Southampton Statistical Sciences Research Institute
University of Southampton

Small Area Estimation Under Linear and Generalized Linear Mixed Models With Time and Area Effects

Ayoub Saei, Ray Chambers

Abstract

This is the theory component of the report on small area estimation theory that was prepared as part of Southampton's involvement in the EURAREA "Enhancing small area estimation techniques to meet European Needs" project IST 2000-26290 in the European Union's Fifth Research And Technological Development Framework Programme.

S[3]RI Methodology Working Paper M03/15

Small Area Estimation Under Linear and Generalized Linear Mixed Models With Time and Area Effects

Ayoub Saei and Ray Chambers

Southampton Statistical Sciences Research Institute

University of Southampton

Highfield, Southampton

SO17 1BJ

United Kingdom

September 2003

Abstract

This is the theory component of the report on small area estimation theory that was prepared as part of Southampton's involvement in the EURAREA "Enhancing small area estimation techniques to meet European Needs" project IST 2000-26290 in the European Union's Fifth Research And Technological Development Framework Programme

1 Estimation Under a Linear Mixed Model With Time and Area Effects

Let the vector $\mathbf{y} = \{y_{dti}\}$ consist of the population values of the survey variable of interest Y. The subscripts d, t and i ($d = 1,2,\ldots,D$; $t = 1,2,\ldots,T$; $i = 1,2,\ldots,N_{dt}$) represent area, time and unit respectively. Let $\mathbf{X}$ be the matrix of population values of the auxiliary covariates $\mathbf{X}_{dti}$. Assuming that the $u_{j(dt)}$, $j = 1, 2, \ldots, J$ represents random effects that are related to area and time, the objective is to estimate/predict the value of the vector $\boldsymbol{\theta} = \mathbf{ay}$ for a given matrix $\mathbf{a}$. Initially we assume that the values in $\mathbf{y}$ follow a linear mixed model with random area and time effects of the form:

$$y_{dti} = \beta_0 + \mathbf{X}'_{dti}\boldsymbol{\beta}_t + \sum_{j=1}^{J} u_{j(dt)} + e_{dti}. \tag{1.1}$$

The vector $\boldsymbol{\beta}_t$ contains regression coefficients at time $t = 1, 2, .. T$, β_0 is the intercept, the random effects $\{u_{j(dt)}\}$ and $\{e_{dti}\}$ are assumed to be mutually independent and to follow normal distributions with zero means and respective variances. The (dt) in $u_{j(dt)}$ indicates that these random effects are related to the specific area d and the specific time t. Linear models with spatially correlated area effects and linear models with independent and autocorrelated time effects are special cases of model (1.1) with $J = 1$ and $J = 2$ respectively. Put $\mathbf{u} = [\mathbf{u}'_1, \mathbf{u}'_2, \ldots \mathbf{u}'_J]'$, and $\mathbf{e} = [e_{dti}]$ and let $\mathbf{Z}_j$ be the "incidence" matrix for the random effects vector $\mathbf{u}_j$. The model (1.1) can then be written in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1.2}$$

where $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_J]$. The random vectors $\mathbf{u}_j$ are assumed to be distributed as multivariate normal with zero mean vectors and variance-covariance matrices $\sigma_j^2\boldsymbol{\Omega}_j(\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ is a vector of correlation/covariance parameters. The vector $\mathbf{e}$ is a normal error vector with zero mean and variance $\sigma^2\mathbf{W}$, where $\mathbf{W}$ is a known square matrix of order N and $\boldsymbol{\Omega}_j(\boldsymbol{\rho})$ is a matrix of order equal to the rank of the matrix $\mathbf{Z}_j$. The covariance matrix of $\mathbf{y}$ is then

$$\sigma^2(\mathbf{W} + \sum_{j=1}^{J}\varphi_j\mathbf{Z}_j\boldsymbol{\Omega}_j\mathbf{Z}'_j) = \sigma^2(\mathbf{W} + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}') = \sigma^2\boldsymbol{\Sigma}$$

where $\varphi_j = \sigma_j^2 / \sigma^2$ and $\boldsymbol{\Omega} = diag(\varphi_j\boldsymbol{\Omega}_j(\boldsymbol{\rho}))$.

The vector $\mathbf{y}$ can be partitioned as $\mathbf{y} = [\mathbf{y}'_s\ \mathbf{y}'_r]'$ after the sample is observed, with the subscripts of s and r corresponding to sample and non-sample population units respectively.

These subscripts will be used below to denote conformable partitions of other vectors and matrices. Thus, the model (1.2) is partitioned conformably as

(1.3) $$\begin{bmatrix}\mathbf{y}_s\\ \mathbf{y}_r\end{bmatrix} = \begin{bmatrix}\mathbf{X}_s\\ \mathbf{X}_r\end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix}\mathbf{Z}_s\\ \mathbf{Z}_r\end{bmatrix}\mathbf{u} + \begin{bmatrix}\mathbf{e}_s\\ \mathbf{e}_r\end{bmatrix} = \begin{bmatrix}\mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s\\ \mathbf{X}_r\boldsymbol{\beta} + \mathbf{Z}_r\mathbf{u} + \mathbf{e}_r\end{bmatrix} = \begin{bmatrix}\boldsymbol{\eta}_s + \mathbf{e}_s\\ \boldsymbol{\eta}_r + \mathbf{e}_r\end{bmatrix}$$

where $\mathbf{X} = [\mathbf{X}_s\ \mathbf{X}_r]$, $\mathbf{Z} = [\mathbf{Z}_s\ \mathbf{Z}_r]$, $\boldsymbol{\eta}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u}$ and $\boldsymbol{\eta}_r = \mathbf{X}_r\boldsymbol{\beta} + \mathbf{Z}_r\mathbf{u}$. Similarly, the matrix $\mathbf{a}$ is partitioned conformably as $\mathbf{a} = [\mathbf{a}_s\ \mathbf{a}_r]$. The vector-valued parameter of interest $\boldsymbol{\theta} = \mathbf{ay}$ can then be written

(1.4) $$\boldsymbol{\theta} = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\mathbf{y}_r.$$

The first term in (1.4) depends only on the sample values and is known after the sample is observed. The second term, which depends on the non-sample values, is unknown. The estimate or predicted value of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, is then obtained by replacing $\mathbf{y}_r$ with its predicted value in (1.4). Let $E(\mathbf{y}_r|\mathbf{u}) = \boldsymbol{\eta}_r = \mathbf{X}_r\boldsymbol{\beta} + \mathbf{Z}_r\mathbf{u}$. The predictor of $\mathbf{y}_r$ is then $\hat{\boldsymbol{\eta}}_r = \mathbf{X}_r\hat{\boldsymbol{\beta}} + \mathbf{Z}_r\hat{\mathbf{u}}$, where the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ are obtained by fitting the linear mixed model

(1.5) $$\mathbf{y}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s$$

to the sample data. This leads to the predicted value

(1.6) $$\hat{\boldsymbol{\theta}} = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\hat{\mathbf{y}}_r = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\hat{\boldsymbol{\eta}}_r = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r(\mathbf{X}_r\hat{\boldsymbol{\beta}} + \mathbf{Z}_r\hat{\mathbf{u}}).$$

2 Best Linear Unbiased Prediction (BLUP)

The linear mixed model (1.5) characterises variation between areas and variation over time in the values of the characteristic Y. The aim is to predict or estimate $\boldsymbol{\theta} = \mathbf{ay}$ using the sample values $\mathbf{y}_s$. For known variance components in the model (1.5), the BLUP method provides a predictor of $\boldsymbol{\theta}$ that is an unbiased linear function of the sample values $\mathbf{y}_s$ and has minimum variance among all other linear unbiased predictors/estimators of $\boldsymbol{\theta}$. The method requires that the covariance between the observable random variable $\mathbf{y}_s$ and the non-observable random variable $\mathbf{u}$ be known. It does not impose a normality assumption on the random effect $\mathbf{u}$. Under (1.5) the covariance between $\mathbf{u}$ and $\mathbf{y}_s$ is $\sigma^2\mathbf{Z}_s'\boldsymbol{\Omega}$. In this case, following Henderson (1963), the best linear unbiased predictor (BLUP) of $\boldsymbol{\eta}_r$ is

(2.1) $$\tilde{\boldsymbol{\eta}}_r = \mathbf{X}_r\hat{\boldsymbol{\beta}} + \mathbf{Z}_r\hat{\mathbf{u}}$$

where

(2.2) $$\hat{\boldsymbol\beta} = (\mathbf{X}_s'\boldsymbol\Sigma_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\boldsymbol\Sigma_s^{-1}\mathbf{y}_s \text{ and } \hat{\mathbf{u}} = \boldsymbol\Omega\mathbf{Z}_s'\boldsymbol\Sigma_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}).$$

Note that $\hat{\boldsymbol\beta}$ is Aitken's generalized least squares (GLS) estimator of $\boldsymbol\beta$ under (1.5). Replacing $\hat{\mathbf{u}}$ in (2.1) by its value from (2.2), and $\hat{\boldsymbol\eta}_r$ by $\tilde{\boldsymbol\eta}_r$ in the right hand side of (1.6) gives

(2.3) $$\tilde{\boldsymbol\theta} = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\tilde{\boldsymbol\eta}_r = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\mathbf{X}_r\hat{\boldsymbol\beta} + \mathbf{a}_r\mathbf{Z}_r\boldsymbol\Omega\mathbf{Z}_s'\boldsymbol\Sigma_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}).$$

An alternative form of (2.3) which is useful in computation can be obtained from following theorem.

Theorem: Put $\mathbf{T}_s^* = (\boldsymbol\Omega^{-1} + \mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s)^{-1} = [\mathbf{T}_{sjj'}^*]$ and $\boldsymbol\Sigma_s = \mathbf{W}_s + \mathbf{Z}_s\boldsymbol\Omega\mathbf{Z}_s'$. Then

a) $\boldsymbol\Sigma_s^{-1} = \mathbf{W}_s^{-1} - \mathbf{W}_s^{-1}\mathbf{Z}_s\mathbf{T}_s^*\mathbf{Z}_s'\mathbf{W}_s^{-1}$

b) $\mathbf{Z}_s'\boldsymbol\Sigma_s^{-1} = \boldsymbol\Omega^{-1}\mathbf{T}_s^*\mathbf{Z}_s'\mathbf{W}_s^{-1}$

c) $\boldsymbol\Omega\mathbf{Z}_s'\boldsymbol\Sigma_s^{-1} = \mathbf{T}_s^*\mathbf{Z}_s'\mathbf{W}_s^{-1}$

d) $\mathbf{Z}_s'\boldsymbol\Sigma_s^{-1}\mathbf{Z}_s = \boldsymbol\Omega^{-1} - \boldsymbol\Omega^{-1}\mathbf{T}_s^*\boldsymbol\Omega^{-1}$

e) $\mathbf{Z}_{sj}'\boldsymbol\Sigma_s^{-1}\mathbf{Z}_{sj'} = \begin{cases} \varphi_j^{-1}\boldsymbol\Omega_j^{-1} - \varphi_j^{-2}\boldsymbol\Omega_j^{-1}\mathbf{T}_{sjj}^*\boldsymbol\Omega_j^{-1} & \text{for } j = j' \\ -\varphi_j^{-1}\varphi_{j'}^{-1}\boldsymbol\Omega_j^{-1}\mathbf{T}_{sjj'}^*\boldsymbol\Omega_{j'}^{-1} & \text{otherwise} \end{cases}$

Using the above Theorem an alternative formulation of (2.3) is

(2.4) $$\tilde{\boldsymbol\theta} = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\mathbf{X}_r\hat{\boldsymbol\beta} + \mathbf{a}_r\mathbf{Z}_r\mathbf{T}_s^*\mathbf{Z}_s'\mathbf{W}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}) = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\hat{\mathbf{y}}_r.$$

The mean cross-product error (MCPE) matrix of the BLUP $\tilde{\boldsymbol\theta}$ is obtained by writing $\tilde{\boldsymbol\theta} - \boldsymbol\theta = \mathbf{a}_r(\hat{\mathbf{y}}_r - \mathbf{y}_r) = \mathbf{a}_r(\tilde{\boldsymbol\eta}_r - \mathbf{y}_r) = \mathbf{a}_r(\tilde{\boldsymbol\eta}_r - \boldsymbol\eta_r + \boldsymbol\eta_r - \mathbf{y}_r) = \mathbf{a}_r[(\tilde{\boldsymbol\eta}_r - \boldsymbol\eta_r) - (\mathbf{y}_r - \boldsymbol\eta_r)]$. Let $\mathbf{X}_r^* = \mathbf{a}_r\mathbf{X}_r$ and $\mathbf{Z}_r^* = \mathbf{a}_r\mathbf{Z}_r$. The mean cross-product error matrix of BLUP estimator $\tilde{\boldsymbol\theta}$ is then

$$MCPE(\tilde{\boldsymbol\theta}) = \mathrm{E}[(\tilde{\boldsymbol\theta} - \boldsymbol\theta)(\tilde{\boldsymbol\theta} - \boldsymbol\theta)'] = \begin{pmatrix}\mathbf{X}_r^* & \mathbf{Z}_r^*\end{pmatrix}\left(E\left(\begin{bmatrix}(\hat{\boldsymbol\beta} - \boldsymbol\beta) \\ (\hat{\mathbf{u}} - \mathbf{u})\end{bmatrix}\begin{bmatrix}(\hat{\boldsymbol\beta} - \boldsymbol\beta) \\ (\hat{\mathbf{u}} - \mathbf{u})\end{bmatrix}'\right)\right)\begin{pmatrix}\mathbf{X}_r^{*\prime} \\ \mathbf{Z}_r^{*\prime}\end{pmatrix} + \sigma^2\mathbf{a}_r\mathbf{W}_r\mathbf{a}_r'$$

(2.5) $$= \sigma^2[\mathbf{X}_r^* - \mathbf{Z}_r^*\mathbf{T}_s^*\mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{X}_s](\mathbf{X}_s'\boldsymbol\Sigma_s^{-1}\mathbf{X}_s)^{-1}[\mathbf{X}_r^{*\prime} - \mathbf{X}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s\mathbf{T}_s^*\mathbf{Z}_r^{*\prime}] + \sigma^2\mathbf{a}_r\mathbf{W}_r\mathbf{a}_r'$$
$$= \mathbf{G}_1(\boldsymbol\omega) + \mathbf{G}_2(\boldsymbol\omega) + \mathbf{G}_4(\boldsymbol\omega)$$

where

$\mathbf{G}_1(\boldsymbol\omega) = \sigma^2\mathbf{Z}_r^*\mathbf{T}_s^*\mathbf{Z}_r^{*\prime}$,

$\mathbf{G}_2(\boldsymbol\omega) = \sigma^2[\mathbf{X}_r^* - \mathbf{Z}_r^*\mathbf{T}_s^*\mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{X}_s](\mathbf{X}_s'\boldsymbol\Sigma_s^{-1}\mathbf{X}_s)^{-1}[\mathbf{X}_r^{*\prime} - \mathbf{X}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s\mathbf{T}_s^*\mathbf{Z}_r^{*\prime}]$,

$\mathbf{G}_4 = \sigma^2\mathbf{a}_r\mathbf{W}_r\mathbf{a}_r'$

$\boldsymbol\omega = (\sigma^2, \boldsymbol\varphi', \boldsymbol\rho')'$.

