The area parameter for domain d at current time instant T is estimated borrowing strength from the T time instants and from the D domains. The call to the function is

```
eblupSTFH(formula, D, T, vardir, proxmat, model = "ST",
    MAXITER = 100, PRECISION = 0.0001, data)
```

Let θ_{dt} be the target area characteristic for area d and time instant t, for $d = 1, \ldots, D$ and $t = 1, \ldots, T$. Let $\hat{\delta}_{dt}^{DIR}$ be a direct estimator of δ_{dt} (left hand side of `formula`) and $\mathbf{x}_{dt}$ a column vector containing the aggregated values of p auxiliary variables related linearly with δ_{dt} (right hand side of `formula`). The spatio-temporal FH (STFH) model is stated as follows. In the first stage, we assume

$$\hat{\delta}_{dt}^{DIR} = \delta_{dt} + e_{dt}, \quad d = 1, \ldots, D, \quad t = 1, \ldots, T, \tag{46}$$

where, given δ_{dt}, sampling errors e_{dt} are assumed to be independent and normally distributed with variances ψ_{dt} known for all d and t (`vardir`). In the second stage, the target parameters for all domains and time points are linked through the model

$$\delta_{dt} = \mathbf{x}_{dt}'\boldsymbol{\beta} + u_{1d} + u_{2dt}, \quad d = 1, \ldots, D, \quad t = 1, \ldots, T. \tag{47}$$

Here, the vectors of area-time random effects $(u_{2d1}, \ldots, u_{2dT})'$ are i.i.d. for each area d, following an AR(1) process with autocorrelation parameter ρ_2, that is,

$$u_{2dt} = \rho_2 u_{2d,t-1} + \epsilon_{2dt}, \quad |\rho_2| < 1, \quad \epsilon_{2dt} \overset{iid}{\sim} N(0, \sigma_2^2). \tag{48}$$

The vector of area effects $(u_{11}, \ldots, u_{1D})'$ follows a SAR(1) process with variance parameter σ_1^2, spatial autocorrelation ρ_1 and row-standardized proximity matrix $\mathbf{W} = (w_{d,\ell})$ defined as in Section 7 (`proxmat`), that is,

$$u_{1d} = \rho_1 \sum_{\ell \neq d} w_{d,\ell}\, u_{1\ell} + \epsilon_{1d}, \quad |\rho_1| < 1, \quad \epsilon_{1d} \overset{iid}{\sim} N(0, \sigma_1^2), \quad d = 1, \ldots, D. \tag{49}$$

Let us define the following vectors and matrices obtained by stacking the elements of the model in columns

$$\mathbf{y} = \underset{1\le d\le D}{\text{col}}\,(\underset{1\le t\le T}{\text{col}}\,(\hat{\delta}_{dt}^{DIR})),\ \mathbf{X} = \underset{1\le d\le D}{\text{col}}\,(\underset{1\le t\le T}{\text{col}}\,(\mathbf{x}_{dt}')),$$
$$\mathbf{e} = \underset{1\le d\le D}{\text{col}}\,(\underset{1\le t\le T}{\text{col}}\,(e_{dt})),\ \mathbf{u}_1 = \underset{1\le d\le D}{\text{col}}\,(u_{1d}) \text{ and } \mathbf{u}_2 = \underset{1\le d\le D}{\text{col}}\,(\underset{1\le t\le T}{\text{col}}\,(u_{2dt}).$$

Defining additionally $\mathbf{Z}_1 = \mathbf{I}_D \bigotimes \mathbf{1}_T$ where $\mathbf{I}_D$ is the $D \times D$ identity matrix, $\mathbf{1}_T$ is a vector of ones of size T and $\bigotimes$ is the Kronecker product, $\mathbf{Z}_2 = \mathbf{I}_n$, where $n = DT$ is the total number of observations, $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, the model can be expressed as a general linear mixed model in the form

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

Let $\boldsymbol{\theta} = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)'$ be the vector of unknown parameters involved in the covariance matrix of $\mathbf{y}$. Observe that here $\mathbf{e} \sim N(\mathbf{0}_n, \boldsymbol{\Psi})$, where $\mathbf{0}_n$ denotes a vector of zeros of size n and $\boldsymbol{\Psi}$ is the diagonal matrix $\boldsymbol{\Psi} = \text{diag}_{1\le d\le D}(\text{diag}_{1\le t\le T}(\psi_{dt}))$.

