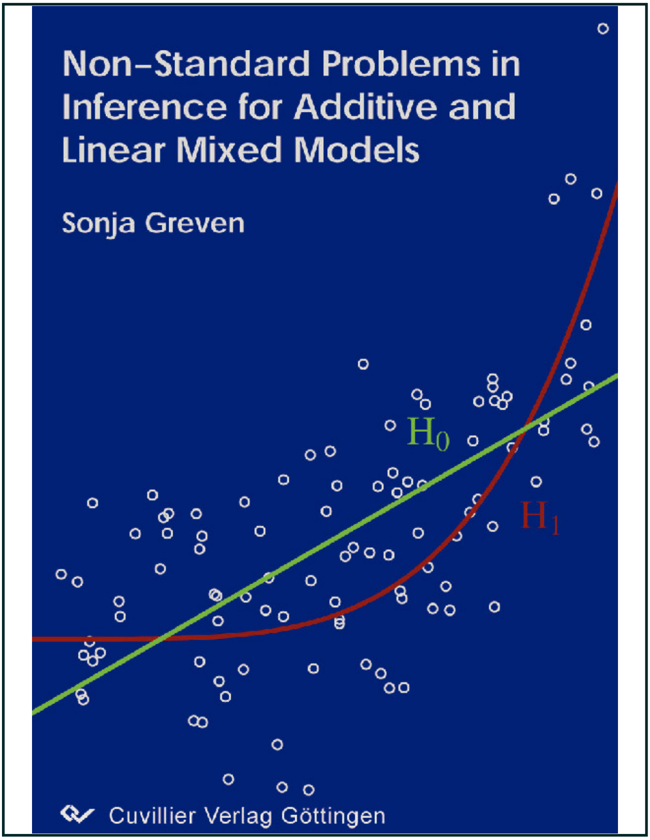
Sonja Greven (Autor)

Non-Standard Problems in Inference for Additive and Linear Mixed Models



https://cuvillier.de/de/shop/publications/1570

Chapter 1

Introduction

Linear mixed models are a powerful inferential tool in modern statistics. They are widely used to model data with different sources of variability, including temporal, spatial and spatio-temporal data. Recent advances utilize the connection between penalized spline smoothing and mixed models for efficient implementation of nonparametric and semiparametric regression techniques.

Nonparametric regression is aimed at reaching more adequate and realistic regression models. Simple linear regression assumes that the relationship between a response variable y and covariates $x_1, \ldots, x_p$ can be described by

$$\mathrm{E}[y] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

with some unknown parameters β_j. However, the linearity and additivity assumption is often too restrictive in realistic settings. Nonparametric regression replaces the linear predictor by an unspecified function $m(\cdot)$, such that the assumption is relaxed to

$$\mathrm{E}[y] = m(x_1, \ldots, x_p).$$

Mixed model penalized splines are a flexible and efficient tool for estimation of the unknown underlying function $m(\cdot)$. They have the added advantage of allowing for straightforward inclusion of additional random effects into the model, such as for longitudinal data. Mixed model penalized splines are widely used and have been extended to many types of models previously limited to linear predictors, such as generalized or survival models.

While linear mixed models have found versatile use in practice, inference for these models is not equally well developed. In particular, inference for random effects is so far limited to certain subclasses of models, or is based on computationally expensive bootstrap procedures. This lag in methodological development is due to the non-standard nature of the testing problem. First, when testing for zero random effects variances, the tested parameter is on the boundary of the parameter space under the null hypothesis. Second, in linear mixed models observations are generally not independent. While in longitudinal linear mixed models there are at least independent subjects or units, such a subdivision of the data is not possible for mixed model penalized spline smoothing.

This dissertation is aimed at developing valid and computationally feasible methodology for inference on random effects in linear mixed models, and at improving our understanding of the effects of boundary setting and lacking independence. We are particularly interested in the important special case of testing for polynomial regression against a general smooth alternative modeled by mixed model penalized splines. All methods are motivated by and applied to the Airgene study on air pollution health effects, where inference on the shape of air pollution dose-response functions is relevant.