3 Empirical Best Linear Unbiased Prediction (EBLUP)

The BLUP development set out in the preceding section assumes variance components are known. In practice of course, this is hardly ever the case. We therefore need to estimate these parameters from the sample data. The empirical best linear unbiased prediction (EBLUP) method replaces them by sample-based estimates in the BLUP.

The two methods most frequently used to estimate variance components are maximum likelihood (ML) and residual maximum likelihood (REML). Kackar and Harville (1981) showed that the ML and REML variance component estimators are translation invariant and even functions of the data. In this section we derive the estimating equations for the variance component vector, $\boldsymbol{\omega} = (\sigma^2, \boldsymbol{\varphi}', \boldsymbol{\rho}')'$, under ML and REML.

3.1 Maximum Likelihood Estimation of the Variance Components (ML)

Maximum likelihood estimation of the regression parameter vector $\boldsymbol{\beta}$ and the variance components vector $\boldsymbol{\omega}$ requires parametric specification of the distribution of the random variable $\mathbf{y}_s$. A standard assumption is that this variable has a multivariate normal distribution. Given that assumption, we can write down the log-likelihood function generated by the observation vector $\mathbf{y}_s$ under the general linear mixed model (1.5) as

$$l = -(1/2)[n \ln(2\pi\sigma^2) + \ln|\boldsymbol{\Sigma}_s^{-1}| + \sigma^{-2}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})'\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})].$$

Differentiation of this log-likelihood function with respect to the parameters of this distribution leads to the ML score functions

(3.1.1) $$\partial l/\partial\boldsymbol{\beta} = \sigma^{-2}\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})$$

(3.1.2) $$\partial l/\partial\sigma^2 = -(1/2)[n\sigma^{-2} - \sigma^{-4}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})'\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})]$$

(3.1.3) $$\begin{aligned}\partial l/\partial\varphi_j = -(1/2)[&\mathrm{tr}(\boldsymbol{\Sigma}_s^{-1}\mathbf{Z}_{sj}\boldsymbol{\Omega}_j\mathbf{Z}_{sj}') \\ &- \sigma^{-2}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})'\boldsymbol{\Sigma}_s^{-1}\mathbf{Z}_{sj}\boldsymbol{\Omega}_j\mathbf{Z}_{sj}'\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})]\end{aligned} \quad \text{for } j = 1, 2, \ldots$$

(3.1.4) $$\begin{aligned}\partial l/\partial\rho_h = -(1/2)[&\mathrm{tr}(\boldsymbol{\Sigma}_s^{-1}\mathbf{Z}_s(\partial\boldsymbol{\Omega}/\partial\rho_h)\mathbf{Z}_s') \\ &- \sigma^{-2}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})'\boldsymbol{\Sigma}_s^{-1}\mathbf{Z}_s(\partial\boldsymbol{\Omega}/\partial\rho_h)\mathbf{Z}_s'\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})]\end{aligned} \quad \text{for } h = 1, 2, \ldots$$

Equating (3.1.1) to zero yields the ML estimating equations for $\boldsymbol{\beta}$. For fixed variance components vector $\boldsymbol{\omega}$, it is clear that the MLE for this parameter is then just its GLS estimator

$$\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{y}_s.$$

Equating the score functions in (3.1.2) – (3.1.4) to zero yield the ML estimating equations for σ^2, φ_j and ρ_h. For fixed φ_j and ρ_h, equating (3.1.2) to zero gives the ML estimate of σ^2 as

(3.1.3) $$\hat{\sigma}^2_{ML} = n^{-1}\mathbf{y}_s'\mathbf{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}) = n^{-1}\mathbf{y}_s'\mathbf{W}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}} - \mathbf{Z}_s\hat{\mathbf{u}})$$

where $\hat{\mathbf{u}}$ is given by (2.2) in the previous section.

The estimating equations for the variance components φ_j and ρ_h have no analytic solution, and so have to be solved numerically. Various methods for computing ML estimates of $\boldsymbol{\omega}$ have been used in literature. Henderson (1963, 1973, 1975), Harville (1977) and Fellner (1986, 1987) have proposed iterative procedures for calculating ML estimates under different normal variance components models. We adapt these methods in the following iterative algorithm for calculating the ML estimates of φ_j, σ^2 and ρ_h.

1. Assign initial values to the variance components φ_j, ρ_h and σ^2.
2. Using the current values for these variance components, calculate $\mathbf{\Sigma}_s$ and $\mathbf{\Omega}$.
3. Update $\boldsymbol{\beta} = (\mathbf{X}_s'\mathbf{\Sigma}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{\Sigma}_s^{-1}\mathbf{y}_s$.
4. Update $\mathbf{u} = \mathbf{\Omega}\mathbf{Z}_s'\mathbf{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})$.
5. Update $\sigma^2 = n^{-1}\mathbf{y}_s'\mathbf{W}_s^{-1}(\mathbf{y}_s - X_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{u})$.
6. Calculate $\mathbf{T}_s^* = (\mathbf{\Omega}^{-1} + \mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s)^{-1} = [\mathbf{T}_{sjj'}^*]$.
7. Update $\varphi_j = v_j^{-1}(\mathrm{tr}(\mathbf{T}_{sjj}^*\mathbf{\Omega}_j^{-1}) + \sigma^{-2}\mathbf{u}_j'\mathbf{\Omega}_j^{-1}\mathbf{u}_j)$ where v_j is the rank of the matrix $\mathbf{Z}_{sj}$.
8. Update $\rho_h = f_h(\boldsymbol{\rho}, \boldsymbol{\varphi}, \mathbf{T}_s^*, \sigma^2, \mathbf{u})$ where f_h is a known function whose specification depends on the parameterization of $\mathbf{\Omega}(\boldsymbol{\rho})$, and current values for variance components are used in the right hand side of this equation.
9. Return to step 2 and repeat the procedure until the values of the different parameters converge.

At convergence the MLE-based empirical best linear unbiased estimate (EBLUP) of $\boldsymbol{\theta}$ is calculated by substituting the converged values of $\boldsymbol{\beta}$ and $\mathbf{u}$ above as the corresponding estimates in (2.1) and then computing (2.3).

3.2. Residual Maximum Likelihood Estimation of the Variance Components (REML)

The Residual Maximum Likelihood (REML) approach is designed to reduce the bias of the ML estimate of the variance components vector $\boldsymbol{\omega}$. In particular, this approach starts by first transforming $\mathbf{y}_s$ into two independent vectors $\mathbf{y}_1 = \mathbf{K}_1\mathbf{y}_s$ and $\mathbf{y}_2 = \mathbf{K}_2\mathbf{y}_s$. The $\mathbf{y}_1$ vector has a distribution that does not depend on the fixed effect $\boldsymbol{\beta}$ and hence satisfies $E(\mathbf{K}_1\mathbf{y}_s) = 0$, i.e. $\mathbf{K}_1\mathbf{X} = \mathbf{0}$, while the $\mathbf{y}_2$ vector is independent of $\mathbf{y}_1$ and satisfies $\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_2' = \mathbf{0}$. The matrix $\mathbf{K}_1$ is chosen to have maximum rank, i.e. n-p, and so the rank of $\mathbf{K}_2$ is p. The likelihood function of $\mathbf{y}_s$ is then the product of the likelihoods generated by $\mathbf{y}_1$ and $\mathbf{y}_2$. Following Patterson and Thompson (1971), the REML estimators of the variance components are then the maximum likelihood estimators of these parameters based on $\mathbf{y}_1$. That is, the REML method estimates the variance components vector $\boldsymbol{\omega}$ by maximising the log-likelihood function

$$l_{REML} = -(1/2)[(n-p)\ln 2\pi\sigma^2 + |\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1| + \sigma^{-2}\mathbf{y}_s'\mathbf{K}_1(\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1)^{-1}\mathbf{K}_1\mathbf{y}_s]$$

where $\mathbf{K}_1 = \mathbf{W}_s^{-1} - \mathbf{W}_s^{-1}\mathbf{X}_s(\mathbf{X}_s'\mathbf{W}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{W}_s^{-1}$ for the model (1.5). Note that if $\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1$ is not of full rank, then $|\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1|$ must be interpreted as the determinant of its linearly independent rows and columns. Given this definition of $\mathbf{K}_1$, the matrix $\mathbf{K}_2$ is defined as $\mathbf{K}_2 = \mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}$. The log-likelihood function defined by $\mathbf{y}_2 = \mathbf{K}_2\,\mathbf{y}_s$ is

$$\begin{aligned} l_L &= -(1/2)[p\ln 2\pi\sigma^2 + \ln|\mathbf{K}_2\boldsymbol{\Sigma}_s\mathbf{K}_2'| + \sigma^{-2}(\mathbf{y}_2 - E(\mathbf{y}_2))'(\mathbf{K}_2\boldsymbol{\Sigma}_s\mathbf{K}_2')^{-1}(\mathbf{y}_2 - E(\mathbf{y}_2))] \\ &= -(1/2)[p\ln 2\pi\sigma^2 + \ln|\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s| + \sigma^{-2}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s(\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})]. \end{aligned}$$

For given values of the variance components, $\boldsymbol{\beta}$ is then estimated by maximizing l_L, leading to $\hat{\boldsymbol{\beta}} = (\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{y}_s$.

Differentiation of l_{REML} with respect to the variance components yields the REML estimating equations

$$\begin{aligned} \partial l_{\mathrm{REML}}/\partial\sigma^2 &= -(1/2)[(n-p)\sigma^{-2} - \sigma^{-4}y_s'\mathbf{K}_1(\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1)^{-1}\mathbf{K}_1\mathbf{y}_s] \\ \partial l_{\mathrm{REML}}/\partial\varphi_j &= -(1/2)[\{\mathrm{tr}(\mathbf{K}_{j\varphi}) - \sigma^{-2}y_s'\mathbf{K}_{j\varphi}\mathbf{K}_1(\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1)^{-1}\mathbf{K}_1\mathbf{y}_s] \\ \partial l_{\mathrm{REML}}/\partial\rho_h &= -(1/2)[\{\mathrm{tr}(\mathbf{K}_{h\rho}) - \sigma^{-2}y_s'\mathbf{K}_{h\rho}\mathbf{K}_1(\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1)^{-1}\mathbf{K}_1\mathbf{y}_s] \end{aligned} \tag{3.2.1}$$

where $\mathbf{K}_{j\varphi} = \mathbf{K}_1(\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1)^{-1}\mathbf{K}_1\partial\boldsymbol{\Sigma}_s/\partial\varphi_j$ and $\mathbf{K}_{h\rho} = \mathbf{K}_1(\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1)^{-1}\mathbf{K}_1\partial\boldsymbol{\Sigma}_s/\partial\rho_h$. For given values of φ_j and ρ_h, equating the REML score function for σ^2 in (3.2.1) to zero leads to a REML estimate of this parameter that satisfies

$$\begin{aligned} \hat{\sigma}^2_{\mathrm{REML}} &= (n-p)^{-1}\mathbf{y}_s'\mathbf{K}_1(\mathbf{K}_1\boldsymbol{\Sigma}_s\mathbf{K}_1)^{-1}\mathbf{K}_1\mathbf{y}_s \\ &= (n-p)^{-1}\mathbf{y}_s'\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}) \\ &= (n-p)^{-1}\mathbf{y}_s'\mathbf{W}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}} - \mathbf{Z}_s\hat{\mathbf{u}}). \end{aligned} \tag{3.2.2}$$

Equating the REML score functions for φ_j and ρ_h in (3.2.1) to zero lead to non-linear functions of φ_j and ρ_h which have to be solved using iterative techniques. Following Fellner

(1986, 1987) we define an iterative algorithm to calculate the REML estimates of the variance components in (1.5). It is identical to the algorithm for calculating the ML estimates defined earlier when one replaces $\mathbf{T}_s^*$ there by $\mathbf{T}_s$ below.

1. Assign initial values to the variance components φ_j, ρ_h and σ^2.
2. Using the current values for these variance components, calculate $\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Omega}$
3. Update $\boldsymbol{\beta} = (\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{y}_s$.
4. Update $\mathbf{u} = \boldsymbol{\Omega}\mathbf{Z}_s'\boldsymbol{\Sigma}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta})$.
5. Update $\sigma^2 = (n-p)^{-1}\mathbf{y}_s'\mathbf{W}_s^{-1}(\mathbf{y}_s - X_s\boldsymbol{\beta} - \mathbf{Z}_s\mathbf{u})$.
6. Calculate $\mathbf{T}_s^* = (\boldsymbol{\Omega}^{-1} + \mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s)^{-1} = [\mathbf{T}_{sjj'}^*]$.
7. Calculate $\mathbf{T}_s = [\mathbf{T}_{sjj'}] = \mathbf{T}_s^* + \mathbf{T}_s^*\mathbf{Z}_s'\mathbf{X}_s(\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{Z}_s\mathbf{T}_s^*$.
8. Update $\varphi_j = v_j^{-1}(\mathrm{tr}(\mathbf{T}_{sjj}\boldsymbol{\Omega}_j^{-1}) + \sigma^{-2}\mathbf{u}_j'\boldsymbol{\Omega}_j^{-1}\mathbf{u}_j)$ where v_j is the rank of the matrix $\mathbf{Z}_{sj}$.
9. Update $\rho_h = f_h(\boldsymbol{\rho},\boldsymbol{\varphi},\mathbf{T}_s,\sigma^2,\mathbf{u})$ where f_h is a known function whose specification depends on the parameterization of $\boldsymbol{\Omega}(\boldsymbol{\rho})$, and current values for variance components are used in the right hand side of this equation.
10. Return to step 2 and repeat the procedure until the values of the different parameters converge.

