Journal of Statistical Software

December 2025, Volume 115, Code Snippet 1. *doi: 10.18637/jss.v115.c01*

Local Influence Diagnostics for Nonlinear Mixed Models under the Case-Weight Perturbation Scheme in SAS

Jhessica Leticia Kirch
University of São Paulo

Geert Molenberghs
Universiteit Hasselt

Geert Verbeke
Catholic University of Leuven

César Gonçalves de Lima
University of São Paulo

Abstract

The nonlinear mixed model is a popular tool for analyzing continuous longitudinal data. This paper is primarily concerned with gauging the sensitivity of nonlinear mixed models to influential observations through local influence, which assesses the impact of small perturbations of the likelihood function. Unlike when case deletion is used, in local influence the model only needs to be fitted once, making it much more computationally appealing. The methodology is illustrated with two datasets, establishing that the local influence diagnostic can easily be applied to nonlinear mixed models through the `NLMIXED` procedure in the SAS software as a tool to identify influential individuals.

Keywords: Individual log-likelihood, influential observations, repeated measures.

1. Introduction

An important step of model building is the application of model diagnostics. This can be carried out, for example, by an influence analysis for detecting outliers and influential observations. Cook (1986) suggests that more confidence can be put in a model that is relatively stable under small perturbations. A well-known perturbation scheme is case-deletion (Cook 1977), where the individual impact of observations on the estimation is measured. This approach is called global influence analysis.

Another approach is local influence, first introduced by Cook (1986) as a general method for assessing the influence of minor perturbations of a statistical model. Beckman, Nachtsheim, and Cook (1987) used local influence to investigate how the parameters change under small perturbations of the error variances, the random-effects variances and the response vector (Verbeke and Molenberghs 2000). An alternative perturbation scheme is case-weight perturbation, where a weight ω_i is assigned to each individual in the calculation of the parameter estimates and it investigates how much the parameter estimates are affected by changes in the weights of the log-likelihood contributions of specific individuals (Verbeke and Molenberghs 1997, 2000). Cook (1986) compared the results of local influence using the case-weight perturbation scheme in normal linear regression with his measure of global influence known as Cook's distance, D_i. The author concludes that local influence and D_i are merely summaries of different characteristics of the influence graph obtained by modifying a single case-weight: Local influence measures the influence of local changes in the case weight, while D_i measures its global changes.

Lesaffre and Verbeke (1998) showed that the local influence approach is useful for the detection of influential individuals in longitudinal data analysis. Nonlinear mixed models have been widely used in longitudinal studies to describe individual response profiles and to take into account heterogeneity within subjects and between subjects as well as non-linear mean structures. General formulations of the nonlinear mixed effects model have been described by Lindstrom and Bates (1990), Pinheiro and Bates (1995), and Davidian and Giltinan (2003), among other authors.

The aim of this paper is to describe the local influence methodology for nonlinear mixed models using a case-weight perturbation scheme. Unlike case-deletion, in which the model needs to be fit $N + 1$ times, once for the entire dataset, and once for each subject deleted, in the local influence approach, the model only needs to be fitted once. This is a great advantage, especially because nonlinear mixed models can pose computational challenges. Also, the case-weight perturbation scheme has the attractive feature of being able to distinguish influence in fixed-effects parameters from that in variance components (Rakhmawati, Molenberghs, Verbeke, and Faes 2017). The general formulation for this perturbation scheme is introduced and the implementation in SAS software (SAS Institute Inc. 2023) through the `PROC NLMIXED` statement is presented. The `PROC NLMIXED` statement makes influence diagnosis computationally inexpensive once the mixed model is fitted.

This paper is organized as follows. The next section presents an overview of the nonlinear mixed effects model in the repeated measurements context. In Section 3, the local influence methodology is reviewed and in Section 4 the SAS implementation is discussed. Sections 5 and 6 illustrate the method and implementation using two different datasets analyzed with a nonlinear mixed-effects model.

2. Nonlinear mixed-effects model

Let $\boldsymbol{Y}_i$ denote the n_i-dimensional vector of repeated measurements for the i-th subject, $i = 1, ..., N$, and Y_{ij} be the j-th outcome measured for subject i. In the nonlinear mixed model, we assume that, conditioned on a q-dimensional vector of random effects $\boldsymbol{b}_i$, $\boldsymbol{Y}_i$ is normally distributed with mean vector $\boldsymbol{\mu}_i$, which can be a nonlinear function in the parameters, and with covariance matrix $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}_i$, where $\boldsymbol{I}_i$ is a n_i-dimensional identity matrix which depends on i only through its dimension n_i and $\boldsymbol{b}_i$ is assumed to be normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{D}$ $(q \times q)$ with element $d_{ij} = d_{ji}$, i.e.,

$$\boldsymbol{Y}_i \mid \boldsymbol{b}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

$$\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{D}).$$

Let $\boldsymbol{\beta}$ be the p-dimensional vector of fixed effects parameters and let $f_i(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\beta}, \sigma^2)$ be the density function of $\boldsymbol{Y}_i \mid \boldsymbol{b}_i$ and $f(\boldsymbol{b}_i \mid \boldsymbol{D})$ be the density function of $\boldsymbol{b}_i$, then the marginal density function of $\boldsymbol{Y}_i$ is given by

$$f_i(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) = \int f_i(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\beta}, \sigma^2) f(\boldsymbol{b}_i \mid \boldsymbol{D}) d\boldsymbol{b}_i. \tag{1}$$