Moreover, $\mathbf{u} \sim N\{\mathbf{0}_n, \mathbf{G}(\boldsymbol{\theta})\}$, where the covariance matrix is the block diagonal matrix $\mathbf{G}(\boldsymbol{\theta}) = \text{diag}\{\sigma_1^2\Omega_1(\rho_1), \sigma_2^2\Omega_2(\rho_2)\}$, with

$$\Omega_1(\rho_1) = \left\{(\mathbf{I}_D - \rho_1\mathbf{W})'(\mathbf{I}_D - \rho_1\mathbf{W})\right\}^{-1}, \tag{50}$$

$$\Omega_2(\rho_2) = \text{diag}_{1\le d\le D}\{\Omega_{2d}(\rho_2)\},$$

$$\Omega_{2d}(\rho_2) = \frac{1}{1-\rho_2^2}\begin{pmatrix} 1 & \rho_2 & \dots & \rho_2^{T-2} & \rho_2^{T-1} \\ \rho_2 & 1 & \ddots & & \rho_2^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho_2^{T-2} & & \ddots & 1 & \rho_2 \\ \rho_2^{T-1} & \rho_2^{T-2} & \dots & \rho_2 & 1 \end{pmatrix}_{T\times T}, \; d = 1, \dots, D. \tag{51}$$

Thus, the covariance matrix of $\mathbf{y}$ is given by

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{G}(\boldsymbol{\theta})\mathbf{Z}' + \boldsymbol{\Psi}.$$

The WLS estimator of $\boldsymbol{\beta}$ and the (componentwise) BLUP of $\mathbf{u}$ obtained by Henderson (1975) are given by

$$\begin{aligned} \tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \left\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{y}, \\ \tilde{\mathbf{u}}(\boldsymbol{\theta}) &= \mathbf{G}(\boldsymbol{\theta})\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\{\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})\}. \end{aligned} \tag{52}$$

Since $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$, the second identity leads to the BLUPs of $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively given by

$$\begin{aligned} \tilde{\mathbf{u}}_1(\boldsymbol{\theta}) &= \sigma_1^2\Omega_1(\rho_1)\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\{\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})\}, \\ \tilde{\mathbf{u}}_2(\boldsymbol{\theta}) &= \sigma_2^2\Omega_2(\rho_2)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\{\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})\}. \end{aligned}$$

Replacing an estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in previous formulas we obtain $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ and the EBLUPs of $\mathbf{u}_1$ and $\mathbf{u}_2$ respectively,

$$\hat{\mathbf{u}}_1 = \tilde{\mathbf{u}}_1(\hat{\boldsymbol{\theta}}) = (\hat{u}_{11}, \dots, \hat{u}_{1D})' \text{ and } \hat{\mathbf{u}}_2 = \tilde{\mathbf{u}}_2(\hat{\boldsymbol{\theta}}) = (\hat{u}_{211}, \dots, \hat{u}_{2DT})'.$$

Finally, the EBLUP of the area characteristic δ_{dt} under the STFH model (46)–(49) returned by function `eblupSTFH` is given by

$$\hat{\delta}_{dt} = \mathbf{x}_{dt}'\hat{\boldsymbol{\beta}} + \hat{u}_{1d} + \hat{u}_{2dt}, \quad d = 1, \dots, D, \quad t = 1, \dots, T.$$

The following subsection describes the REML model fitting procedure applied by function `eblupSTFH` to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$.

Remark 1. *Computation of the inverse of the $n \times n$ matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ involved in (52) can be too time consuming for large n. This is replaced by the inversion of two smaller matrices as follows. Observe that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be expressed as*

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma_1^2\mathbf{Z}_1\Omega_1(\rho_1)\mathbf{Z}_1' + \boldsymbol{\Gamma}(\boldsymbol{\theta}),$$

where $\mathbf{\Gamma}(\theta) = diag_{1\leq d\leq D}\{\Gamma_d(\theta)\}$ *and* $\Gamma_d(\theta) = \sigma_2^2\Omega_{2d}(\rho_2) + diag_{1\leq t\leq T}(\psi_{dt})$, $d = 1,\ldots,D$. *Applying the inversion formula*

$$(A+CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1} \tag{53}$$

with $A = \mathbf{\Gamma}(\boldsymbol{\theta})$, $B = \sigma_1^2\Omega_1(\rho_1)$, $C = \mathbf{Z}_1$ *and* $D = \mathbf{Z}_1'$, *we obtain*

$$\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}) = \mathbf{\Gamma}^{-1}(\boldsymbol{\theta}) - \mathbf{\Gamma}^{-1}(\boldsymbol{\theta})\mathbf{Z}_1\left\{\sigma_1^{-2}\Omega_1^{-1}(\rho_1) + \mathbf{Z}_1'\mathbf{\Gamma}^{-1}(\boldsymbol{\theta})\mathbf{Z}_1\right\}^{-1}\mathbf{Z}_1'\mathbf{\Gamma}^{-1}(\boldsymbol{\theta}),$$

where $\mathbf{\Gamma}^{-1}(\boldsymbol{\theta}) = diag_{1\leq d\leq D}\{\Gamma_d^{-1}(\theta)\}$. *Here,* $\Gamma_d(\theta)$ *is inverted using again (53). This procedure only requires inversion of the* $T\times T$ *matrix* $\Omega_{2d}(\rho_2)$ *given in (51), which is constant for all* d, *and the* $D\times D$ *matrix* $\Omega_1(\rho_1)$ *given in (50).*