As with the MLE, the REML-based empirical best linear unbiased estimate (EBLUP) of $\boldsymbol{\theta}$ is then calculated by substituting the converged values of $\boldsymbol{\beta}$ and $\mathbf{u}$ above as the corresponding estimates in (2.1) and then computing (2.3).

3.3 Estimating the Mean Squared Error of the EBLUP

To start, we assume that the variance components are estimated via ML. The prediction error of $\hat{\boldsymbol{\theta}}$ is $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathbf{a}_r(\hat{\mathbf{y}}_r - \mathbf{y}_r) = \mathbf{a}_r(\hat{\boldsymbol{\eta}}_r - \mathbf{y}_r) = \mathbf{a}_r(\hat{\boldsymbol{\eta}}_r - \boldsymbol{\eta}_r + \boldsymbol{\eta}_r - \mathbf{y}_r) = \mathbf{a}_r[(\hat{\boldsymbol{\eta}}_r - \boldsymbol{\eta}_r) - (\mathbf{y}_r - \boldsymbol{\eta}_r)]$. The mean cross-product error matrix is therefore $MCPE(\hat{\boldsymbol{\theta}}) = \mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})']$. After some algebra, it can be shown that this matrix is

$$MCPE(\hat{\boldsymbol{\theta}}) \cong MCPE(\tilde{\boldsymbol{\theta}}) + \mathbf{G}_3(\boldsymbol{\omega}) = \mathbf{G}_1(\boldsymbol{\omega}) + \mathbf{G}_2(\boldsymbol{\omega}) + 2\mathbf{G}_3(\boldsymbol{\omega}) + \mathbf{G}_4(\boldsymbol{\omega}) \quad (3.3.1)$$

where $\mathbf{G}_1(\boldsymbol{\omega})$, $\mathbf{G}_2(\boldsymbol{\omega})$ and $\mathbf{G}_4(\boldsymbol{\omega})$ are given in section 2, and $\mathbf{G}_3(\boldsymbol{\omega})$ is the square matrix of the same order as $\boldsymbol{\theta}$, with (α,α') component $\sigma^2[\mathrm{tr}(\mathbf{G}_{\alpha\alpha'}\mathbf{B})]$, $\mathbf{B}$ is the asymptotic covariance matrix of the ML estimates of the variance components $\boldsymbol{\gamma} = (\hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\rho}}')'$ and, taking $\tilde{\theta}_\alpha$ to be α^{th} element of the BLUP $\tilde{\boldsymbol{\theta}}$, $\mathbf{G}_{\alpha\alpha'} = \mathrm{Cov}(\frac{\partial\tilde{\theta}_\alpha}{\partial\boldsymbol{\gamma}}, \frac{\partial\tilde{\theta}_{\alpha'}}{\partial\boldsymbol{\gamma}})$. Note that **B** is defined by the "$\boldsymbol{\gamma}$-component" of the inverse of the Fisher information matrix for the variance components $\boldsymbol{\omega}$. Under the model (1.5) this information matrix is

$$\frac{1}{2}\begin{bmatrix} n\sigma^{-4} & \sigma^{-2}[\varphi_j^{-1}(v_j - r_j^*)] & \sigma^{-2}[\sum_{j=1}^{J}(-v_j^{(h)} + r_j^{*(h)})] \\ & [\varphi_j^{-2}\{(v_j - 2r_j^*)\delta_{jj'} + \varphi_j^{-2}\varphi_{j'}^{-2}r_{jj'}^*\}] & [\varphi_j^{-1}(2r_j^{*(h)} - v_j^{(h)} - \sum_{j'=1}^{J}\varphi_j^{-1}\varphi_{j'}^{-1}r_{jj'}^{*(h)}] \\ & & [\sum_{j=1}^{J}(\{\sum_{j'=1}^{J}\varphi_j^{-1}\varphi_{j'}^{-1}r_{jj'}^{*(hh')}\} + v_j^{(hh')} - 2r_j^{*(hh')})] \end{bmatrix}$$

where δ_{ij} is a Kronecker delta,

$$\begin{aligned} r_j^* &= \varphi_j^{-1}\mathrm{tr}(\boldsymbol{\Omega}_j^{-1}\mathbf{T}_{sjj}^*), \\ r_{jj'}^* &= \mathrm{tr}(\mathbf{T}_{sj'j}^*\boldsymbol{\Omega}_j^{-1}\mathbf{T}_{sjj'}^*\boldsymbol{\Omega}_{j'}^{-1}), \\ r_j^{*(h)} &= \mathrm{tr}[\varphi_j^{-1}(\partial\boldsymbol{\Omega}_j^{-1}/\partial\rho_h)\mathbf{T}_{sjj}^*, \\ v_j^{(h)} &= \mathrm{tr}[(\partial\boldsymbol{\Omega}_j^{-1}/\partial\rho_h)\boldsymbol{\Omega}_j], \qquad (3.3.2) \\ r_{jj'}^{*(h)} &= \mathrm{tr}[\mathbf{T}_{sjj'}^*(\partial\boldsymbol{\Omega}_{j'}^{-1}/\partial\rho_h)\mathbf{T}_{sj'j}^*\boldsymbol{\Omega}_j^{-1}], \\ v_j^{(hh')} &= \mathrm{tr}[(\partial\boldsymbol{\Omega}_j^{-1}/\partial\rho_h)\boldsymbol{\Omega}_j(\partial\boldsymbol{\Omega}_j^{-1}/\partial\rho_{h'})\boldsymbol{\Omega}_j], \\ r_{jj'}^{*(hh')} &= \mathrm{tr}[\mathbf{T}_{sjj'}^*(\partial\boldsymbol{\Omega}_{j'}^{-1}/\partial\rho_h)\mathbf{T}_{sj'j}^*(\partial\boldsymbol{\Omega}_j^{-1}/\partial\rho_{h'})], \\ r_j^{*(hh')} &= \varphi_j^{-1}\mathrm{tr}[(\partial\boldsymbol{\Omega}_j^{-1}/\partial\rho_h)\mathbf{T}_{sjj}^*(\partial\boldsymbol{\Omega}_j^{-1}/\partial\rho_{h'})\boldsymbol{\Omega}_j] \end{aligned}$$

and ν_j is the rank of the matrix $\mathbf{Z}_{sj}$. Recollect that $\mathbf{X}_r^* = \mathbf{a}_r\mathbf{X}_r$, $\mathbf{Z}_r^* = \mathbf{a}_r\mathbf{Z}_r$, $\mathbf{X}_s^* = \mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{X}_s$ and $y_s^* = \mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{y}_s$. The EBLUP of $\boldsymbol{\theta}$ can then be written

$$\hat{\boldsymbol{\theta}} = \mathbf{a}_s\mathbf{y}_s + \mathbf{X}_r^*\hat{\boldsymbol{\beta}} + \mathbf{Z}_r^*\mathbf{T}_s^*(\mathbf{y}_s^* - \mathbf{X}_s^*\hat{\boldsymbol{\beta}}). \qquad (3.3.3)$$

Let $\boldsymbol{\Delta} = [\boldsymbol{\Delta}_1', \boldsymbol{\Delta}_2', ..., \boldsymbol{\Delta}_n']'$ be the coefficient matrix of $(\mathbf{y}_s^* - \mathbf{X}_s^*\hat{\boldsymbol{\beta}})$ in (3.3.3), i.e., $\boldsymbol{\Delta} = \mathbf{Z}_r^*\mathbf{T}_s^*$ and, when $\mathbf{Z}_\alpha^*$ is the α^{th} row of the matrix $\mathbf{Z}_r^*$, put $\partial\boldsymbol{\Delta}_\alpha/\partial\boldsymbol{\gamma} = \partial(\mathbf{Z}_\alpha^*\mathbf{T}_s^*)/\partial\boldsymbol{\gamma} = \nabla_\alpha$. The mean cross-product error matrix of the EBLUP $\hat{\boldsymbol{\theta}}$ then has $\mathbf{G}_3(\boldsymbol{\omega}) = \sigma^2[\mathrm{tr}(\boldsymbol{\nabla}_\alpha\boldsymbol{\Sigma}_s^*\boldsymbol{\nabla}_{\alpha'}'\mathbf{B})]$ where $\boldsymbol{\Sigma}_s^* = \mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s + \mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s\boldsymbol{\Omega}\mathbf{Z}_s'\mathbf{W}_s^{-1}\mathbf{Z}_s$.

For model (1.5) we have $\boldsymbol{\nabla}_\alpha = -(\mathbf{Z}_\alpha^*\mathbf{T}_s^* \otimes \mathbf{I}_{J+H})(\partial\boldsymbol{\Omega}^{-1}/\partial\boldsymbol{\gamma})\mathbf{T}_s^*$ where $\otimes$ denotes direct product and $\mathbf{I}_{J+H}$ denotes an identity matrix of dimension equal to $J+H$. An estimate of the

$MCPE$ matrix of the EBLUP $\hat{\boldsymbol{\theta}}$ is therefore $M\hat{C}\hat{P}E(\hat{\boldsymbol{\theta}}) = \mathbf{G}_1(\hat{\boldsymbol{\omega}}) + \mathbf{G}_2(\hat{\boldsymbol{\omega}}) + 2\mathbf{G}_3(\hat{\boldsymbol{\omega}}) + \mathbf{G}_4(\hat{\boldsymbol{\omega}})$, where $\hat{\boldsymbol{\omega}}$ is the ML estimate of the variance components vector $\boldsymbol{\omega}$.

The preceding development still holds when the variance components in (1.5) are estimated via REML. The only change is that in this case the Fisher information matrix of the REML estimators $\hat{\sigma}^2, \hat{\boldsymbol{\varphi}}$ and $\hat{\boldsymbol{\rho}}$ is now of the form

$$\frac{1}{2}\begin{bmatrix} (n-p)\sigma^{-4} & \sigma^{-2}[\varphi_j^{-1}(\nu_j - r_j)] & \sigma^{-2}[\sum_{j=1}^{J}(-v_j^{(h)} + r_j^{(h)})] \\ & [\varphi_j^{-2}\{(\nu_j - 2r_j)\delta_{jj'} + \varphi_j^{-2}\varphi_{j'}^{-2} r_{jj'}\}] & [\varphi_j^{-1}(2r_j^{(h)} - v_j^{(h)} - \sum_{j'=1}^{J}\varphi_j^{-1}\varphi_{j'}^{-1} r_{jj'}^{(h)}] \\ & & [\sum_{j=1}^{J}(\{\sum_{j'=1}^{J}\varphi_j^{-1}\varphi_{j'}^{-1} r_{jj'}^{(hh')}\} + v_j^{(hh')} - 2r_j^{(hh')})] \end{bmatrix}$$

where ν_j, $v_j^{(h)}$, $v_j^{(hh')}$ are defined in (3.3.2) and r_j, $r_j^{(h)}$, $r_j^{(hh')}$, $r_{jj'}^{(h)}$ and $r_{jj'}^{(hh')}$ are obtained by simply by deleting the star (*) from all notation used in (3.3.2).

4 Application to Unit Level Linear Mixed Models

In this section we illustrate the application of the preceding theory to a number of commonly used unit level linear mixed models. These are appropriate when individual level data (both for Y and $\mathbf{X}$) are available within the small areas.

4.1 A Model with IID Area Effects and IID Time Effects

A special case of (1.1) is a linear mixed model with independent and identically distributed (*iid*) area effects and corresponding *iid* time effects. Note that this corresponds to an assumption of a constant covariance between two observations from the same small area at two different points in time. In term of the general model (1.1), we have two random components, an area effect and a time effect. The model in this special case is

(4.1.1) $$y_{dti} = \beta_0 + \mathbf{X}'_{dti}\boldsymbol{\beta}_t + u_{1t} + u_{2d} + e_{dti}.$$

The u_{1t} and u_{2d} are assumed to be mutually independent and to follow normal distributions with zero means and respective variances. The e_{dti} is a random error, independent of random components u_{1t} and u_{2d} and is assumed to follow a normal distribution with zero mean and variance σ^2. Let $\mathbf{u}_1 = [u_{11}, u_{12},...u_{1T}]'$, $\mathbf{u}_2 = [u_{21}, u_{22},...u_{2D}]'$ and $\mathbf{e} = [e_{dti}]$ and let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ to be the incidence matrices for the random effect vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ respectively. The model (4.1.1) can then be written in matrix form as

(4.1.2) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}$.

Here $\mathbf{Z}_1 = \mathbf{Z}_1^* // \mathbf{Z}_2^* // ... // \mathbf{Z}_D^*$ where // denotes the "stacking" operator, with $\mathbf{Z}_d^* = diag(\mathbf{1}_{N_{dt}}; t = 1, \cdots, T)$, where $\mathbf{1}_{N_{dt}}$ is a vector of dimension N_{dt} with all elements equal to one. Similarly, $\mathbf{Z}_2 = diag(\mathbf{1}_{N_d}; d = 1, \cdots, D)$, where $N_d = \sum_{t=1}^{T} N_{dt}$. The random vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{e}$ are assumed to be distributed as multivariate normal with zero mean vectors and variance-covariance matrices given by $\sigma_1^2\mathbf{I}_T$, $\sigma_2^2\mathbf{I}_D$ and $\sigma^2\mathbf{I}_N$ respectively, where N is the sum of the population sizes at t = 1, 2, …, T and the matrices $\mathbf{I}_D$, $\mathbf{I}_N$ and $\mathbf{I}_T$ are identity matrices with dimensions equal to D, N and T respectively. In the context of the notation introduced in section 1, we therefore have $J = 2$, $\varphi_1 = \frac{\sigma_1^2}{\sigma^2}$, $\varphi_2 = \frac{\sigma_2^2}{\sigma^2}$, $\boldsymbol{\Omega}_1 = \mathbf{I}_T$, $\boldsymbol{\Omega}_2 = \mathbf{I}_D$, $\mathbf{W} = \mathbf{I}_N$ and therefore $\boldsymbol{\Omega} = \begin{bmatrix} \varphi_1\mathbf{I}_T & \mathbf{0} \\ \mathbf{0} & \varphi_2\mathbf{I}_D \end{bmatrix}$ and $\boldsymbol{\Sigma} = \mathbf{I}_N + \varphi_1\mathbf{Z}_1\mathbf{Z}_1' + \varphi_2\mathbf{Z}_2\mathbf{Z}_2'$. Note that there is no parameter $\boldsymbol{\rho}$. Setting $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, (4.1.2) further simplifies to

(4.1.3) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$

with corresponding model for the sample values $\mathbf{y}_s$ given by

(4.1.4) $\mathbf{y}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s$.