The log-likelihood for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \text{vec}\,(\boldsymbol{D}), \sigma^2)^\top$, where $\text{vec}\,(\boldsymbol{D})$ is the vector of unique elements in $\boldsymbol{D}$, is derived as

$$\ell(\boldsymbol{\theta}) = \log\left[\prod_{i=1}^{N} f_i(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)\right] = \log\left[\prod_{i=1}^{N} \int f_i(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) f(\boldsymbol{b}_i \mid \boldsymbol{D}) d\boldsymbol{b}_i\right]. \tag{2}$$

From (2), it follows that the log-likelihood function can be rewritten as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}), \tag{3}$$

where $\ell_i(\boldsymbol{\theta})$ is the contribution of the i-th individual to the log-likelihood.

We proceed by maximum likelihood. In general, it is not possible to find the analytic expressions for the integrals in $\ell(\boldsymbol{\theta})$ and numerical approximations are needed (Pinheiro and Bates 1995; Davidian and Giltinan 2003; Molenberghs and Verbeke 2005). It is also useful to calculate estimates for the random effects $\boldsymbol{b}_i$ as well. This can be done using empirical Bayes estimates. Such estimates reflect how much the subject-specific profiles deviate from the overall average profile (Molenberghs and Verbeke 2005).

3. Local influence

Let $\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega})$ denote a perturbed version of $\ell(\boldsymbol{\theta})$, depending on a vector $\boldsymbol{\omega}$ of weights. For the detection of influential subjects, the perturbed log-likelihood will be given by

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega}) = \sum_{i=1}^{N} \omega_i \ell_i(\boldsymbol{\theta}), \tag{4}$$

where $\boldsymbol{\omega}$ is now a $(N \times 1)$ vector of weights. When $\omega_i = 1 \;\forall\; i$, we have the classical log-likelihood (3). Also, the log-likelihood with the i-th individual removed corresponds to the vector $\boldsymbol{\omega}$ with $\omega_i = 0$ and $\omega_j = 1 \;\forall\; j \neq i$ (Verbeke and Molenberghs 2000).

Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator for $\boldsymbol{\theta}$ obtained by maximizing $\ell(\boldsymbol{\theta})$ and let $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ be the maximum likelihood estimator for $\boldsymbol{\theta}$ under $\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega})$, the local influence approach compares $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ based on the likelihood displacement $LD(\boldsymbol{\omega}) = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})]$, so that the variability of $\hat{\boldsymbol{\theta}}$ is taken into consideration. $LD(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\theta})$ is strongly curved at $\hat{\boldsymbol{\theta}}$ and small if $\ell(\boldsymbol{\theta})$ is fairly flat at $\hat{\boldsymbol{\theta}}$ (Verbeke and Molenberghs 2000).

A graph of $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ contains essential information on the influence of case-weight perturbations. Cook (1986) refers to this graph as an influence graph, which represents a geometric surface formed by the values of the $(N+1)$-dimensional vector $\boldsymbol{\zeta}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, LD(\boldsymbol{\omega}))^\top$. Graphically depicting the influence graph is only possible in cases where the number of weights $\boldsymbol{\omega}$ does not exceed two, so alternative methods are needed to extract the most relevant information from an influence graph (Verbeke and Molenberghs 2000).

Cook (1986) proposes studying the local behavior $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0 = (1, 1, ..., 1)^\top$ as this describes how sensitive $\ell(\hat{\boldsymbol{\theta}})$ is to small perturbations of the case weights. This was done using the normal curvature $C_{\boldsymbol{l}}$ of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0 = (1, 1, ..., 1)^\top$, in the direction of a vector $\boldsymbol{l}$ of unit length. Let $\boldsymbol{\Delta}_i$ be the s-dimensional vector defined by

$$\boldsymbol{\Delta}_i = \left.\frac{\partial^2 \ell_i(\boldsymbol{\theta} \mid \omega_i)}{\partial \omega_i \partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

where $s = p + 1 + q(q+1)/2$. Then, $\boldsymbol{\Delta}$ is the matrix with i-th column equal to $\boldsymbol{\Delta}_i$. Also, let $\ddot{\boldsymbol{L}}$ be the $(s \times s)$ matrix of second-order derivatives of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, i.e.,

$$\ddot{\boldsymbol{L}} = \left.\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

Cook (1986) showed that, for any unit vector $\boldsymbol{l}$, $C_{\boldsymbol{l}}$ is given by $C_{\boldsymbol{l}} = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}^{-1} \boldsymbol{\Delta} \boldsymbol{l}|$. Also note that using the perturbed log-likelihood for the detection of influential subjects in (4), we have

$$\boldsymbol{\Delta}_i = \left.\frac{\partial^2 \omega_i \ell_i(\boldsymbol{\theta})}{\partial \omega_i \partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \left.\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

A possible choice for $\boldsymbol{l}$ corresponds to the direction of the i-th individual, that is, the vector $\boldsymbol{l}_i$ that contains zeros everywhere except on the i-th position, where there is a one. The resulting local influence is then given by

$$C_{\boldsymbol{l}_i} = C_i = 2|\boldsymbol{\Delta}_i^\top \ddot{\boldsymbol{L}}^{-1} \boldsymbol{\Delta}_i|,$$

where a large C_i indicates a large impact of the i-th individual, in a local sense, on the total parameter vector. Lesaffre and Verbeke (1998) suggest the use of cut-off values at which individuals with twice the average value should be considered influential $(2\sum_{i=1}^N C_i/N)$.