11.1 REML fitting method

REML fitting method maximizes the restricted likelihood, which is the joint p.d.f. of a vector of $n-p$ linearly independent contrasts $\mathbf{F}'\mathbf{y}$, where $\mathbf{F}$ is an $n\times(n-p)$ full column rank matrix satisfying $\mathbf{F}'\mathbf{F} = \mathbf{I}_{n-p}$ and $\mathbf{F}'\mathbf{X} = \mathbf{0}_{n-p}$. It holds that $\mathbf{F}'\mathbf{y}$ is independent of $\hat{\boldsymbol{\beta}}$ given in (52). Consequently, the p.d.f. of $\mathbf{F}'\mathbf{y}$ does not depend on $\boldsymbol{\beta}$ and is given by

$$f_R(\boldsymbol{\theta};\mathbf{y}) = (2\pi)^{-(n-p)/2}|\mathbf{X}'\mathbf{X}|^{1/2}|\mathbf{\Sigma}(\boldsymbol{\theta})|^{-1/2}|\mathbf{X}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}(\boldsymbol{\theta})\mathbf{y}\right\},$$

where

$$\mathbf{P}(\boldsymbol{\theta}) = \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}) - \mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\left\{\mathbf{X}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}).$$

Observe that $\mathbf{P}(\boldsymbol{\theta})$ satisfies $\mathbf{P}(\boldsymbol{\theta})\mathbf{\Sigma}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta}) = \mathbf{P}(\boldsymbol{\theta})$ and $\mathbf{P}(\boldsymbol{\theta})\mathbf{X} = \mathbf{0}_n$.

The REML estimator of $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_4)' = (\sigma_1^2,\rho_1,\sigma_2^2,\rho_2)'$ is the maximizer of $\ell_R(\boldsymbol{\theta};\mathbf{y}) = \log f_R(\boldsymbol{\theta};\mathbf{y})$. This maximum is computed using the Fisher-scoring algorithm. Let $\mathbf{S}_R(\boldsymbol{\theta}) = \partial\ell_R(\boldsymbol{\theta};\mathbf{y})/\partial\boldsymbol{\theta} = (S_1^R(\boldsymbol{\theta}),\ldots,S_4^R(\boldsymbol{\theta}))'$ be the scores vector and $\mathcal{I}_R(\boldsymbol{\theta}) = -E\{\partial^2\ell_R(\boldsymbol{\theta};\mathbf{y})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\} = (\mathcal{I}_{rs}^R(\boldsymbol{\theta}))$ the Fisher information matrix associated with $\boldsymbol{\theta}$. Using the fact that

$$\frac{\partial\mathbf{P}(\boldsymbol{\theta})}{\partial\theta_r} = -\mathbf{P}(\boldsymbol{\theta})\frac{\partial\mathbf{\Sigma}(\boldsymbol{\theta})}{\partial\theta_r}\mathbf{P}(\boldsymbol{\theta}),\quad r=1,\ldots,4,$$

the first order partial derivative of $\ell_R(\boldsymbol{\theta};\mathbf{y})$ with respect to θ_r is

$$S_r^R(\boldsymbol{\theta}) = -\frac{1}{2}\,\mathrm{tr}\left\{\mathbf{P}(\boldsymbol{\theta})\frac{\partial\mathbf{\Sigma}(\boldsymbol{\theta})}{\partial\theta_r}\right\} + \frac{1}{2}\,\mathbf{y}'\mathbf{P}(\boldsymbol{\theta})\frac{\partial\mathbf{\Sigma}(\boldsymbol{\theta})}{\partial\theta_r}\mathbf{P}(\boldsymbol{\theta})\mathbf{y},\quad r=1,\ldots,4.$$

The element (r,s) of the Fisher information matrix is the expected value of the negative second order partial derivative of $\ell_R(\boldsymbol{\theta};\mathbf{y})$ with respect to θ_r and θ_s, which yields

$$\mathcal{I}_{rs}^R(\boldsymbol{\theta}) = \frac{1}{2}\mathrm{tr}\left\{\mathbf{P}(\boldsymbol{\theta})\frac{\partial\mathbf{\Sigma}(\boldsymbol{\theta})}{\partial\theta_r}\mathbf{P}(\boldsymbol{\theta})\frac{\partial\mathbf{\Sigma}(\boldsymbol{\theta})}{\partial\theta_s}\right\},\quad r,s=1,\ldots,4.$$