In this case $\boldsymbol{\Sigma}_s = \mathbf{I}_n + \varphi_1\mathbf{Z}_{s1}\mathbf{Z}_{s1}' + \varphi_2\mathbf{Z}_{s2}\mathbf{Z}_{s2}'$, where n denotes the overall number of sample values of Y (across all times and all areas) and $\mathbf{Z}_{sj}$ denotes the sample version of $\mathbf{Z}_j$.

The ML and REML estimation algorithms described in sections 3.1 and 3.2 and the MCPE estimator described in section 3.3 can then be used (with these versions of $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_s$) to calculate the EBLUPs of any set of linear combinations of the values in the population vector $\mathbf{y}$. Note that in both the ML and REML estimation algorithms we ignore the step that updates ρ_h since this parameter does not exist in the model (4.1.1). We also have $\nu_1 = T$ and $\nu_2 = D$.

The observed information matrix for the ML estimates of $\boldsymbol{\omega} = (\sigma^2, \varphi_1, \varphi_2)'$ is then

$$\hat{\mathbf{I}}_{ML} = \frac{1}{2}\begin{bmatrix} n\hat{\sigma}^{-4} & \hat{\sigma}^{-2}[\hat{\varphi}_j^{-1}(\nu_j - \hat{r}_j^*)] \\ & [\hat{\varphi}_j^{-2}\{(\nu_j - 2\hat{r}_j^*)\delta_{jj'} + \hat{\varphi}_j^{-2}\hat{\varphi}_{j'}^{-2}\hat{r}_{jj'}^*\}] \end{bmatrix}$$

where a "hat" denotes an ML estimate, $\delta_{jj'}$ = Kroneker delta, $\hat{r}_j^* = \hat{\varphi}_j^{-1}\mathrm{tr}(\hat{\mathbf{T}}_{sjj'}^*)$, for j = 1, 2 and $\hat{r}_{jj'}^* = \mathrm{tr}(\hat{\mathbf{T}}_{sj'j}^*\hat{\mathbf{T}}_{sjj'}^*)$. Note that the estimate of the matrix $\mathbf{B}$ needed to evaluate the $\mathbf{G}_3$ term of the MCPE matrix of $\hat{\boldsymbol{\theta}}$ in this case is obtained from the 2 × 2 submatrix in the bottom right hand corner of the inverse of $\hat{\mathbf{I}}_{ML}$ above. The estimated MCPE matrix for the ML-based $\hat{\boldsymbol{\theta}}$ is therefore $M\hat{C}\hat{P}E(\hat{\boldsymbol{\theta}}) = \mathbf{G}_1(\hat{\boldsymbol{\omega}}) + \mathbf{G}_2(\hat{\boldsymbol{\omega}}) + 2\mathbf{G}_3(\hat{\boldsymbol{\omega}}) + \mathbf{G}_4(\hat{\boldsymbol{\omega}})$, where

$$\mathbf{G}_1(\hat{\boldsymbol{\omega}}) = \sigma^2 \mathbf{Z}_r^* \hat{\mathbf{T}}_s^* \mathbf{Z}_r^{*\prime}$$

$$\mathbf{G}_2(\hat{\boldsymbol{\omega}}) = \hat{\sigma}^2 [\mathbf{X}_r^* - \mathbf{Z}_r^* \hat{\mathbf{T}}_s^* \mathbf{Z}_s' \mathbf{X}_s](\mathbf{X}_s' \hat{\boldsymbol{\Sigma}}_s^{-1} \mathbf{X}_s)^{-1} [\mathbf{X}_r^{*\prime} - \mathbf{X}_s' \mathbf{Z}_s \hat{\mathbf{T}}_s^* \mathbf{Z}_r^{*\prime}]$$

$$\mathbf{G}_4(\hat{\boldsymbol{\omega}}) = \hat{\sigma}^2 \mathbf{a}_r \mathbf{a}_r'$$

$$\mathbf{G}_3(\hat{\boldsymbol{\omega}}) = \hat{\sigma}^2 [\mathrm{tr}(\hat{\boldsymbol{\nabla}}_\alpha \hat{\boldsymbol{\Sigma}}_s^* \hat{\boldsymbol{\nabla}}_{\alpha'}' \hat{\mathbf{B}})]$$

$$\hat{\boldsymbol{\Sigma}}_s^* = \mathbf{Z}_s' \mathbf{Z}_s + \mathbf{Z}_s' \mathbf{Z}_s \hat{\boldsymbol{\Omega}} \mathbf{Z}_s' \mathbf{Z}_s$$

$$\hat{\boldsymbol{\nabla}}_\alpha = -(\mathbf{Z}_\alpha^* \hat{\mathbf{T}}_s^* \otimes \mathbf{I}_2)(\partial \hat{\boldsymbol{\Omega}}^{-1} / \partial \hat{\boldsymbol{\gamma}}) \hat{\mathbf{T}}_s^*$$

$$\partial \hat{\boldsymbol{\Omega}}^{-1} / \partial \hat{\boldsymbol{\gamma}} = -diag\left\{ \hat{\varphi}_1^{-2} \mathbf{I}_T \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ \hat{\varphi}_2^{-2} \mathbf{I}_D \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

Here $\hat{\boldsymbol{\theta}} = (\hat{\theta}_\alpha)$, $\hat{\boldsymbol{\omega}} = (\hat{\sigma}^2, \hat{\varphi}_1, \hat{\varphi}_2)'$ and $\hat{\boldsymbol{\gamma}} = (\hat{\varphi}_1, \hat{\varphi}_2)'$. The only change when calculating the estimate of the MCPE matrix for the REML-based $\hat{\boldsymbol{\theta}}$ is the estimate of **B**. This is given by the 2 × 2 submatrix in the bottom right hand corner of the inverse of the estimated information matrix for the REML estimates of the variance components, which itself is defined by

$$\hat{\mathbf{I}}_{REML} = \frac{1}{2} \begin{bmatrix} (n-p)\hat{\sigma}^{-4} & \hat{\sigma}^{-2}[\hat{\varphi}_j^{-1}(\nu_j - \hat{r}_j)] \\ & [\hat{\varphi}_j^{-2}\{(\nu_j - 2\hat{r}_j)\delta_{jj'} + \hat{\varphi}_j^{-2}\hat{\varphi}_{j'}^{-2}\hat{r}_{jj'}\}] \end{bmatrix}$$

where $\hat{r}_j = \hat{\varphi}_j^{-1} \mathrm{tr}(\hat{\mathbf{T}}_{sjj})$, for j = 1, 2 and $\hat{r}_{jj'} = \mathrm{tr}(\hat{\mathbf{T}}_{sj'j} \hat{\mathbf{T}}_{sjj'})$.

A common application is to predict the population totals of Y in each of the small areas at time T, in which case $\boldsymbol{\theta} = \mathbf{ay}$ where $\mathbf{a} = diag[\mathbf{a}_d'; d = 1,2,\cdots,D]$, with $\mathbf{a}_d$ the N_d vector with zeros everywhere except for the last N_{dT} of its elements, which are one.

4.2 A Model with IID Area Effects and Autocorrelated Time Effects

In the previous section we assumed a model that implied a constant correlation between observations at different points in time in the same area. A more reasonable assumption is that the time effect is autocorrelated in some way. The simplest model for this type of behaviour is to allow this effect to be the realisation of a first order autoregressive AR(1) process. That is, the specifications (4.1.1) and (4.1.2) are unchanged, but now $\mathbf{u}_1 = [u_{11}, u_{12}, \ldots u_{1T}]'$ represents the outcome of an AR(1) process. In particular, this means that this vector is distributed as multivariate normal with zero mean vector and covariance matrix proportional to

$$\mathbf{\Omega}_1 = \left(\frac{1}{1-\rho^2}\right)\begin{bmatrix} 1 & \rho & . & . & \rho^{\mathrm{T}-1} \\ \rho & 1 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \rho \\ \rho^{\mathrm{T}-1} & . & . & \rho & 1 \end{bmatrix}$$

where ρ is the correlation parameter. The inverse of this matrix is $\mathbf{\Omega}_1^{-1} = I_T + \rho^2\mathbf{E} + \rho\mathbf{F}$, where $\mathbf{E}$ is a diagonal matrix of dimension T with the diagonal elements of (0,1,1,...1,0) and $\mathbf{F}$ is a $T \times T$ matrix with elements on the principle diagonal equal to zero, elements in the diagonals immediately above and below the principle diagonal equal to minus one and all other elements zero. In this case $\mathbf{\Omega} = \begin{bmatrix} \varphi_1\mathbf{\Omega}_1 & \mathbf{0} \\ \mathbf{0} & \varphi_2\mathbf{I}_D \end{bmatrix}$ so $\mathbf{\Sigma}_s = \mathbf{I}_n + \varphi_1\mathbf{Z}_1\mathbf{\Omega}_1\mathbf{Z}_1' + \varphi_2\mathbf{Z}_2\mathbf{Z}_2'$.

ML and REML estimation of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\omega} = (\sigma^2, \varphi_1, \varphi_2, \rho)'$ of this model follows along the same lines as in section 4.1 above. The main difference is that in this case H = 1 (so we drop the h subscript) and we need to update the estimate of ρ in the iterative estimation process. This is done by replacing step 8 in the ML algorithm (see section 3.1) by the identity

$$\rho = -\frac{\varphi_1^{-1}[\mathrm{tr}(\mathbf{T}_{s11}^*\mathbf{F}) + \sigma^{-2}\mathbf{u}_1'\mathbf{F}\mathbf{u}_1]}{\{(2/(1-\rho^2)) + 2\varphi_1^{-1}[\mathrm{tr}(\mathbf{T}_{s11}^*\mathbf{E}) + \sigma^{-2}\mathbf{u}_1'\mathbf{E}\mathbf{u}_1]\}}.$$

Step 9 in the corresponding REML algorithm (see section 3.2) also uses this identity, but with $\mathbf{T}_{s11}^*$ replaced by $\mathbf{T}_{s11}$.

Estimation of the MCPE of the small area estimates requires the Fisher information matrix for the estimated variance components $\hat{\boldsymbol{\omega}} = (\hat{\sigma}^2, \hat{\varphi}_1, \hat{\varphi}_2, \hat{\rho})'$

$$\hat{\mathbf{I}}_{ML} = \frac{1}{2}\begin{bmatrix} n\hat{\sigma}^{-4} & \hat{\sigma}^{-2}\hat{\varphi}_1^{-1}(\nu_1 - \hat{r}_1^*) & \hat{\sigma}^{-2}\hat{\varphi}_2^{-1}(\nu_2 - \hat{r}_2^*) & \hat{\sigma}^{-2}(\hat{r}_1^{*(1)} - v_1^{(1)}) \\ & \hat{\varphi}_1^{-2}(\nu_1 - 2\hat{r}_1^*) + \hat{\varphi}_1^{-4}\hat{r}_{11}^* & \hat{\varphi}_1^{-2}\hat{\varphi}_2^{-2}\hat{r}_{12}^* & \hat{\varphi}_1^{-1}(2\hat{r}_1^{*(1)} - v_1^{(1)} - \hat{\varphi}_1^{-1}\hat{r}_{11}^{*(1)}) \\ & & \hat{\varphi}_2^{-2}(\nu_2 - 2\hat{r}_2^*) + \hat{\varphi}_2^{-4}\hat{r}_{22}^* & -\hat{\varphi}_1^{-1}\hat{\varphi}_2^{-2}\hat{r}_{12}^{*(1)} \\ & & & (\hat{\varphi}_1^{-2}\hat{r}_{11}^{*(11)} + v_1^{(11)} - 2\hat{r}_1^{*(11)}) \end{bmatrix}$$

where the individual terms are defined in (3.3.2). In order to calculate them we replace unknown parameter values by ML estimates and make use of the fact that $\partial\hat{\mathbf{\Omega}}_1^{-1}/\partial\hat{\rho}$ = $2\hat{\rho}\mathbf{E} + \mathbf{F}$ in this case. The matrix $\mathbf{B}$ can be calculated as the 3 × 3 submatrix in the bottom right hand corner of the inverse of the corresponding estimate of this information matrix. Estimation of the MCPE matrix of the MLE-based EBLUP $\hat{\boldsymbol{\theta}}$ under this model then follows along exactly the same lines as in section 4.1, the only difference being that we now calculate,

$$\hat{\mathbf{\nabla}}_\alpha = -(\mathbf{Z}_\alpha^*\hat{\mathbf{T}}_s^* \otimes \mathbf{I}_3)(\partial\hat{\mathbf{\Omega}}^{-1}/\partial\hat{\boldsymbol{\gamma}})\hat{\mathbf{T}}_s^*$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\varphi}_1,\ \hat{\varphi}_2,\ \hat{\rho})'$, and

$$\partial\hat{\boldsymbol{\Omega}}^{-1}/\partial\hat{\boldsymbol{\gamma}} = -diag\{[(\hat{\varphi}_1^{-2}\hat{\boldsymbol{\Omega}}_1^{-1} \otimes \mathbf{E}_1) - \hat{\varphi}_1^{-1}(\partial\hat{\boldsymbol{\Omega}}_1^{-1}/\partial\hat{\rho}) \otimes \mathbf{E}_3],\ \ \hat{\varphi}_2^{-2}\mathbf{I}_D \otimes \mathbf{E}_2\}$$

where

$$\mathrm{E}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix},\ \mathrm{E}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix},\ \mathrm{E}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

The only change when calculating the corresponding estimate of the MCPE matrix for the REML-based version of $\hat{\boldsymbol{\theta}}$ is estimation of **B**. This is calculated as the 3 × 3 submatrix in the bottom right hand corner of the inverse of the estimated Fisher information matrix for the REML estimates of the variance components under this model, given by

$$\hat{\mathbf{I}}_{REML} = \frac{1}{2}\begin{bmatrix}(n-p)\hat{\sigma}^{-4} & \hat{\sigma}^{-2}\hat{\varphi}_1^{-1}(v_1-\hat{r}_1) & \hat{\sigma}^{-2}\hat{\varphi}_2^{-1}(v_2-\hat{r}_2) & \hat{\sigma}^{-2}(\hat{r}_1^{(1)} - v_1^{(1)}) \\ & \hat{\varphi}_1^{-2}(v_1-2\hat{r}_1)+\hat{\varphi}_1^{-4}\hat{r}_{11} & \hat{\varphi}_1^{-2}\hat{\varphi}_2^{-2}\hat{r}_{12} & \hat{\varphi}_1^{-1}(2\hat{r}_1^{(1)} - v_1^{(1)} - \hat{\varphi}_1^{-1}\hat{r}_{11}^{(1)}) \\ & & \hat{\varphi}_2^{-2}(v_2-2\hat{r}_2)+\hat{\varphi}_2^{-4}\hat{r}_{22} & -\hat{\varphi}_1^{-1}\hat{\varphi}_2^{-2}\hat{r}_{12}^{(1)} \\ & & & (\hat{\varphi}_1^{-2}\hat{r}_{11}^{(11)} + v_1^{(11)} - 2\hat{r}_1^{(11)})\end{bmatrix}$$

where we refer again to (3.3.2) for definitions of the various components of this matrix, and note that this requires removal of all star (*) superscripts in (3.3.2).