This dissertation consists of two parts. The first part, comprising Chapters 2 and 3, lays the ground work. We give a short review of linear mixed model methodology and its connection to longitudinal data and penalized spline smoothing. We also introduce the longitudinal Airgene study and discuss implementation and application of appropriate additive mixed models.

The second and main part, consisting of Chapters 4 to 7, is concerned with methodology for inference in linear mixed models. In particular, we deal with inference on random effects, and the important special case of testing for polynomial regression using mixed model penalized splines.

In Chapter 4, we investigate the asymptotics of restricted likelihood ratio testing for polynomial regression using mixed model penalized splines. We consider two commonly used penalized spline bases, namely truncated polynomials and B-splines. We find that the two are equivalent for restricted maximum likelihood estimation with corresponding penalty and knots, but are not equivalent for maximum likelihood estimation. For both mixed model penalized splines, we show that the asymptotic results on boundary testing for independent observations do not hold, even when the number of spline knots increases to infinity with the sample size. This is due to the asymptotic non-normality of the score statistic. Fundamentally, this is caused by the dependence of observations induced by mixed model penalized splines. We find that this dependence structure cannot be avoided in penalized spline smoothing, as it is inherently necessary for the attainment of smooth curves. A different approach to this testing problem is thus necessary.

In Chapter 5, we therefore develop finite sample alternatives for testing for zero random effect variances in linear mixed models. The class of models we consider is more general than has previously been covered. In particular, it includes nonparametric smoothing as well as models with moderate numbers of clusters or unbalanced designs. We also allow more than one random effect in the model. We propose two approximations to the finite sample null distribution of the restricted likelihood ratio test statistic. Extensive simulations show that both outperform the chi-square mixture approximation and parametric bootstrap currently used. In Chapter 6, we compare the procedures based on the restricted likelihood ratio test to several other tests. We find that our *fast finite sample approximation* is comparable to the best bootstrap-based competitors with regard to power and adherence to the alpha-level, while reducing computation time from hours to seconds.

Lastly, we discuss alternatives to testing for mixed model penalized splines in Chapter 7. In model selection, information criteria are also often used to decide between polynomial and smooth terms. We investigate the Akaike Information Criterion (AIC) based on the marginal likelihood. We show that the AIC is not asymptotically unbiased for the expected relative Kullback-Leibler distance. In fact, it is biased towards the simpler model. There is a close correspondence between the boundary effects on likelihood ratio testing and on the AIC. Contrary to these tests, however, the AIC cannot be adapted to the boundary setting. An alternative is provided using our results on restricted likelihood ratio testing.

In our conclusion, Chapters 8 and 9, we summarize the main findings gained from applying our methods to the Airgene study, and close with a discussion and outlook.

The main results of this dissertation can be summarized as follows:

1. The asymptotics for testing for polynomial regression using mixed model penalized splines are different from the asymptotics for boundary testing with independent observations. This is due to the non-ignorable and unavoidable dependence structure induced by penalized splines.

2. We provide a finite sample alternative for inference on random effects in general, and on mixed model penalized splines in particular. Our approximation is computationally efficient, exact for models with one random effect, and shows power and adherence to the alpha-level comparable to the best bootstrap-based competitors.

3. Our method is also an alternative to model selection for random effects based on the marginal Akaike Information Criterion. This criterion is not asymptotically unbiased for the expected relative Kullback-Leibler distance, and is in fact biased towards the simpler model.

This dissertation is based on

- Greven S, Crainiceanu C, Küchenhoff H, Peters A (2007). Restricted Likelihood Ratio Testing for Zero Variance Components in Linear Mixed Models. *Journal of Computational and Graphical Statistics*, to appear. (Chapter 5)

- Scheipl F, Greven S, Küchenhoff H (2007). Size and Power of Tests for a Zero Random Effect Variance or Polynomial Regression in Additive and Linear Mixed Models. *Computational Statistics & Data Analysis*, to appear, doi:10.1016/j.csda.2007.10.022. (Chapter 6)