Then, if $\boldsymbol{\theta}^{(k)}$ is the value of the estimator at iteration k, the updating formula of the Fisher-scoring algorithm is given by

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

Finally, the partial derivatives of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ with respect to the components of $\boldsymbol{\theta}$, involved in $\mathbf{S}_R(\boldsymbol{\theta})$ and $\mathcal{I}_R(\boldsymbol{\theta})$, are given by

$$\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \sigma_1^2} = \mathbf{Z}_1\Omega_1(\rho_1)\mathbf{Z}_1', \qquad \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \rho_1} = -\sigma_1^2\mathbf{Z}_1\Omega_1(\rho_1)\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1}\Omega_1(\rho_1)\mathbf{Z}_1',$$
$$\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \sigma_2^2} = \operatorname*{diag}_{1\le d\le D}\left\{\Omega_{2d}(\rho_2)\right\}, \quad \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \rho_2} = \sigma_2^2 \operatorname*{diag}_{1\le d\le D}\left\{\frac{\partial \Omega_{2d}(\rho_2)}{\partial \rho_2}\right\},$$

where

$$\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1} = -\mathbf{W} - \mathbf{W}' + 2\rho_1\mathbf{W}'\mathbf{W}$$

and

$$\frac{\partial \Omega_{2d}(\rho_2)}{\partial \rho_2} = \frac{1}{1-\rho_2^2}\begin{pmatrix} 0 & 1 & \dots & \dots & (T-1)\rho_2^{T-2} \\ 1 & 0 & \ddots & & (T-2)\rho_2^{T-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ (T-2)\rho_2^{T-3} & & \ddots & 0 & 1 \\ (T-1)\rho_2^{T-2} & \dots & \dots & 1 & 0 \end{pmatrix} + \frac{2\rho_2\Omega_{2d}(\rho_2)}{1-\rho_2^2}.$$

12 Function pbmseSTFH

Function `pbmseSTFH` gives parametric bootstrap MSE estimates for the EBLUPs of the domain parameters under the STFH model (46)–(49) as in Marhuenda, Molina & Morales (2013). The call to the function is

```
pbmseSTFH(formula, D, T, vardir, proxmat, B = 100, model = "ST",
          MAXITER = 100, PRECISION = 0.0001, data)
```

where the arguments are the same as in `eblupSTFH`, together with the number of bootstrap replicates `B`. The parametric bootstrap procedure is described below:

(1) Using the available data $\{(\hat{\delta}_{dt}^{DIR}, \mathbf{x}_{dt}), t = 1,\dots,T, d = 1,\dots,D\}$, fit the STFH model (46)–(49) and obtain model parameter estimates $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_1^2$, $\hat{\rho}_1$, $\hat{\sigma}_2^2$ and $\hat{\rho}_2$.

(2) Generate bootstrap area effects $\{u_{1d}^{*(b)}, d = 1,\dots,D\}$, from the SAR(1) process given in (49), using $(\hat{\sigma}_1^2, \hat{\rho}_1)$ as true values of parameters (σ_1^2, ρ_1).

(3) Independently of $\{u_{1d}^{*(b)}\}$ and independently for each d, generate bootstrap time effects $\{u_{2dt}^{*(b)}, t = 1,\dots,T\}$, from the AR(1) process given in (48), with $(\hat{\sigma}_2^2, \hat{\rho}_2)$ acting as true values of parameters (σ_2^2, ρ_2).

(4) Calculate true bootstrap quantities,

$$\delta_{dt}^{*(b)} = \mathbf{x}'_{dt}\hat{\boldsymbol{\beta}} + u_{1d}^{*(b)} + u_{2dt}^{*(b)}, \quad t = 1, \ldots, T, \quad d = 1, \ldots, D.$$

(5) Generate errors $e_{dt}^{*(b)} \overset{ind.}{\sim} N(0, \psi_{dt})$ and obtain bootstrap data from the sampling model,

$$\hat{\delta}_{dt}^{DIR*(b)} = \delta_{dt}^{*(b)} + e_{dt}^{*(b)}, \quad t = 1, \ldots, T, \quad d = 1, \ldots, D.$$

(6) Using the new bootstrap data $\{(\hat{\delta}_{dt}^{DIR*(b)}, \mathbf{x}_{dt}), t = 1, \ldots, T, d = 1, \ldots, D\}$, fit the STFH model (46)–(49) and obtain the bootstrap EBLUPs,

$$\hat{\delta}_{dt}^{*(b)} = \mathbf{x}'_{dt}\hat{\boldsymbol{\beta}}^{*(b)} + \hat{u}_{1d}^{*(b)} + \hat{u}_{2dt}^{*(b)}, \quad t = 1, \ldots, T, \quad d = 1, \ldots, D.$$

(7) Repeat steps (1)-(6) for $b = 1, \ldots, B$, where B is a large number.