4.3 A Model with a Time Varying Area Effect

The AR(1) model in introduced in section 4.2 requires a reasonable number of observations from different times to obtain an efficient estimate of the correlation parameter ρ. If there are only data available for a few time periods then the ML and REML estimates of this correlation parameter will be negatively biased. Furthermore, the model assumes that the area effects do not themselves evolve over time, with the only change over time arising because of the evolving (global) time effect. An alternative AR(1) model that uses more information and reduces biases and allows area random effects to vary over time is therefore of interest. Such a model is specified by

(4.3.1) $$y_{dti} = \beta_0 + \mathbf{X}'_{dti}\boldsymbol{\beta}_t + u_{dt} + e_{dti}$$

where the $\{u_{dt}\}$ are random effects that follow independent AR(1) processes for $d = 1, 2,\ldots,D$. As usual, the area level random effects u_{dt} and the individual level random effects e_{dti} are assumed to be mutually independent and to follow normal distributions with zero means and with variances and covariances as specified below. Let $\mathbf{u} = [u_{11},\ u_{12},\ldots u_{dt}\ ,\ldots,\ u_{DT}]'$, and

$\mathbf{e} = [e_{dti}]$ and let $\mathbf{Z}$ to be the incidence matrix for the random effect vector $\mathbf{u}$. The model (4.3.1) can then be written in matrix form as

(4.3.2) $\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}$.

The random vectors $\mathbf{u}$ and $\mathbf{e}$ are distributed as independent multivariate normal with zero mean vectors and covariance matrices given by $\sigma_u^2\mathbf{\Omega}_1$ and $\sigma^2\mathbf{I}_N$ respectively. Assuming that the random effects for the same area and different points in time can be modelled as a realisation of an AR(1) process, and the same process applies in all areas, the matrix $\mathbf{\Omega}_1$ is then

$$\mathbf{\Omega}_1 = \left(\frac{1}{1-\rho^2}\right)\mathbf{I}_D \otimes \begin{bmatrix} 1 & \rho & . & . & \rho^{\mathrm{T}-1} \\ \rho & 1 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \rho \\ \rho^{\mathrm{T}-1} & . & . & \rho & 1 \end{bmatrix}$$

where ρ is the (common) autocorrelation parameter. The matrix $\mathbf{\Omega}_1$ is a block-diagonal matrix with $\mathbf{\Omega}_1^{-1} = diag(\mathbf{I}_T + \rho^2\mathbf{E} + \rho\mathbf{F})$, where $\mathbf{E}$ and $\mathbf{F}$ were defined in section 4.2. The variance-covariance of $\mathbf{y}$ is then $\sigma^2\mathbf{I}_N + \sigma_u^2\mathbf{Z\Omega}_1\mathbf{Z}'$ = $\sigma^2\mathbf{\Sigma}$ where $\mathbf{\Sigma} = \mathbf{I}_N + \mathbf{Z\Omega Z}'$, $\mathbf{\Omega} = \varphi\mathbf{\Omega}_1$ and $\varphi = \sigma_u^2/\sigma^2$. The corresponding model for the sample values is therefore

(4.3.3) $\mathbf{y}_s = \mathbf{X}_s\mathbf{\beta} + \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s$

with the covariance of $\mathbf{y}_s$ given by $\sigma^2\mathbf{\Sigma}_s$ where $\mathbf{\Sigma}_s = \mathbf{I}_n + \mathbf{Z}_s\mathbf{\Omega Z}_s'$. Here n is the total sample size.

Again, we observe that ML and REML estimation of the parameters $\mathbf{\beta}$ and $\mathbf{\omega} = (\sigma^2, \varphi, \rho)'$ of this model follow the same lines as in section 4.1. In this case we update the estimate of ρ in the iterative estimation process by replacing step 8 in the ML algorithm (see section 3.1) by the identity

$$\rho = -\frac{\varphi^{-1}[\mathrm{tr}(\mathbf{T}_s^*\mathbf{F}^*) + \sigma^{-2}\mathbf{u}'\mathbf{F}^*\mathbf{u}]}{\{(2D/(1-\rho^2) + 2\varphi^{-1}[\mathrm{tr}(\mathbf{T}_s^*\mathbf{E}^*) + \sigma^{-2}\mathbf{u}'\mathbf{E}^*\mathbf{u}]\}}.$$

Here $\mathbf{E}^*$ and $\mathbf{F}^*$ are block-diagonal matrices of order D made up of blocks $\mathbf{E}$ and $\mathbf{F}$ respectively. Step 9 in the corresponding REML algorithm (see section 3.2) also uses this identity, but with $\mathbf{T}_s^*$ replaced by $\mathbf{T}_s$, where $\mathbf{T}_s = \mathbf{T}_s^* + \mathbf{T}_s^*\mathbf{Z}_s'\mathbf{X}_s(\mathbf{X}_s'\mathbf{\Sigma}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{Z}_s\mathbf{T}_s^*$.

Turning now to estimation of the MCPE of the small area estimates, we note that the estimated Fisher information matrix for the MLEs of the variance components $\hat{\boldsymbol{\omega}} = (\hat{\sigma}^2, \hat{\varphi}, \hat{\rho})'$ in this case is

$$\hat{\mathbf{I}}_{ML} = \frac{1}{2}\begin{bmatrix} n\hat{\sigma}^{-4} & \hat{\sigma}^{-2}\hat{\varphi}^{-1}(v - \hat{r}_1^*) & \hat{\sigma}^{-2}(\hat{r}_1^{*(1)} - v_1^{(1)}) \\ & \hat{\varphi}^{-2}(v - 2\hat{r}_1^*) + \hat{\varphi}^{-4}\hat{r}_{11}^* & \hat{\varphi}^{-1}(2\hat{r}_1^{*(1)} - v_1^{(1)} - \hat{\varphi}^{-1}\hat{r}_{11}^{*(1)}) \\ & & (\hat{\varphi}^{-2}\hat{r}_{11}^{*(11)} + v_1^{(11)} - 2\hat{r}_1^{*(11)}) \end{bmatrix}$$

where (3.3.2) provides definitions of the individual terms. This matrix can be evaluated by replacing unknown parameter values by ML estimates and making use of the fact that $\partial\hat{\boldsymbol{\Omega}}_1^{-1}/\partial\hat{\rho} = 2\hat{\rho}\mathbf{E}^* + \mathbf{F}^*$ in this case. As usual, we estimate of **B** by the 2 × 2 submatrix in the bottom right hand corner of the inverse of this estimated information matrix, with estimation of the MCPE matrix of the MLE-based $\hat{\boldsymbol{\theta}}$ under this model then as in section 4.1, except that

$$\hat{\boldsymbol{\nabla}}_\alpha = -(\mathbf{Z}_\alpha^*\hat{\mathbf{T}}_s^* \otimes \mathbf{I}_2)(\partial\hat{\boldsymbol{\Omega}}^{-1}/\partial\hat{\boldsymbol{\gamma}})\hat{\mathbf{T}}_s^*$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\varphi}, \hat{\rho})'$, and

$$\partial\hat{\boldsymbol{\Omega}}^{-1}/\partial\hat{\boldsymbol{\gamma}} = -diag\left\{\hat{\varphi}^{-2}\hat{\boldsymbol{\Omega}}_1^{-1} \otimes \begin{bmatrix}1\\0\end{bmatrix} - \hat{\varphi}^{-1}(\partial\hat{\boldsymbol{\Omega}}_1^{-1}/\partial\hat{\rho}) \otimes \begin{bmatrix}0\\1\end{bmatrix}\right\}.$$

The corresponding estimate of the MCPE of the REML-based EBLUP $\hat{\boldsymbol{\theta}}$ differs only in computation of the estimate of the matrix **B**. Here this is obtained as the 2 × 2 submatrix in the bottom right hand corner of the inverse of the estimated Fisher information matrix for the REML estimates of the variance components, which is given by

$$\hat{\mathbf{I}}_{REML} = \frac{1}{2}\begin{bmatrix} (n-p)\hat{\sigma}^{-4} & \hat{\sigma}^{-2}\hat{\varphi}^{-1}(v - \hat{r}_1) & \hat{\sigma}^{-2}(\hat{r}_1^{(1)} - v_1^{(1)}) \\ & \hat{\varphi}^{-2}(v - 2\hat{r}_1) + \hat{\varphi}^{-4}\hat{r}_{11} & \hat{\varphi}^{-1}(2\hat{r}_1^{(1)} - v_1^{(1)} - \hat{\varphi}^{-1}\hat{r}_{11}^{(1)}) \\ & & (\hat{\varphi}^{-2}\hat{r}_{11}^{(11)} + v_1^{(11)} - 2\hat{r}_1^{(11)}) \end{bmatrix}.$$

As usual, (3.3.2) provides definitions of individual terms, after star (*) superscripts used there are discarded.

4.4 A Model with Spatial Correlated Area Effects

So far all the models we have considered have assumed that area effects from any two different areas are independent. This assumption may be hard to justify in some situations, particularly when there are underlying variables common to different small areas that can be expected to lead to some spatial correlation. A linear model with spatially correlated area effects is a special case of (1.1). Since we are not concerned with data from different time points in this case, we drop the 't' subscript and assume all our data relate to a single time point. Generalization to the models with both spatially and temporally correlated random effects is straightforward.

Let y_{di} denote the i^{th} population value for a characteristic of interest within an area d (i = 1,2,…N_i; d = 1,2,…D). The vector $\mathbf{x}_{di}$ represents the corresponding values of auxiliary population information (covariates). The objective is to estimate/predict the value of the small area characteristic $\boldsymbol{\theta}$ which is a linear function of the population values y_{di}. Let u_d be the d^{th} area effect, y_{di} be the population response variable. The model of interest is then

(4.4.1) $$y_{di} = \beta_0 + \mathbf{x}'_{di}\boldsymbol{\beta} + u_d + e_{di}$$

where β_0 is an intercept, $\boldsymbol{\beta}$ is a vector of regression coefficients, the e_{di} are independent random errors with $E(e_{di}) = 0$ and $Var(e_{di}) = \sigma^2$ and the u_d are normally distributed variables with zero mean and covariances given by

(4.4.2) $$Cov(u_d, u_{d'}) = \sigma_u^2 f(Dist(d,d');\rho)$$

where ρ is an unknown parameter and $Dist(d,d')$) is an appropriate measure of the "distance" between areas d and d'.

Let $\mathbf{y}$ = $\{y_{di}\}$ be the vector of population values of the response variable, with $\mathbf{y}_s$ denoting the values observed in the sample**,** $\mathbf{X}$ =$\{\mathbf{x}_{di}\}$ be the matrix of regression variables (covariates), with $\mathbf{X}_s$ denoting the corresponding sample values, and **u** and **e** the random area effect and error vectors respectively. Let **Z** be the incidence matrix for the random effect vector **u**. The population model (4.4.1) can then be written

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

The error vector **e** and area effect vector **u** have independent multivariate normal distributions with zero mean vectors and covariance matrices given $\sigma^2\mathbf{I}_N$ and $\sigma_u^2\boldsymbol{\Omega}_1$ respectively, with the matrix $\boldsymbol{\Omega}_1$ reflecting the spatial autocorrelation of the area effects. For example, this is achieved via a model of the form

$$\boldsymbol{\Omega}_1 = \left[[1 + \delta_{dd'} \exp(\frac{Dist(d,\ d')}{\rho})]^{-1} \right]$$

where ρ is an unknown parameter and $\delta_{dd'}$ is zero for $d = d'$ and 1 otherwise. The model for the sample vector $\mathbf{y}_s$ is

(4.4.4) $$\mathbf{y}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s$$

where the incidence matrix $\mathbf{Z}_s = \begin{bmatrix} 1_{n_1} & . & . & 0 \\ 0 & . & . & 0 \\ 0 & . & . & 0 \\ 0 & . & . & 1_{n_D} \end{bmatrix}$ where 1_{n_d} is a vector of dimension n_d with all elements equal to one. It follows that the covariance matrix of $\mathbf{y}_s$ is then $\sigma^2 \mathbf{\Sigma}_s$ where $\mathbf{\Sigma}_s = \mathbf{I}_n + \mathbf{Z}_s \mathbf{\Omega} \mathbf{Z}'_s$, $\mathbf{\Omega} = \varphi \mathbf{\Omega}_1$ and $\varphi = \sigma_u^2 / \sigma^2$.