- Greven S, Küchenhoff H and Peters A (2006). Additive mixed models with P-Splines. In J. Hinde, J. Einbeck and J. Newell (Eds), *Proceedings of the 21st International Workshop on Statistical Modelling*, 201-207. Statistical Modelling Society. (Chapter 3)

- Peters A, Schneider A, Greven S, Bellander T, Forastiere F, Ibald-Mulli A, Illig T, Jacquemin B, Katsouyanni K, Koenig W, Lanki T, Pekkanen J, Pershangen G, Piccioto S, Rückerl R, Schaffrath Rosario A, Stefanadis C, Sunyer J (2006). Air Pollution and Inflammatory Response in Myocardial Infarction Survivors: Gene-Environment-Interactions in a High-Risk Group. Study Design of the Airgene Study. *Inhalation Toxicology*, 19 (Suppl.1): 161-175. (Section 3.2)

- Rückerl R*, Greven S*, Ljungman P, Aalto P, Antoniades C, Bellander T, Berglind N, Chrysohoou C, Forastiere F, Jacquemin B, von Klot S, Koenig W, Küchenhoff H, Lanki T, Pekkanen J, Perucci CA, Schneider A, Sunyer J, Peters A (2007). Air Pollution and Inflammation (IL6, CRP, Fibrinogen) in Myocardial Infarction Survivors. *Environmental Health Perspectives*, 115 (7): 1072-1080. *equal contribution. (Section 8.1)

- Peters A, Greven S, Heid I, Baldari F, Breitner S, Bellander T, Chrysohoou C, Illig T, Jacquemin B, Koenig W, Lanki T, Nyberg F, Pekkanen J, Pistelli R, Rückerl R, Stefanadis C, Schneider A, Sunyer J, Wichmann HE (2007). Single Nucleotide Polymorphisms in the Fibrinogen Gene Cluster Modify Fibrinogen Response to Ambient Particulate Matter. Submitted to *Lancet*. (Section 8.2)

and two working papers (Chapters 4 and 7).

Chapter 2

Mixed Models, Longitudinal Data and Penalized Spline Smoothing

This chapter gives a brief introduction to the topics this dissertation is based on. We will introduce the linear mixed model, define notation necessary for later chapters, and shed light on the connections of linear mixed models with longitudinal data and penalized spline smoothing. Section 2.1 introduces the linear mixed model and discusses the special case of longitudinal mixed models. Section 2.2 discusses the nonparametric regression problem and shows how mixed models can be used for penalized spline smoothing.

2.1 Mixed Models and Longitudinal Data

2.1.1 The Linear Mixed Model

The linear mixed model can be defined as the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$ is a vector of n observable random variables, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known matrices containing explanatory covariates, $\boldsymbol{\beta}$ is a vector of unknown fixed parameters, $\boldsymbol{b}$ a vector of random effects, and $\boldsymbol{\varepsilon}$ is a vector of unobservable random errors. The assumptions are independence of $\boldsymbol{b}$ and $\boldsymbol{\varepsilon}$, and

$$\mathrm{E}\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \text{Cov}\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R} \end{pmatrix}. \tag{2.2}$$

Model (2.1) with assumption (2.2) can also been seen as the conditional formulation of the mixed model, stating the assumptions for both $\boldsymbol{b}$ and $\boldsymbol{y}|\boldsymbol{b}$: $\boldsymbol{b} \sim (\mathbf{0}, \boldsymbol{D})$ and $\boldsymbol{y}|\boldsymbol{b} \sim (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}, \boldsymbol{R})$, where $\boldsymbol{z} \sim (\boldsymbol{\mu}, \Sigma)$ denotes $\mathrm{E}(\boldsymbol{z}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{z}) = \Sigma$. For inference in model (2.1), however, often the marginal model formulation

$$\boldsymbol{y} \sim (\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}' + \boldsymbol{R}) \tag{2.3}$$

is used. The two models are not equivalent, but (2.1) implies (2.3). Also, for inference in model (2.1), $\boldsymbol{b}$ and $\boldsymbol{\varepsilon}$ are typically assumed to follow multivariate normal distributions with means and covariance matrices as specified in (2.2).