(8) The parametric bootstrap MSE estimates returned by function `pbmseSTFH` are given by

$$mse(\hat{\delta}_{dt}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\delta}_{dt}^{*(b)} - \delta_{dt}^{*(b)}\right)^2, \quad t = 1, \ldots, T, \quad d = 1, \ldots, D. \quad (54)$$

13 Function eblupBHF

Function `eblupBHF` estimates the area means $\bar{Y}_d$, $d = 1, \ldots, D$, under the unit level model introduced by Battese, Harter & Fuller (1988) (BHF model). The call to the function is

```
eblupBHF(formula, dom, selectdom, meanxpop, popnsize,
   method = "REML", data)
```

The function allows to select a subset of domains for estimation through the argument `selectdom`, but dropping this argument it estimates in all domains. Let Y_{dj} be the value of the target variable for unit j in domain d (left hand side of `formula`). The BHF model assumes

$$\begin{aligned} &Y_{dj} = \mathbf{x}'_{dj}\boldsymbol{\beta} + u_d + e_{dj}, \quad j = 1, \ldots, N_d, \quad d = 1, \ldots, D, \\ &u_d \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{dj} \overset{iid}{\sim} N(0, \sigma_e^2), \end{aligned} \quad (55)$$

where $\mathbf{x}_{dj}$ is a vector containing the values of p explanatory variables for the same unit (right hand side of `formula`), u_d is the area random effect and e_{dj} is the individual error, where area effects u_d and errors e_{dj} are independent. Let us define vectors and matrices obtained by stacking in columns 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

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

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 $\mathbf{X}_d$ and $\mathbf{V}_d$ as

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

If σ_u^2 and σ_e^2 are known, the BLUP of the small area mean $\bar{Y}_d$ is given by

$$\tilde{\bar{Y}}_d = \frac{1}{N_d}\left(\sum_{j\in s_d} Y_{dj} + \sum_{j\in r_d} \tilde{Y}_{dj}\right), \tag{56}$$

where $\tilde{Y}_{dj} = \mathbf{x}'_{dj}\tilde{\boldsymbol{\beta}} + \tilde{u}_d$ is the BLUP of Y_{dj}. Here, $\tilde{\boldsymbol{\beta}}$ is the WLS estimator of $\boldsymbol{\beta}$ and $\tilde{u}_d$ is the BLUP of u_d, given respectively by

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{d=1}^{D} \mathbf{X}_d\mathbf{V}_{ds}^{-1}\mathbf{X}'_d\right)^{-1} \sum_{d=1}^{D} \mathbf{X}_d\mathbf{V}_{ds}^{-1}\mathbf{y}_d, \tag{57}$$

$$\tilde{u}_d = \gamma_d(\bar{y}_{ds} - \bar{\mathbf{x}}'_{ds}\tilde{\boldsymbol{\beta}}), \tag{58}$$

where $\bar{y}_{ds} = n_d^{-1}\sum_{j\in s_d} Y_{dj}$, $\bar{\mathbf{x}}_{ds} = n_d^{-1}\sum_{j\in s_d} \mathbf{x}_{dj}$ and $\gamma_d = \sigma_u^2/(\sigma_u^2 + \sigma_e^2/n_d)$, $d = 1, \ldots, D$.

Let $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ be consistent estimators of σ_u^2 and σ_e^2 respectively, such as those obtained by ML or REML. The EBLUP is

$$\hat{\bar{Y}}_d = \frac{1}{N_d}\left(\sum_{j\in s_d} Y_{dj} + \sum_{j\in r_d} \hat{Y}_{dj}\right), \tag{59}$$

where $\hat{Y}_{dj} = \mathbf{x}'_{dj}\hat{\boldsymbol{\beta}} + \hat{u}_d$ is the EBLUP of Y_{dj}, $\hat{\boldsymbol{\beta}}$ and $\hat{u}_d$ are given respectively by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{d=1}^{D} \mathbf{X}_d\hat{\mathbf{V}}_{ds}^{-1}\mathbf{X}'_d\right)^{-1} \sum_{d=1}^{D} \mathbf{X}_d\hat{\mathbf{V}}_{ds}^{-1}\mathbf{y}_d \tag{60}$$