Given this set-up we are back with the model (1.1) and so can apply the ML and REML estimation theory set out in section 3 to prediction of $\boldsymbol{\theta}$. In this context we observe that both the ML estimation algorithm described in section 3.1 and the REML estimation algorithm described in section 3.2 still apply, with the updating step 8 in the ML algorithm (step 9 in the REML algorithm) replaced by:

> Put $\rho^{new} = \rho^{old} + \theta(\partial l / \partial \rho^{old})$, where
>
> $$\partial l / \partial \rho = -(1/2)[(\mathrm{tr}(\mathbf{\Omega}_1^{-1}(\partial \mathbf{\Omega}_1 / \partial \rho))) - (\varphi^{-1}\mathrm{tr}(\mathbf{\Omega}_1^{-1}\mathbf{T}_s^{*}\mathbf{\Omega}_1^{-1}(\partial \mathbf{\Omega}_1 / \partial \rho))) \\ - \sigma^{-2}\varphi^{-1}\hat{\mathbf{u}}'\mathbf{\Omega}_1^{-1}(\partial \mathbf{\Omega}_1 / \partial \rho)\mathbf{\Omega}_1^{-1}\hat{\mathbf{u}}]$$
>
> and θ is the (3,3) element of the inverse of the information matrix of the ML/REML estimate of the variance components $\hat{\boldsymbol{\omega}} = (\hat{\sigma}^2, \hat{\varphi}, \hat{\rho})'$.

Note that for the spatial correlation model defined above, we have

$$\partial \mathbf{\Omega}_1 / \partial \rho = \left[\frac{Dist(d,d')}{\rho^2} \delta_{dd'} \exp\left(\frac{Dist(d,d')}{\rho} \right) \left(1 + \delta_{dd'} \exp\left(\frac{Dist(d,d')}{\rho} \right) \right)^{-2} \right].$$

For estimation of the MCPE matrix in this case, we observe that the estimated Fisher information matrix of the ML estimates of the variance components $\hat{\boldsymbol{\omega}} = (\hat{\sigma}^2, \hat{\varphi}, \hat{\rho})'$ is

$$\hat{\mathbf{I}}_{ML} = \frac{1}{2} \begin{bmatrix} n\hat{\sigma}^{-4} & \hat{\sigma}^{-2}\hat{\varphi}^{-1}(\nu - \hat{r}_1^{*}) & \hat{\sigma}^{-2}(\hat{r}_1^{*(1)} - v_1^{(1)}) \\ & \hat{\varphi}^{-2}(\nu - 2\hat{r}_1^{*}) + \hat{\varphi}^{-4}\hat{r}_{11}^{*} & \hat{\varphi}^{-1}(2\hat{r}_1^{*(1)} - v_1^{(1)} - \hat{\varphi}^{-1}\hat{r}_{11}^{*(1)}) \\ & & (\hat{\varphi}^{-2}\hat{r}_{11}^{*(11)} + v_1^{(11)} - 2\hat{r}_1^{*(11)}) \end{bmatrix}$$

while that of the REML estimates is

$$\hat{\mathbf{I}}_{REML} = \frac{1}{2} \begin{bmatrix} (n-p)\hat{\sigma}^{-4} & \hat{\sigma}^{-2}\hat{\varphi}^{-1}(\nu - \hat{r}_1) & \hat{\sigma}^{-2}(\hat{r}_1^{(1)} - v_1^{(1)}) \\ & \hat{\varphi}^{-2}(\nu - 2\hat{r}_1) + \hat{\varphi}^{-4}\hat{r}_{11} & \hat{\varphi}^{-1}(2\hat{r}_1^{(1)} - v_1^{(1)} - \hat{\varphi}^{-1}\hat{r}_{11}^{(1)}) \\ & & (\hat{\varphi}^{-2}\hat{r}_{11}^{(11)} + v_1^{(11)} - 2\hat{r}_1^{(11)}) \end{bmatrix}.$$

In both cases see (3.3.2) for the definitions of the various components, remembering that in the REML case we need to discard the star (*) superscripts. Note also that evaluation of these components depends on $\partial \hat{\mathbf{\Omega}}_1^{-1} / \partial \hat{\rho}$ and is therefore dependent on the actual spatial correlation structure assumed for the area level random effect. Finally we observe that, given the estimate of the matrix **B** defined by the 2 × 2 submatrix in the bottom right hand corner of the inverse of the estimated information matrix for either the ML or REML estimates, we can then estimate the MCPE matrix of the EBLUP $\hat{\boldsymbol{\theta}}$ following the same steps as outlined in section 4.3 above.

5 Application to Area Level Linear Mixed Models

There are many applications of small area estimation where the survey information is available at area level rather that at individual level. For example, survey-based estimates of small area averages (the so-called direct estimates) are available, but individual level data are not. The theory developed in sections 1 – 3 can be adapted to this situation provided it is reasonable to assume these direct estimates can also be modelled linearly. We now illustrate this via application of area level versions of the models considered in section 4.

5.1 A Model with IID Area and IID Time Effects

Let the vector $\mathbf{y} = \{y_{dt};\ t = 1,2,\ldots, T;\ d = 1,2,\ldots,D\}$ consist of the direct survey estimates of the survey variable Y. The subscripts d and t ($t = 1,2,\ldots,T$; $d = 1,2,\ldots,D$) represent area and time respectively. Let $\mathbf{X}$ be the matrix of the auxiliary covariates $\mathbf{X}_{dt}$, all measured at area level. Assuming that u_{1t} and u_{2d} represent time and area effects respectively, let η_{dt} denote the mean of Y in small area d at time t (i.e. after conditioning on these effects). The objective then is to predict the value of $\boldsymbol{\theta} = \mathbf{a}\boldsymbol{\eta}$, where $\boldsymbol{\eta} = (\eta_{dt})$, given the following model

(5.1.1) $$\eta_{dt} = \beta_0 + \mathbf{X}'_{dt}\boldsymbol{\beta}_t + u_{1t} + u_{2d}\,.$$

The vector $\boldsymbol{\beta}_t$ contains regression coefficients at time $t = 1,2, .., T$, β_0 is the intercept and the random effects u_{1t} and u_{2d} are assumed to be mutually independent and normally distributed with zero means and variances as defined below. To illustrate, let $T = 2$ and $D = 3$. The **a** matrix that defines the small area means at last time period ($t = 2$) is then

$$\mathbf{a} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The linear predictor η_{dt} is related to direct estimator y_{dt} via following model

(5.1.2) $$y_{dt} = \eta_{dt} + e_{dt}$$

where e_{dt} represents sampling error. This is often referred to as a Fay-Herriot-type model. Let $\mathbf{u}_1 = [u_{11}, u_{12},...u_{1T}]'$, $\mathbf{u}_2 = [u_{21}, u_{22},...u_{1D}]'$ and $\mathbf{e} = [e_{dt}]$ and let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ be the incidence matrices for the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ respectively. Put $\mathbf{Z}_1' = \mathbf{1}_D \otimes \mathbf{I}_T = (\mathbf{Z}_1'^{*}, \mathbf{Z}_2'^{*},...,\mathbf{Z}_D'^{*})$ with $\mathbf{Z}_d^{*} = \mathbf{I}_T$ for $d = 1,\cdots,D)$ and $\mathbf{Z}_2 = \mathbf{I}_D \otimes \mathbf{1}_T = diag(\mathbf{1}_T;\, d = 1,\cdots,D)$ where $\mathbf{1}_T$ is a vector of dimension T with all elements equal to one and $\mathbf{I}_D$ and $\mathbf{I}_T$ are identity matrices of order D and T respectively. The model (5.1.2) can then be written in matrix form as

(5.1.3) $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}.$$

The random vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{e}$ are assumed to be mutually independent with covariance matrices given by $\sigma_1^2\mathbf{I}_T$, $\sigma_2^2\mathbf{I}_D$ and $\sigma^2\mathbf{W}$ respectively. The matrix $\mathbf{W}$ is a known positive definte square matrix of order $n = T \times D$. Typically $\mathbf{W}$ is a diagonal matrix with elements that are functions of the sample sizes n_{dt}. Here n_{dt} denotes the sample size in small area d at time t. The covariance matrix of $\mathbf{y}$ is then $Var(\mathbf{y}) = \sigma^2\mathbf{W} + \sigma_1^2\mathbf{Z}_1\mathbf{Z}_1' + \sigma_2^2\mathbf{Z}_2\mathbf{Z}_2'$. Setting $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, the model (5.1.3) further simplifies to

(5.1.4) $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

in which case $\mathrm{Var}(\mathbf{y}) = \sigma^2\boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma} = \mathbf{W} + \varphi_1\mathbf{Z}_1\mathbf{Z}_1' + \varphi_2\mathbf{Z}_2\mathbf{Z}_2'$ and $\varphi_j = \sigma_j^2/\sigma^2$.

ML and REML estimation of the parameters of (5.1.4), as well as calculation of the corresponding EBLUP and its estimated MCPE matrix, then follows along exactly the same lines as in section 4.1. The only difference is that we replace $\mathbf{y}_s$ and $\boldsymbol{\Sigma}_s$ there by $\mathbf{y}$ and $\boldsymbol{\Sigma}$ here. Note that this model can only be fitted if $D > 1$ and $T > 1$. If either of these conditions are not met, then σ^2 is not identifiable. In such a case we need to estimate this parameter using other methods and then substitute this estimate in the preceding development.

5.2 A Model with IID Area and Autocorrelated Time Effects

Here we extend the model in the previous section to allow the time effect to be the outcome of a stochastic process. In particular, we assume that this process is AR(1). The model for the observed direct estimates is therefore

(5.2.1) $$y_{dt} = \eta_{dt} + e_{dt}$$

where $\eta_{dt} = \beta_0 + \mathbf{X}_{dt}'\boldsymbol{\beta}_t + u_{1t} + u_{2d}$ and e_{dt} is sampling error. The vector $\boldsymbol{\beta}_t$ contains the regression coefficients for time $t = 1, 2, .., T$, β_0 is the intercept and the random effects u_{1t} and u_{2d} are assumed to be mutually independent and to follow normal distributions with zero means and variances as set out below. Let $\mathbf{u}_1 = [u_{11}, u_{12}, \ldots u_{1T}]'$, $\mathbf{u}_2 = [u_{21}, u_{22}, \ldots u_{1D}]'$ and $\mathbf{e} = [e_{dt}]$ and let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ to be the incidence matrices for the random effect vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ respectively. The model (5.2.1) can then be written in matrix form as

(5.2.2) $\mathbf{y} = \eta + \mathbf{e}$

where $\eta = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ were defined in the previous section. The random vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{e}$ are assumed to have covariance matrices given by $\sigma_1^2\boldsymbol{\Omega}_1$, $\sigma_2^2\mathbf{I}_D$ and $\sigma^2\mathbf{W}$ respectively, where $\boldsymbol{\Omega}_1$ is the AR(1) correlation matrix defined in section 4.2 and $\mathbf{W}$ has the same interpretation as in the previous section. It immediately follows that we can write

(5.2.3) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$

where the covariance matrix of $\mathbf{y}$ is given by $\sigma^2\boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma} = \mathbf{W} + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}'$, with $\boldsymbol{\Omega} = diag(\varphi_1\boldsymbol{\Omega}_1, \varphi_2\boldsymbol{\Omega}_2) = diag(\varphi_1\boldsymbol{\Omega}_1, \varphi_2\mathbf{I}_D)$ and $\varphi_j = \sigma_j^2/\sigma^2$.

Again, we observe that this is identical to the set up in section 4.2 once we replace $\mathbf{y}_s$ and $\boldsymbol{\Sigma}_s$ there by $\mathbf{y}$ and $\boldsymbol{\Sigma}$ here. Parameter estimation and prediction of $\boldsymbol{\theta}$ follow directly.

5.3 A Model with a Time Varying Area Effect

The motivation for this type of model has already been provided in section 4.3. Here we note that at the area level such a model can be expressed as

(5.3.1) $y_{dt} = \eta_{dt} + e_{dt}$

where $\eta_{dt} = \beta_0 + \mathbf{X}'_{dt}\boldsymbol{\beta}_t + u_{dt}$ and u_{dt} is a random effect that follows independent AR(1) processes for $d = 1, 2, \ldots, D$. As usual, the random effect u_{dt} and the sampling error e_{dt} are assumed to be mutually independent and to follow normal distributions with zero means and appropriate variances. Let $\mathbf{u} = [u_{11}, u_{12}, \ldots u_{dt}, \ldots, u_{DT}]'$, and $\mathbf{e} = [e_{dt}]$ and let $\mathbf{Z}$ be the incidence matrix for the random effect vector $\mathbf{u}$. The model (5.3.1) can then be written in matrix form as

(5.3.2) $\mathbf{y} = \boldsymbol{\eta} + \mathbf{e}$

where $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$. The random vectors $\mathbf{u}$ and $\mathbf{e}$ mutually independent realisations from multivariate normal distributions with zero mean vectors and covariance matrices given by $\sigma_u^2\boldsymbol{\Omega}_1$ and $\sigma^2\mathbf{W}$ respectively. Here $\boldsymbol{\Omega}_1$ has the same specification as in section 4.3 and $\mathbf{W}$ is the same as in section 5.1. The covariance matrix of **y** is then $\sigma^2(\mathbf{W}+\mathbf{Z\Omega Z}')$ = $\sigma^2\mathbf{\Sigma}$ where $\mathbf{\Omega}=\varphi\mathbf{\Omega}_1$ and $\varphi=\sigma_u^2/\sigma^2$.

Once more we observe that this is identical to the set up in section 4.3 when we replace $\mathbf{y}_s$ and $\mathbf{\Sigma}_s$ there by **y** and $\mathbf{\Sigma}$ here. Parameter estimation and prediction of $\boldsymbol{\theta}$ follow directly provided σ^2 is known. This is because this parameter is not identifiable given the specification above.