Extensions to generalized linear mixed models will not be discussed here; see for example McCulloch and Searle (2001); Molenberghs and Verbeke (2005).

In the following, let $\boldsymbol{\theta}$ denote the vector of unknown parameters contained in $\boldsymbol{D}$ and $\boldsymbol{R}$.

2.1.2 Longitudinal Data

Mixed models are commonly used to model longitudinal data, where a response variable y is measured for a number of subjects or units repeatedly over time (see, for example, Verbeke and Molenberghs, 2000, for a good overview, and Diggle et al., 1994, also for a contrast with other methods for longitudinal data analysis).

Models for longitudinal data have to take into account the correlation that is present among data on the same unit, which might also depend on the time interval between the measurements. They should adequately model the different sources of variability in the data, namely between units, within units over time, and additional variability due for example to measurement error. A mixed model for longitudinal data is often formulated on the individual level,

$$\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} \tag{2.4}$$

(Laird and Ware, 1982), where $\boldsymbol{y_i} = (y_{i1}, \ldots, y_{in_i})$ contains the n_i responses of unit i, $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ its covariates, and $\boldsymbol{\varepsilon_i} = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i}) \sim (\boldsymbol{0}, \boldsymbol{R_i})$ or $\boldsymbol{\varepsilon_i} \sim N(\boldsymbol{0}, \boldsymbol{R_i})$. While $\boldsymbol{\beta}$ are common regression parameters, the $\boldsymbol{b_i}$ are individual regression parameters for the ith unit assumed to be distributed as $\boldsymbol{b_i} \sim (\boldsymbol{0}, \tilde{\boldsymbol{D}})$ or $\boldsymbol{b_i} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{D}})$. Random effects $\boldsymbol{b_i}$ are commonly used to model differences between units when these units can be regarded as a random sample from an underlying population, while fixed effects are often, though not always, used otherwise (McCulloch and Searle, 2001).

(2.4) is a special case of model (2.1), where $\boldsymbol{y}$ can be divided into independent subvectors $\boldsymbol{y_i}$, and $\boldsymbol{D}$ and $\boldsymbol{R}$ are block diagonal matrices with diagonal blocks $\tilde{\boldsymbol{D}}$ respectively $\boldsymbol{R_i}$. The form (2.1) can be obtained by stashing the $\boldsymbol{y_i}$, $\boldsymbol{X_i}$, $\boldsymbol{b_i}$ and $\boldsymbol{\varepsilon_i}$ on top of each other to obtain $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{b}$ and ε, and letting $\boldsymbol{Z}$ be the block diagonal matrix with blocks $\boldsymbol{Z_i}$ on the diagonal. Then, $\boldsymbol{V} := \text{Cov}(\boldsymbol{y}) = \boldsymbol{ZDZ}' + \boldsymbol{R}$ is also block diagonal, indicating the independence of different units i.

Together, $\boldsymbol{Z}, \boldsymbol{D}$ and $\boldsymbol{R}$ define the correlation structure among observations on the same unit. Common examples for $\boldsymbol{Z}$ include a random intercept design matrix, where $\boldsymbol{Z}$ contains a separate intercept column for each unit. Then, $\tilde{\boldsymbol{D}} = \sigma_b^2$. If $\boldsymbol{R_i} = \sigma_\varepsilon^2 \boldsymbol{I_{n_i}}$ for each i, $\boldsymbol{V}$ has a *compound symmetry* structure, with values $\sigma_b^2 + \sigma_\varepsilon^2$ on the diagonal, and values σ_b^2 on the off-diagonal entries within each block. This means that all pairs of observations on the same unit have the same correlation $(\sigma_b^2 + \sigma_\varepsilon^2)/\sigma_b^2$. Common extensions include columns for random slopes in $\boldsymbol{Z}$, where $\tilde{\boldsymbol{D}}$ can also include a correlation parameter between the random intercept and slope. Additionally, $\boldsymbol{R_i}$ can be used to model correlations between observations on the same unit that are not constant, but rather decreasing with distance between measurements.