$$\hat{u}_d = \hat{\gamma}_d(\bar{y}_{ds} - \bar{\mathbf{x}}'_{ds}\hat{\boldsymbol{\beta}}), \tag{61}$$

with $\hat{\gamma}_d = \hat{\sigma}_u^2/(\hat{\sigma}_u^2 + \hat{\sigma}_e^2/n_d)$ and $\hat{\mathbf{V}}_{ds} = \hat{\sigma}_u^2\mathbf{1}_{n_d}\mathbf{1}'_{n_d} + \hat{\sigma}_e^2\mathbf{I}_n$, $d = 1, \ldots, D$. Replacing (60) and (61) in (59), we obtain the expression for the EBLUP of $\bar{Y}_d$ returned by function `eblupBHF` ,

$$\hat{\bar{Y}}_d = f_d\,\bar{y}_{ds} + \left(\bar{\mathbf{X}}_d - f_d\,\bar{\mathbf{x}}_{ds}\right)'\hat{\boldsymbol{\beta}} + (1 - f_d)\hat{u}_d,$$

where $f_d = n_d/N_d$ is the sampling fraction. Note that the EBLUP requires the vector of population means of the auxiliary variables $\bar{\mathbf{X}}_d$ (`meanxpop`) and the population sizes (`popnsize`) apart from the sample data (specified in `formula`), but the individual values of the auxiliary variables for each population unit are not needed.

14 Function pbmseBHF

Function `pbmseBHF` gives a parametric bootstrap MSE estimate for the EBLUP under the BHF model (55). The call to the function is

```
pbmseBHF(formula, dom, selectdom, meanxpop, popnsize, B = 200,
   method = "REML", data)
```

The function applies the parametric bootstrap procedure for finite populations introduced by González-Manteiga et al. (2008a) particularized to the estimation of means. The estimated MSEs are obtained as follows:

1) Fit the BHF model (55) to sample data $\mathbf{y}_s = (\mathbf{y}'_{1s}, \ldots, \mathbf{y}'_{Ds})'$ and obtain model parameter estimates $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2_u$ and $\hat{\sigma}^2_e$.

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

3) Generate, independently of the random effects $u^{*(b)}_d$, bootstrap errors for sample elements $e^{*(b)}_{dj} \overset{iid}{\sim} N(0, \hat{\sigma}^2_e)$, $j \in s_d$, and error domain means $\bar{E}^{*(b)}_d \overset{iid}{\sim} N(0, \hat{\sigma}^2_e/N_d)$, $d = 1, \ldots, D$. ,

4) Compute the true domain means of this bootstrap population, given by

$$\bar{Y}^{*(b)}_d = \bar{\mathbf{X}}'_d \hat{\boldsymbol{\beta}} + u^{*(b)}_d + \bar{E}^{*(b)}_d, \quad d = 1, \ldots, D.$$

   Observe that computation of $\bar{Y}^{*(b)}_d$ does not require the individual values $\mathbf{x}_{dj}$, for each out-of-sample unit $j \in r_d$.

5) Using the known sample vectors $\mathbf{x}_{dj}$, $j \in s_d$, generate the model responses for sample elements from the model

$$Y^{*(b)}_{dj} = \mathbf{x}'_{dj} \hat{\boldsymbol{\beta}} + u^{*(b)}_d + e^{*(b)}_{dj}, \quad j \in s_d, \quad d = 1, \ldots, D.$$

   Let $\mathbf{y}^{*(b)}_s = ((\mathbf{y}^{*(b)}_{1s})', \ldots, (\mathbf{y}^{*(b)}_{Ds})')'$ be the bootstrap sample data vector.

6) Fit the BHF model (55) to bootstrap data $\mathbf{y}^{*(b)}_s$ and obtain the bootstrap EBLUPs $\hat{\bar{Y}}^{*(b)}_d$, $d = 1, \ldots, D$.

7) Repeat steps 2)–7) for $b = 1, \ldots, B$. Let $\bar{Y}_d^{*(b)}$ be the true mean and $\hat{\bar{Y}}_d^{*(b)}$ the corresponding EBLUP of domain d for bootstrap replicate b. The parametric bootstrap estimates of the MSEs of the EBLUPs $\hat{\bar{Y}}_d$ returned by function `pbmseBHF` are given by

$$mse(\hat{\bar{Y}}_d) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\bar{Y}}_d^{*(b)} - \bar{Y}_d^{*(b)}\right)^2, \quad d = 1, \ldots, D. \tag{62}$$