6 Generalized Linear Mixed Models

Many variables of interest in small area estimation are not normally distributed, and therefore cannot be adequately modelled via the linear mixed model theory set out in sections 1 to 3. For such variables we can instead consider using a generalized linear mixed model (GLMM). Under this type of model, the distribution of the vector **y** of population values of the variable of interest is assumed to depend on a vector quantity $\boldsymbol{\eta}$ that is related to regression covariates and random components through the equation $\boldsymbol{\eta}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Zu}$. The linear predictor $\boldsymbol{\eta}$ is connected to **y** via a known function *h*, defined by $\mathrm{E}(\mathbf{y}\,|\,\mathbf{u})=h(\boldsymbol{\eta})$. As usual, we partition the linear predictor $\boldsymbol{\eta}$ and the conditional mean $\mathrm{E}(\mathbf{y}\,|\,\mathbf{u})=h(\boldsymbol{\eta})$ as $\boldsymbol{\eta}=[\boldsymbol{\eta}_s',\boldsymbol{\eta}_r']'$ and $\mathrm{E}(\mathbf{y}\,|\,\mathbf{u})=[h'(\boldsymbol{\eta}_s),\,h'(\boldsymbol{\eta}_r)]'$ where $\boldsymbol{\eta}_s=\mathbf{X}_s\boldsymbol{\beta}+\mathbf{Z}_s\mathbf{u}$, $\boldsymbol{\eta}_r=\mathbf{X}_r\boldsymbol{\beta}+\mathbf{Z}_r\mathbf{u}$, $\mathrm{E}(\mathbf{y}_s|\mathbf{u})=h(\boldsymbol{\eta}_s)$ and $\mathrm{E}(\mathbf{y}_r|\mathbf{u})=h(\boldsymbol{\eta}_r)$. The parameter of interest is $\boldsymbol{\theta}=\mathbf{ay}$ which can be written as

(6.1) $$\boldsymbol{\theta}=\mathbf{a}_s\mathbf{y}_s+\mathbf{a}_r\mathbf{y}_r.$$

As in the preceding linear model development, the first term in (6.1) depends only on the sample values and is known after the sample is observed. The second term, which depends on the non-sample values, is unknown. The predicted value of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, is then obtained by replacing $\mathbf{y}_r$ with its predicted value in (6.1). This predictor of $\mathbf{y}_r$ is $h(\hat{\boldsymbol{\eta}}_r)$ where $\hat{\boldsymbol{\eta}}_r$ = $\mathbf{X}_r\hat{\boldsymbol{\beta}}+\mathbf{Z}_r\hat{\mathbf{u}}$, and where the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ are obtained by fitting the model

(6.2) $$\boldsymbol{\eta}_s=\mathbf{X}_s\boldsymbol{\beta}+\mathbf{Z}_s\mathbf{u}$$

using the sample data. This leads to the predicted value

(6.3) $$\hat{\boldsymbol{\theta}}=\mathbf{a}_s\mathbf{y}_s+\mathbf{a}_r\hat{\mathbf{y}}_r=\mathbf{a}_s\mathbf{y}_s+\mathbf{a}_r h(\hat{\boldsymbol{\eta}}_r)=\mathbf{a}_s\mathbf{y}_s+\mathbf{a}_r h(\mathbf{X}_r\hat{\boldsymbol{\beta}}+\mathbf{Z}_r\hat{\mathbf{u}}).$$

6.1 Generalized Best Linear Unbiased Prediction via Penalised Quasi-Likelihood

Let $\mathbf{y}_s = \{y_{dti}\}$ denotes vector of sample values of the survey variable Y. The subscripts d, t and i ($d = 1, 2, \ldots, D$; $t = 1, 2, \ldots, T$; $i = 1, 2, \ldots, n_{dt}$) represent area, time and unit respectively. As in the linear case, we assume that the random effect $\mathbf{u}_j$ is multivariate normal with zero mean vector and covariance matrix $\varphi_j\mathbf{\Omega}_j(\boldsymbol{\rho})$ and so $\mathbf{u} = [\mathbf{u}_1', \mathbf{u}_2', \ldots \mathbf{u}_J']'$ has covariance $\mathbf{\Omega}$ = $diag(\varphi_j\mathbf{\Omega}_j(\boldsymbol{\rho}))$. Let $f_1(\mathbf{y}_s|\mathbf{u})$ be the probability density function of $\mathbf{y}_s$ conditional on $\mathbf{u}$ and let $f_2(\mathbf{u})$ be the probability density function of $\mathbf{u}$. The loglikelihood function defined by the vector $\mathbf{y}_s$ conditional on fixed $\mathbf{u}$ and the logarithm of the probability density function of $\mathbf{u}$ are then

(6.1.1) $$l_1 = \ln(f_1(\mathbf{y}_s|\mathbf{u}))$$

(6.1.2) $$l_2 = -(1/2)[\text{Const.} + \ln|\mathbf{\Omega}| + \mathbf{u}'\mathbf{\Omega}^{-1}\mathbf{u}].$$

The $\boldsymbol{\beta}$ and $\mathbf{u}$ values that jointly maximise $l = l_1 + l_2$ are called the maximum penalised quasi-likelihood (MPQL) estimates. Substituting these penalised quasi-likelihood estimates into (6.3) yields the MPQL estimate of $\boldsymbol{\theta}$.

Given $\mathbf{\Omega}$, an iterative procedure can used to obtain the MPQL estimates of $\boldsymbol{\beta}$ and $\mathbf{u}$. This is:

1. Assign initial values to $\boldsymbol{\beta}$ and $\mathbf{u}$.
2. Update these values via $\begin{bmatrix}\boldsymbol{\beta}_{new}\\ \mathbf{u}_{new}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\beta}_{old}\\ \mathbf{u}_{old}\end{bmatrix} + \mathbf{V}_s^{-1}\begin{bmatrix}\mathbf{X}_s'\\ \mathbf{Z}_s'\end{bmatrix}(\partial l_1/\partial\boldsymbol{\eta}_s|_{\boldsymbol{\beta}_{old},\mathbf{u}_{old}}) - \mathbf{V}_s^{-1}\begin{bmatrix}\mathbf{0}\\ \mathbf{\Omega}^{-1}\mathbf{u}_{old}\end{bmatrix}$, where $\mathbf{V}_s = \begin{bmatrix}\mathbf{X}_s'\\ \mathbf{Z}_s'\end{bmatrix}(-\partial^2 l_1/\partial\boldsymbol{\eta}_s\partial\boldsymbol{\eta}_s')[\mathbf{X}_s\ \mathbf{Z}_s] + \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Omega}^{-1}\end{bmatrix}$ and $\partial l_1/\partial\boldsymbol{\eta}_s$, $\partial^2 l_1/\partial\boldsymbol{\eta}_s\partial\boldsymbol{\eta}_s'$ are the first and second order derivatives of l_1 with respect to $\boldsymbol{\eta}_s$.
3. Return to step 2.

At convergence the MPQL estimate of $\boldsymbol{\theta}$ is obtained by substituting the converged values of $\boldsymbol{\beta}$ and $\mathbf{u}$ (denoted by a "tilde" below) in the right hand side of (6.3), leading to

(6.1.2) $$\tilde{\boldsymbol{\theta}} = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\tilde{\mathbf{y}}_r = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r h(\tilde{\boldsymbol{\eta}}_r) = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r h(\mathbf{X}_r\tilde{\boldsymbol{\beta}} + \mathbf{Z}_r\tilde{\mathbf{u}}).$$

Define $\mathbf{B}_s = -(\partial^2 l_1/\partial\boldsymbol{\eta}_s\partial\boldsymbol{\eta}_s'|_{\boldsymbol{\eta}_s=\tilde{\boldsymbol{\eta}}_s})$ as the matrix of second order derivatives of l_1 with respect to $\boldsymbol{\eta}_s$ and evaluated at $\tilde{\boldsymbol{\eta}}_s = \mathbf{X}_s\tilde{\boldsymbol{\beta}} + \mathbf{Z}_s\tilde{\mathbf{u}}$ and put $\mathbf{T}_s^* = (\mathbf{\Omega}^{-1} + \mathbf{Z}_s'\mathbf{B}_s\mathbf{Z}_s)^{-1}$. Define $\mathbf{B}_r$ in the same way. Let $\mathbf{V} = \begin{bmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{bmatrix}$ and $\mathbf{V}^{-1} = \begin{bmatrix}\mathbf{T}_{11} & \mathbf{T}_{12}\\ \mathbf{T}_{21} & \mathbf{T}_{22}\end{bmatrix}$ denote the partitioning of the matrix $\mathbf{V}$ and its inverse defined by the dimensions of $\boldsymbol{\beta}$ and $\mathbf{u}$. Let $\mathbf{H}_r = \mathbf{H}(\tilde{\boldsymbol{\eta}}_r) = \partial h(\boldsymbol{\eta}_r)/\partial\boldsymbol{\eta}_r|_{\boldsymbol{\eta}_r=\tilde{\boldsymbol{\eta}}_r}$,

$\mathbf{X}_r^+ = \mathbf{a}_r\mathbf{H}_r\mathbf{X}_r$ and $\mathbf{Z}_r^+ = \mathbf{a}_r\mathbf{H}_r\mathbf{Z}_r$. A first order approximation to the MCPE matrix of the PL estimator of $\boldsymbol{\theta}$ is then $MCPE(\tilde{\boldsymbol{\theta}}) = \mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_4$, where $\mathbf{G}_1 = \mathbf{Z}_r^+\mathbf{T}_s^*\mathbf{Z}_r^{+\prime}$, $\mathbf{G}_4 = \mathbf{a}_r\mathbf{B}_r\mathbf{a}_r'$ and $\mathbf{G}_2 = [\mathbf{X}_r^+ - \mathbf{Z}_r^+\mathbf{T}_s^*\mathbf{Z}_s'\mathbf{B}_s\mathbf{X}_s]\mathbf{T}_{11}[\mathbf{X}_r^{+\prime} - \mathbf{X}_s'\mathbf{B}_s'\mathbf{Z}_s\mathbf{T}_s^*\mathbf{Z}_r^{+\prime}]$.

In practice the variance components parameters defining the matrix $\boldsymbol{\Omega}$ are unknown and have to be estimated from the sample data. The MPQL estimate of these variance components are biased and so this approach is not recommended in the practice. Alternative estimates based on maximum likelihood (ML) and restricted maximum likelihood (REML) can be defined. In particular, the bias in the REML estimates is typically small.

6.2 ML and REML Estimation of Variance Components

For the normal error model, the interrelationship between BLUP (MPQL) with maximum likelihood (ML) and restricted or residual maximum likelihood (REML) estimators was developed in Harville (1977) and investigated further in Thompson (1981), Fellner (1986, 1987) and Speed (1991). McGilchrist (1994) extends this approach to generalized linear mixed models. This method has elements in common with Schall (1991), Breslow and Clayton (1993), Wolfinger (1993), Nelder and Lee (1996) and Saei and McGilchrist (1998). Lee and Nelder (2001a, 2001b) further extend the work of Nelder and Lee (1996) to correlated non-normal data. Below we outline the extension of McGilchrist (1994) to estimation of variance components for the small area estimation problem.

Let l_1 be the loglikelihood function of the sample values $\mathbf{y}_s$ conditional on $\mathbf{u}$ and let l_2 be the logarithm of the probability density function of $\mathbf{u}$. Expressions for l_1 and l_2 are given in (6.1.1) and (6.1.2) in the previous section. An iterative procedure that combines the MPQL estimation of $\boldsymbol{\beta}$ and $\mathbf{u}$ with ML estimation of $\boldsymbol{\Omega}$ is then given by:

1. Assign initial values to $\boldsymbol{\beta}$, $\mathbf{u}$, $\boldsymbol{\varphi} = (\varphi_j)$ and $\boldsymbol{\rho} = (\rho_h)$.
2. Update $\boldsymbol{\Omega}$.
3. Update $\boldsymbol{\beta}$ and $\mathbf{u}$ using the iterative PL estimation procedure described in the previous section.
4. Update $\boldsymbol{\eta}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u}$.
5. Update $\mathbf{B}_s = -(\partial^2 l_1 / \partial\boldsymbol{\eta}_s \partial\boldsymbol{\eta}_s')$.
6. Update $\mathbf{T}_s^* = (\boldsymbol{\Omega}^{-1} + \mathbf{Z}_s'\mathbf{B}_s\mathbf{Z}_s)^{-1} = [\mathbf{T}_{sjj'}^*]$.
7. Update $\varphi_j = v_j^{-1}(\mathrm{tr}(\mathbf{T}_{sjj}^*\boldsymbol{\Omega}_j^{-1}) + \sigma^{-2}\mathbf{u}_j'\boldsymbol{\Omega}_j^{-1}\mathbf{u}_j)$ where v_j is the rank of the matrix $\mathbf{Z}_{sj}$.
8. Update $\rho_h = f_h(\boldsymbol{\rho}, \boldsymbol{\varphi}, \mathbf{T}_s^*, \mathbf{u})$ where f_h is a known function whose specification depends on the parameterization of $\boldsymbol{\Omega}(\boldsymbol{\rho})$, and current values for variance components are used in the right hand side of this equation.
9. Return to step 2 and repeat the procedure until the values of the different parameters converge.

The corresponding iterative procedure used to obtain the REML estimators is exactly the same except that the role of $\mathbf{T}_s^*$ in steps 7 and 8 of the ML algorithm is played by the $\mathbf{T}_{22}$ submatrix of the matrix $\mathbf{T}$ defined in section 6.1 above. We note that this matrix can be explicitly calculated using the formulae

$$\mathbf{T}_{22} = \mathbf{T}_s^* + \mathbf{T}_s^*\mathbf{Z}_s'\mathbf{B}_s\mathbf{X}_s\mathbf{T}_{11}\mathbf{X}_s'\mathbf{B}_s'\mathbf{Z}_s\mathbf{T}_s^*$$

$$\mathbf{T}_{11} = (\mathbf{X}_s'\mathbf{B}_s\mathbf{X}_s - \mathbf{X}_s'\mathbf{B}_s\mathbf{Z}_s\mathbf{T}_s^*\mathbf{Z}_s'\mathbf{B}_s\mathbf{X}_s)^{-1}.$$

This procedure is similar to the hierarchical likelihood approach of Nelder and Lee (1996) and is based on the method described in Saei and McGilchrist (1998).