2.1.3 Estimation and Prediction

Inference in model (2.1) is usually carried out using the marginal model (2.3). For a given parameter $\boldsymbol{\theta}$, the fixed effects vector $\boldsymbol{\beta}$ can be estimated as

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}, \tag{2.5}$$

provided the respective inverses exist. This will be assumed in the following. This estimate is the weighted (or generalized) least squares estimate, and it is also the maximum likelihood estimate under the normality assumption. Additionally, $\hat{\boldsymbol{\beta}}$ is the *best linear unbiased estimator*

(BLUE) of $\boldsymbol{\beta}$, and the *best unbiased estimator* under the normality assumption (see Zyskind and Martin, 1969; Harville, 1976, for more general results).

For the random effects $\boldsymbol{b}$, the *best linear unbiased predictor (BLUP)* is

$$\hat{\boldsymbol{b}} = \boldsymbol{D}\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}), \tag{2.6}$$

which is also the *best unbiased predictor* under the normality assumption (Harville, 1976). Here, unbiasedness refers to $\mathrm{E}(\hat{\boldsymbol{b}}) = \mathrm{E}(\boldsymbol{b}) = \boldsymbol{0}$, linearity is linearity in $\boldsymbol{y}$, and the best (linear) unbiased predictor is the one minimizing the mean squared error $\mathrm{E}[(\tilde{\boldsymbol{b}} - \boldsymbol{b})^2]$ among all (linear) unbiased predictors $\tilde{\boldsymbol{b}}$ for $\boldsymbol{b}$. The term predictor is used to denote that the target of $\hat{\boldsymbol{b}}$ is random, to distinguish it from estimators of fixed effects. However, we will also use the term as an umbrella term for both BLUP and BLUE. Harville (1976) also discusses best linear unbiased prediction of linear combinations of $\boldsymbol{\beta}$ and $\boldsymbol{b}$. Robinson (1991) reviews several other possible derivations for the BLUP, including *Henderson's Justification* (Henderson, 1950). Henderson maximizes the joint density of $\boldsymbol{y}$ and $\boldsymbol{b}$ assuming normality, yielding the minimization problem

$$\min_{\boldsymbol{\beta},\boldsymbol{b}} \ (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b}) + \boldsymbol{b}\boldsymbol{D}^{-1}\boldsymbol{b}, \tag{2.7}$$

which can be shown to result in the BLUPs (2.5) and (2.6) using

$$\boldsymbol{V}^{-1} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1} \tag{2.8}$$

(Henderson et al., 1959).

Both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}$ are derived from model (2.1) for known parameters $\boldsymbol{\theta}$. As $\boldsymbol{\theta}$ is generally unknown, it has to be estimated as well. Two main methods are typically used for estimation of $\boldsymbol{\theta}$, namely *maximum likelihood (ML)* or *restricted (residual) maximum likelihood (REML)*.

The ML estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ under the normality assumption is obtained by maximizing the log-likelihood for model (2.3),

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(\boldsymbol{V})) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Substituting the ML estimate $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ as defined in (2.5) for $\boldsymbol{\beta}$, the resulting profile log-likelihood $\ell_P(\boldsymbol{\theta})$ can be maximized to obtain the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. The ML estimate for $(\boldsymbol{\theta}, \boldsymbol{\beta})$ then is $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}))$. In general, there is no closed form solution for $\hat{\boldsymbol{\theta}}$ and the maximization has to be done numerically (Harville, 1977).

ML estimation, however, is known to be biased downward for variances. This property is already known from linear regression, where the ML estimate for the residual variance is the sum of squares divided by the sample size n. The usually used unbiased estimate conversely divides the sum of squares by $n - p$, with p the number of columns in the design matrix $\boldsymbol{X}$. (We assume that $\boldsymbol{X}$ has full rank for simplicity.) This estimator can also be derived as the restricted maximum likelihood estimator, a likelihood-based estimator that takes the loss in degrees of freedom resulting from estimation of the fixed effects into account. Patterson and Thompson (1971) discussed the problem of variance parameter estimation and proposed to maximize not the full likelihood, but the likelihood of certain error contrasts. The method was later termed restricted maximum likelihood estimation (Harville, 1977). The idea is to use the likelihood of $n - p$ linearly independent error contrasts $\boldsymbol{A}\boldsymbol{y}$, such that $\mathrm{E}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$ and the resultant likelihood does not depend on the fixed effects. Harville (1974) shows that the resulting log-likelihood is independent, up to an additive constant, of the precise error contrast used. Possibilities include $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}$, hence the second name residual maximum likelihood. Harville (1974) also derives the corresponding log-likelihood,