15 Function ebBHF

Function `ebBHF` estimates non linear area parameters $\delta_d = h_d(\mathbf{y}_d)$, $d = 1, \ldots, D$ under the BHF model (55), using the empirical best/Bayes (EB) method of Molina & Rao (2010). The call to the function is

```
ebBHF(formula, dom, selectdom, Xnonsample, MC = 100, data,
     transform = "BoxCox", lambda = 0, constant = 0, indicator)
```

where the function $h_d()$ is specified by the user (`indicator`). This function assumes that model responses Y_{dj} are obtained by a transformation of the values E_{dj} of a quantitative variable as $Y_{dj} = T(E_{dj})$. The transformation $T()$ (`transform`) must be selected by the user between the Box-Cox family or the power family of transformations, to achieve approximate normality of the Y_{dj} values. Both families contain two parameters, an additive constant m and a power λ. The Box-Cox family is given by

$$T(E_{dj}) = \begin{cases} \left\{(E_{dj}+m)^{\lambda} - 1\right\}/\lambda, & \lambda \neq 0; \\ \log(E_{dj}+m), & \lambda = 0, \end{cases}$$

and the power family is

$$T(E_{dj}) = \begin{cases} (E_{dj}+m)^{\lambda}, & \lambda \neq 0; \\ \log(E_{dj}+m), & \lambda = 0. \end{cases}$$

The parameters m (`constant`) and λ (`lambda`) must be specified by the user. Note that setting $m = 0$ and $\lambda = 1$ means no transformation. Function `ebBHF` assumes that the transformed variables $Y_{dj} = T(E_{dj})$ follow the BHF model (55).

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. The best predictor of $\delta_d = h_d(\mathbf{y}_d)$ is given by

$$\tilde{\delta}_d = E_{\mathbf{y}_{dr}}\left[h_d(\mathbf{y}_d)|\mathbf{y}_{ds}\right] = \int h_d(\mathbf{y}_d) f(\mathbf{y}_{dr}|\mathbf{y}_{ds})\, d\mathbf{y}_{dr}, \tag{63}$$

where $f(\mathbf{y}_{dr}|\mathbf{y}_{ds})$ is the joint density of $\mathbf{y}_{dr}$ given the observed data vector $\mathbf{y}_{ds}$. The expectation in (63) is approximated by Monte Carlo. For this, function

`ebBHF` generates L replicates $\{\mathbf{y}_{dr}^{(\ell)}; \ell = 1, \ldots, L\}$ of $\mathbf{y}_{dr}$ from the estimated conditional distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$, where L can be specified by the user (`MC`). The elements of $\mathbf{y}_{dr}$ or non-sample values $Y_{dj}^{(\ell)}$ are generated from the estimated model

$$Y_{dj}^{(\ell)} = \mathbf{x}_{dj}'\hat{\boldsymbol{\beta}} + \hat{u}_d + v_d + \varepsilon_{di}, \tag{64}$$

$$v_d \sim N(0, \hat{\sigma}_u^2(1-\hat{\gamma}_d)),\ \varepsilon_{dj} \sim N(0, \hat{\sigma}_e^2),\ j \in r_d,\ d = 1, \ldots, D, \tag{65}$$

where $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ are the estimated model parameters. Attaching the sample values $\mathbf{y}_{ds}$ to the generated out-of-sample vector $\mathbf{y}_{dr}^{(\ell)}$, full population vectors $\mathbf{y}_d^{(\ell)} = ((\mathbf{y}_{dr}^{(\ell)})', \mathbf{y}_{ds}')'$ are obtained. Then, function `ebBHF` returns the Monte Carlo approximation to the EB predictor of δ_d,

$$\hat{\delta}_d = \frac{1}{L}\sum_{\ell=1}^{L} h_d(\mathbf{y}_d^{(\ell)}). \tag{66}$$

Examples of non linear area parameters are the members of the FGT family of poverty indicators defined by Foster, Greer & Thorbecke (1984), which for domain d are given by

$$F_{\alpha d} = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z - E_{dj}}{z}\right)^{\alpha} I(E_{dj} < z), \quad \alpha \geq 0, \tag{67}$$

where E_{dj} is in this case a welfare measure such as income or expenditure, z is the poverty line defined for the population and $I(\text{condition})$ is the indicator function with value 1 when condition is true and 0 otherwise. If a transformation $T()$ is specified through the arguments `transform`, `lambda` and `constant`, the function `ebBHF` calculates the EB estimates (66) of the parameters

$$\delta_d = h_d(\mathbf{y}_d) = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z - T^{-1}(Y_{dj})}{z}\right)^{\alpha} I(T^{-1}(Y_{dj}) < z), \quad d = 1, \ldots, D.$$

16 Function pbmseebBHF

Function `pbmseEB` gives parametric bootstrap MSE estimates for the EB estimators (66) under the BHF model. The call to the function is

```
pbmseebBHF(formula, dom, selectdom, Xnonsample, B = 100, MC = 100,
  data, transform = "BoxCox", lambda = 0, constant = 0, indicator)
```

where the arguments are as in `ebBHF`, together with the number of bootstrap replicates `B`. The function uses the parametric bootstrap of González-Manteiga et al. (2008a), which proceeds as follows:

1) Fit the BHF model (55), deriving estimates $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$.