The final "EBLUP-type" predictor $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by substituting these parameter estimates (MPQL/ML-based or MPQL/REML-based) into the expression (6.3), similarly to (6.1.2). In order to estimate the MCPE matrix $\hat{\boldsymbol{\theta}}$ we note that it can be approximated in the same way as the MCPE matrix of the "BLUP-type" predictor $\tilde{\boldsymbol{\theta}}$ defined by (6.1.2). Since the variance components parameters have been estimated, however, we obtain an extra term in this approximation, leading to an approximate MCPE matrix for the estimator $\hat{\boldsymbol{\theta}}$ of the form

(6.2.1) $$MCPE(\hat{\boldsymbol{\theta}}) \cong MCPE(\tilde{\boldsymbol{\theta}}) + 2\mathbf{G}_3 = \mathbf{G}_1 + \mathbf{G}_2 + 2\mathbf{G}_3 + \mathbf{G}_4.$$

The $\mathbf{G}_1$, $\mathbf{G}_2$ and $\mathbf{G}_4$ terms here are defined in exactly the same way as in the MCPE matrix for $\tilde{\boldsymbol{\theta}}$. In order to obtain $\mathbf{G}_3$, put $\boldsymbol{\Delta} = \mathbf{Z}_r^+\mathbf{T}_s^*$ and let $\mathbf{Z}_\alpha^+$ be the α^{th} row of the matrix $\mathbf{Z}_r^+$, with $\partial\boldsymbol{\Delta}_\alpha/\partial\boldsymbol{\gamma} = \partial(\mathbf{Z}_\alpha^+\mathbf{T}^*)/\partial\boldsymbol{\gamma} = \boldsymbol{\nabla}_\alpha$. Here $\boldsymbol{\gamma} = (\boldsymbol{\varphi}'\ \boldsymbol{\rho}')'$ is the variance components parameter vector. Then $\mathbf{G}_3 = [\mathrm{tr}(\nabla_i\boldsymbol{\Sigma}_s^+\nabla'_j\mathbf{I}^{-1})]$ where $\mathbf{I}$ denotes the inverse of the appropriate Fisher information matrix for $\boldsymbol{\gamma}$ (this will vary depending on whether ML or REML is used in estimation), $\boldsymbol{\Sigma}_s^+ = \mathbf{Z}_s'\mathbf{B}_s\mathbf{Z}_s + \mathbf{Z}_s'\mathbf{B}_s\mathbf{Z}_s\boldsymbol{\Omega}\mathbf{Z}_s'\mathbf{B}_s\mathbf{Z}_s$ and $\boldsymbol{\nabla}_\alpha = -(\mathbf{Z}_\alpha^+\mathbf{T}_s^* \otimes \mathbf{I}_{J+H})(\partial\boldsymbol{\Omega}^{-1}/\partial\boldsymbol{\gamma})\mathbf{T}_s^*$. An estimate of the MCPE matrix of $\hat{\boldsymbol{\theta}}$ is finally obtained by substituting estimates for the unknown parameters in all these expressions.

6.3 Application to Binomial Response Data

We consider data for a single time period and so drop the t subscript. Let

$$\mathbf{y}_s = \{y_{sdi}\} = (y_{s11}, y_{s12}, ..., y_{s1I}, ..., y_{sd1}, y_{sd2}, ..., y_{sdI}, ..., y_{sD1}, y_{sD2}, ..., y_{sDI})'$$

$$\mathbf{y}_r = \{y_{rdi}\} = (y_{r11}, y_{r12}, ..., y_{r1I}, ..., y_{rd1}, y_{rd2}, ..., y_{rdI}, ..., y_{rD1}, y_{rD2}, ..., y_{rDI})'$$

respectively denote the vectors of sample and non-sample values of a binomial survey variable Y. Here the index i indicates a stratification of the population into I mutually exclusive "groups" (e.g. defined by age and sex) and the survey variable Y is a count of the number of people in a group with a characteristic of interest (e.g. being unemployed). In terms of these examples, y_{sdi} corresponds to the sample count of unemployed individuals in age-sex group i in small area d while y_{rdi} is the corresponding non-sample count.

To provide a focus for the following discussion, we shall from now on assume these examples apply, with the parameter of interest then being $\boldsymbol{\theta} = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\mathbf{y}_r$. We also assume that between area variation in these counts can be characterised in terms of the value of a random effect vector $\mathbf{u}$ that is distributed as multivariate normal with zero mean vector and covariance matrix $\varphi\mathbf{I}_D$. Let p_{di} be the probability that an individual in age-sex group i from area d is unemployed and let N_{di} (n_{di}) be the number of total number of individuals in group i from area d in the population (sample) and put $M_{di} = N_{di} - n_{di}$. A popular model for this type of data is then where the sample (non-sample) counts y_{sdi} (y_{rdi}) are distributed as independent binomial variables with mean $n_{di}p_{di}$ ($M_{di}p_{di}$) with the common age-sex by area unemployment probabilities p_{di} following the linear logistic mixed model,

$$\text{logit}(p_{di}) = \ln\left(\frac{p_{di}}{1-p_{di}}\right) = \mathbf{x}'_{di}\boldsymbol{\beta} + u_d = \eta_{di}$$

where $\mathbf{x}_{di}$ is a vector of area by age-sex group level covariates and u_d is the area effect. It follows that the means of $\mathbf{y}_s$ and $\mathbf{y}_r$ conditional on fixed $\mathbf{u}$ are

$$\text{E}(\mathbf{y}_s \mid \mathbf{u}) = h(\boldsymbol{\eta}_s) = \left[n_{di}\frac{\exp(\eta_{di})}{1+\exp(\eta_{di})}\right] = \left[n_{di}\frac{\exp(x'_{di}\boldsymbol{\beta} + u_d)}{1+\exp(x'_{di}\boldsymbol{\beta} + u_d)}\right],$$

$$\text{E}(\mathbf{y}_r \mid \mathbf{u}) = h(\boldsymbol{\eta}_r) = \left[M_{di}\frac{\exp(\eta_{di})}{1+\exp(\eta_{di})}\right] = \left[M_{di}\frac{\exp(x'_{di}\boldsymbol{\beta} + u_d)}{1+\exp(x'_{di}\boldsymbol{\beta} + u_d)}\right]$$

and the loglikelihood function of the binomial vector $\mathbf{y}_s$ conditional on fixed $\mathbf{u}$ and the log of probability density function of $\mathbf{u}$ are

(6.3.2) $$l_1 = \text{Const.} + \sum_{d=1}^{D}\sum_{i=1}^{I}[y_{sdi}\eta_{di} - n_{di}\ln(1+\exp(\eta_{di}))]$$

(6.3.3) $$l_2 = -(1/2)[Const. + D\ln\varphi + \varphi^{-1}\mathbf{u}'\mathbf{u}].$$

For fixed φ, the $\boldsymbol{\beta}$ and $\mathbf{u}$ values that jointly maximise $l = l_1 + l_2$ are called the penalised likelihood estimates of these quantities. See Saei and McGilchrist (1998). Replacement of these penalised likelihood estimates of $\boldsymbol{\beta}$ and $\mathbf{u}$ into (2.3) yields the MPQL (BLUP-type) estimate of $\boldsymbol{\theta}$. The iterative procedure that yields these MPQL estimates is described in section 6.1. To apply it here we note that now $\boldsymbol{\Omega} = \varphi \text{I}_D$, $\mathbf{Z}_s = \mathbf{Z} = diag(\mathbf{1}_I; d = 1, .., D)$, where $\mathbf{1}_I$ is a vector of ones of length I, $\mathbf{X}_s = \mathbf{X} = [\mathbf{x}_{11}\cdots\mathbf{x}_{1I}\cdots\mathbf{x}_{d1}\cdots\mathbf{x}_{di}\cdots\mathbf{x}_{dI}\cdots\mathbf{x}_{DI}]'$,

$$\partial l_1/\partial\boldsymbol{\eta}_s = [y_{sdi} - n_{di}\frac{\exp(\eta_{di})}{1+\exp(\eta_{di})}] = [y_{sdi} - n_{di}p_{di}]$$

and

$$\partial^2 l_1/\partial\boldsymbol{\eta}_s\partial\boldsymbol{\eta}'_s = diag[-n_{di}\frac{\exp(\eta_{di})}{(1+\exp(\eta_{di}))^2}] = diag[-n_{di}p_{di}(1-p_{di})].$$

In order to estimate φ we can use either an ML or a REML approach. Algorithms for implementing both jointly with PL estimation of $\boldsymbol{\beta}$ and $\mathbf{u}$ are described in the previous section. In this case we replace steps 5 – 8 of the ML algorithm by:

5. Update $\mathbf{B}_s = diag[n_{di}p_{di}(1-p_{di})]$.
6. Update $\mathbf{T}_s^* = (\varphi^{-1}\mathbf{I}_D + \mathbf{Z}'\mathbf{B}_s\mathbf{Z})^{-1}$.
7. Update $\varphi = D^{-1}(\text{tr}(\mathbf{T}_s^*) + \mathbf{u}'\mathbf{u})$.

The corresponding REML estimate of φ is obtained by using this $\mathbf{T}_s^*$ in the calculation of the $\mathbf{T}_{22}$ and $\mathbf{T}_{11}$ submatrices defined in the REML modifications to the ML algorithm. The final predictor of $\boldsymbol{\theta}$ is then $\hat{\boldsymbol{\theta}} = \mathbf{a}_s\mathbf{y}_s + \mathbf{a}_r\hat{\mathbf{y}}_r$, where

$$\hat{\mathbf{y}}_r = \left[M_{di}\frac{\exp(\hat{\eta}_{di})}{1+\exp(\hat{\eta}_{di})}\right] = \left[M_{di}\frac{\exp(x'_{di}\hat{\boldsymbol{\beta}} + \hat{u}_d)}{1+\exp(x'_{di}\hat{\boldsymbol{\beta}} + \hat{u}_d)}\right].$$

Finally, we note that the MCPE matrix of this predictor can be estimated using (6.2.1). When ML estimates are used in this predictor we put $\mathbf{H}_r = diag[M_{di}\hat{p}_{di}(1-\hat{p}_{di})]$, $\mathbf{X}_r^+ = \mathbf{a}_r\mathbf{H}_r\mathbf{X}$, $r_1 = \hat{\varphi}^{-1}tr(\hat{\mathbf{T}}_s^*)$, $r_{11} = \mathrm{tr}(\hat{\mathbf{T}}_s^*\hat{\mathbf{T}}_s^*)$, $\hat{\mathrm{var}}(\hat{\varphi}) = 2(\hat{\varphi}^{-2}(D-2r_1) + \hat{\varphi}^{-4}r_{11})^{-1}$, $\hat{\nabla}_\alpha = \hat{\varphi}^{-2}\mathbf{Z}_\alpha^+\hat{\mathbf{T}}_s^*\hat{\mathbf{T}}_s^*$, where $\mathbf{Z}_\alpha^+$ is the α^{th} row of the matrix $\mathbf{Z}_r^+ = \mathbf{a}_r\mathbf{H}_r\mathbf{Z}$ and $\hat{\boldsymbol{\Sigma}}_s^+ = \mathbf{Z}'\hat{\mathbf{B}}_s\mathbf{Z} + \hat{\varphi}\mathbf{Z}'\hat{\mathbf{B}}_s\mathbf{Z}\mathbf{Z}'\hat{\mathbf{B}}_s\mathbf{Z}$. The estimated MCPE matrix of $\hat{\boldsymbol{\theta}}$ is then $M\hat{C}\hat{P}E(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{G}}_1 + \hat{\mathbf{G}}_2 + 2\hat{\mathbf{G}}_3 + \hat{\mathbf{G}}_4$, where

$$\hat{\mathbf{G}}_1 = \mathbf{Z}_r^+\hat{\mathbf{T}}_s^*\mathbf{Z}_r^{+\prime}$$

$$\hat{\mathbf{G}}_2 = [\mathbf{X}_r^+ - \mathbf{Z}_r^+\hat{\mathbf{T}}_s^*\mathbf{Z}'\hat{\mathbf{B}}_s\mathbf{X}]\hat{\mathbf{T}}_{11}[\mathbf{X}_r^{+\prime} - \mathbf{X}'\hat{\mathbf{B}}_s'\mathbf{Z}\hat{\mathbf{T}}_s^*\mathbf{Z}_r^{+\prime}]$$

$$\hat{\mathbf{G}}_3 = \left[\mathrm{tr}(\hat{\boldsymbol{\nabla}}_\alpha\hat{\boldsymbol{\Sigma}}_s^+\hat{\boldsymbol{\nabla}}'_{\alpha'})\,\hat{\mathrm{var}}(\hat{\varphi})\right]$$

$$\hat{\mathbf{G}}_4 = \mathbf{a}_r\hat{\mathbf{B}}_r\mathbf{a}_r'.$$

When REML estimates are used, the only change to these expressions is that the estimate $\hat{\mathbf{T}}_s^*$ in the definitions of r_1 and r_{11} above is replaced by the corresponding estimate of the submatrix $\mathbf{T}_{22}$.

References

Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalised linear mixed models. *Journal of the American Statistical Association* , 88, 9-25.

Fellner, W.H. (1986). Robust estimation of variance components. *Technometrics* **28**, 51 – 60

Fellner, W.H. (1987). Spare matrices, and the estimation of variance components by likelihood methods. *Communication in Statistics and Simulation.* **16**, 439 - 463

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320-340.

Henderson, C.R. (1963) Selection index and expected genetic advance. *In Statistical Genetics and Plant Breeding* (W.D. Hanson and H.F. Robinson, eds.), 141-163. National

Academy of Sciences and National Research Council Publication No. 982, Washington, D.C.

Henderson, C.R. (1973) Sire evaluation and genetic trends. *In Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush* 10-41. Amer. Soc. Animal Sci.-Amer. Dairy Sci. Assn.-Poultry Sci. Assn. Champaigne, Illinois.

Henderson, C.R. (1973) Maximum likelihood estimation of variance components, Unpublished manuscript.

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

Kacker, R.N. and Harville, D.A. (1981). Unbiased of two-stage estimation and prediction procedure for mixed linear models. *Communications in Statistics- Theory and Methods* **A10**, 1249-1261.

Lee, Y. and Nelder, J.A. (2001a) Hierarchical generalized linear models: A synthesis of generalised linear models, random-effect model and structured dispersions. *Biometrika*, **88**, 987-1006.

Lee, Y. and Nelder, J.A. (2001b) Modelling and analysing correlated non-normal data. *Statistical Modelling*, **1**, 3-16

McCullagh, P. and Nelder, J.A. (1989) *Generalized linear models.* Second Ed. London: Chapman and Hall.

McGilchrist, C.A. (1994) Estimation in generalized mixed models. *Journal of the Royal Statistical Society Series B*, **56**, 61-69.

Nelder, J.A. and Lee, Y. (1996) Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B*, **58**, 619-678.

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545-554.

Saei, A. and McGilchrist, C. (1998) Longitudinal threshold models with random components. *The Statistician* (JRSS Series D), **47**, 365-375.

Schall, R (1991) Estimation in generalised linear models with random effects. *Biometrika* **78**, 719-727.

Speed, T. (1991) Comment on Robinson: Estimation of random effect. *Statistical Science* **6**, 42-44.

Thompson , R. (1981). Maximum likelihood estimation of variance components. *Math. Operforch. Statist. Ser. Statist*. **11**, 125 – 131.

Wolfinger, R. (1993) Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791-796.