$$l(\boldsymbol{\theta}) = \text{const} - \frac{1}{2}\log(\det(\boldsymbol{V})) - \frac{1}{2}\log(\det(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),$$

where $\hat{\boldsymbol{\beta}}$ is defined in (2.5). Thus, up to an additive constant,

$$l(\boldsymbol{\theta}) = \ell(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \frac{1}{2}\log(\det(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})) = \ell_P(\boldsymbol{\theta}) - \frac{1}{2}\log(\det(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})).$$

Cressie and Lahiri (1993) note that the estimating equations for $\boldsymbol{\theta}$ resulting from maximization of $l(\boldsymbol{\theta})$ are unbiased, while the corresponding equations using $\ell(\boldsymbol{\theta})$ are not. This helps explaining the smaller bias of REML estimates compared to ML estimates in smaller samples.

The ML or REML estimate $\hat{\boldsymbol{\theta}}$ can be used subsequently to obtain values for the BLUE and BLUP defined in (2.5) and (2.6). If we denote by $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{D}}$ the matrices $\boldsymbol{V}$ and $\boldsymbol{D}$ with $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\theta}}$, the estimated BLUE (EBLUE) and estimated BLUP (EBLUP) are defined as

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) &= \left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y} \quad \text{and} \\ \hat{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) &= \hat{\boldsymbol{D}}\boldsymbol{Z}'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}(\hat{\boldsymbol{\theta}}))\end{aligned}$$

respectively. Variability in the EBLUE and EBLUP thus stems from both estimation of $\boldsymbol{\beta}$ and $\boldsymbol{b}$, as well as estimation of $\boldsymbol{\theta}$. Both sources of variability should be taken into account for inference.

2.1.4 Inference for Fixed Effects

In this and the next section, we will assume normality of $\boldsymbol{\varepsilon}$ and $\boldsymbol{b}$. For known $\boldsymbol{\theta}$, we then have that $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1})$. The simplest approximation to the variance of $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ thus is $\left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}$, although this does not take into account the variability inherent in estimation of $\boldsymbol{\theta}$ and underestimates standard errors for the $\hat{\beta}_i(\hat{\boldsymbol{\theta}})$ coefficients.

An approximate Wald test for the linear hypothesis

$$H_0 : \boldsymbol{L\beta} = \boldsymbol{0} \quad \text{versus} \quad H_A : \boldsymbol{L\beta} \neq \boldsymbol{0} \tag{2.9}$$

could then be constructed using the Wald test statistic

$$W = \hat{\boldsymbol{\beta}}'\boldsymbol{L}'\left[\boldsymbol{L}\left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)\boldsymbol{L}'\right]^{-1}\boldsymbol{L}\hat{\boldsymbol{\beta}}$$

with an asymptotic chi-square distribution with $\text{rank}(\boldsymbol{L})$ degrees of freedom (Verbeke and Molenberghs, 2000). However, not only might $\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}$ be a poor estimate of Cov$(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}))$ especially for small sample sizes, but also is the asymptotic distribution

$$\hat{\boldsymbol{\beta}}'\boldsymbol{L}'\left[\boldsymbol{L}\left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)\boldsymbol{L}'\right]^{-1}\boldsymbol{L}\hat{\boldsymbol{\beta}} \overset{a}{\sim} \chi^2_{\text{rank}(\boldsymbol{L})}$$

so far lacking a theoretical foundation save in special cases (Ruppert et al., 2003). This is due to the complications arising from the random effect induced dependence in $\boldsymbol{y}$ (see model (2.1)).