2) Generate bootstrap domain effects as

$$u_d^{*(b)} \overset{iid}{\sim} N(0, \hat{\sigma}_u^2), \quad d = 1, \ldots, D.$$

3) Generate, independently of $u_1^{*(b)}, \ldots, u_D^{*(b)}$, model errors

$$e_{dj}^{*(b)} \overset{iid}{\sim} N(0, \hat{\sigma}_e^2), \quad j = 1, \ldots, N_d, \ d = 1, \ldots, D$$

4) Generate a bootstrap population of Y_{dj} values from the model

$$Y_{dj}^{*(b)} = \mathbf{x}_{dj}'\hat{\boldsymbol{\beta}} + u_d^{*(b)} + e_{dj}^{*(b)}, \quad j = 1, \ldots, N_d, \ d = 1, \ldots, D.$$

5) Let us define the area vector $\mathbf{y}_d^{*(b)} = (Y_{d1}^{*(b)}, \ldots, Y_{dN_d}^{*(b)})'$. Calculate target area quantities for the bootstrap population

$$\delta_d^{*(b)} = h_d(\mathbf{y}_d^{*(b)}), \quad d = 1, \ldots, D.$$

6) For the original sample $s = s_1 \cup \cdots \cup s_D$, let $\mathbf{y}_s^{*(b)}$ be the vector containing the bootstrap observations whose indices are in the sample, that is, containing $Y_{dj}^{*(b)}$, $j \in s_d$, $d = 1, \ldots, D$. Fit again the BHF model (55) to bootstrap sample data $\mathbf{y}_s^{*(b)}$ and obtain bootstrap estimates $\hat{\sigma}_u^{2*(b)}$, $\hat{\sigma}_e^{2*(b)}$ and $\hat{\boldsymbol{\beta}}^{*(b)}$.

7) Using the bootstrap sample data $\mathbf{y}_s^{*(b)}$, obtain the bootstrap EB estimators $\hat{\delta}_d^{*(b)}$, $d = 1, \ldots, D$, through the Monte Carlo approximation (66).

8) Repeat 2)–7) for $b = 1, \ldots, B$, obtaining true value $\delta_d^{*(b)}$ and EB estimate $\hat{\delta}_d^{*(b)}$ for each area $d = 1, \ldots, D$ and bootstrap sample $b = 1, \ldots, B$.

9) The bootstrap MSE estimates returned by function `pbmseEB` are given by

$$\text{mse}_B(\hat{\delta}_d) = B^{-1} \sum_{b=1}^{B} \left(\hat{\delta}_d^{*(b)} - \delta_d^{*(b)} \right)^2, \quad d = 1, \ldots, D.$$

References

Anselin, L. (1988). *Spatial Econometrics. Methods and Models*. Boston: Kluwer Academic Publishers.

Banerjee, S., Carlin, B. & Gelfand, A.(2004). *Hierarchical Modeling and Analysis for Spatial Data*. New York: Chapman and Hall.

Battese, G. E., Harter, R. M. & 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.

Cressie, N. (1993). *Statistics for Spatial Data.*New York: John Wiley & Sons.

Datta, G. S. & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica* **10**, 613–627.

Datta, G. S. & Rao, J. N. K. & Smith D. D.(2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika* **92**, 183–196.

Drew, D., Singh, M.P. & Choudhry, G.H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methodology* **8**, 17–47.

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

Foster, J., Greer, J. & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica* **52**, 761–766.

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

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

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447.

Jiang, J. (1996). REML estimation: asymptotic behavior and related topics. *Annals of Statistics* **24**, 255-286.

Marhuenda, Y., Molina, I. & Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics and Data Analysis* **58**, 308-325.

Molina, I., Salvati, N. & Pratesi, M. (2009). Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics* **24**, 441-458.

Molina, I. & Rao, J.N.K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics* **38**, 369-385.

Prasad, N. & 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.

Petrucci, A. & Salvati, N. (2006). Small area estimation for spatial correlation in watershed erosion assessment. *Journal of Agricultural, Biological and Environmental Statistics* **11**, 169–182.

Singh, B., Shukla, G. & Kundu, D.(2005). Spatio-temporal models in small area estimation. *Survey Methodology* **31**, 183–195.