It is possible to extend the approach to measure the local influences of individual i on the fixed effects and on the variance components, separately. This decomposition of C_i suggests a practical procedure to find an explanation for the influential nature of an individual (Verbeke and Molenberghs 2000). Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, where $\boldsymbol{\theta}_1$ contains all the fixed effects parameters and $\boldsymbol{\theta}_2$ contains all variance components of the model, and partitioning $\ddot{\boldsymbol{L}}$ as

$$\ddot{\boldsymbol{L}} = \begin{bmatrix} \ddot{\boldsymbol{L}}_{11} & \ddot{\boldsymbol{L}}_{12} \\ \ddot{\boldsymbol{L}}_{21} & \ddot{\boldsymbol{L}}_{22} \end{bmatrix},$$

according to the dimension of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Then, the local influence of subject i on the estimation of the fixed effects parameters, denoted by $C_i(\boldsymbol{\theta}_1)$, and for the variance components, denoted by $C_i(\boldsymbol{\theta}_2)$, are given by

$$C_i(\boldsymbol{\theta}_1) = 2\left|\boldsymbol{\Delta}_i^\top \left(\ddot{\boldsymbol{L}}^{-1} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\boldsymbol{L}}_{22}^{-1} \end{bmatrix}\right) \boldsymbol{\Delta}_i\right|,$$

$$C_i(\boldsymbol{\theta}_2) = 2\left|\boldsymbol{\Delta}_i^\top\left(\ddot{\boldsymbol{L}}^{-1} - \begin{bmatrix}\ddot{\boldsymbol{L}}_{11}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\right)\boldsymbol{\Delta}_i\right|.$$

4. SAS implementation

Nonlinear mixed models can be fitted using the SAS software (SAS Institute Inc. 2023) through the `PROC NLMIXED` statement. Using the `HESS` and `SUBGRADIENT` options, the `PROC NLMIXED` statement enables the display of the Hessian matrix and the subject-specific gradients of the integrated, marginal log-likelihood concerning all parameters.

The Hessian matrix corresponds to the second-order derivative of the negative marginal log-likelihood function, i.e., $-\ddot{\boldsymbol{L}}$, and the subject-specific gradient corresponds to the first-order derivative of the contribution of the i-th subject to the negative marginal log-likelihood, $-\boldsymbol{\Delta}_i$. The calculation of local influence can then be done within the `PROC IML` statement as a matrix multiplication.

We have implemented the steps discussed above in SAS version 9.4. and SAS/IML 15.1.

5. Analysis of the orange tree dataset

Consider the orange tree data of Draper and Smith (1981), which consists of seven measurements of the trunk circumference (in millimeters) on each of five orange trees. Lindstrom and Bates (1990) and Pinheiro and Bates (1995) suggested the following nonlinear mixed-effects model:

$$Y_{ij} = \frac{\beta_1 + b_i}{1 + \exp\left[-(t_{ij} - \beta_2)/\beta_3\right]} + \varepsilon_{ij},$$

with $b_i \sim N(0, d)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$, where Y_{ij} represents the j-th circumference measurement on the i-th tree and t_{ij} is the time in days ($i = 1, ..., N = 5$ and $j = 1, ..., n_i = n = 7$). Parameters estimates for this model can be found in Pinheiro and Bates (1995). The individual growth profiles and the fitted values of the five trees are presented in Figure 1.

The analytical expressions of the individual log-likelihood and of the first-order derivative of $\ell_i(\boldsymbol{\theta})$ with respect to each of the elements of $\boldsymbol{\theta}$ are available in Appendices A and B. Using the `NLMIXED` procedure, we can obtain the second derivative of the negative log-likelihood with respect to β_1, β_2, β_3, d and σ^2, using the `HESS` statement and the first-order derivative of the individual negative log-likelihood with respect to each one of the parameters using the `SUBGRADIENT` statement, producing the $-\ddot{\boldsymbol{L}}$ and $-\boldsymbol{\Delta}_i$ matrices, respectively. For the orange tree dataset, we use the following code to fit the model and simultaneously store the Hessian matrix into `h` and subgradient into `gradi`:

```
proc nlmixed data = OrangeTree HESS SUBGRADIENT = gradi;
    parms beta1 = 190 beta2 = 700 beta3 = 350 d = 1000 sigma2 = 60;
    num = b + beta1;
    den = 1 + exp(-(Days - beta2) / beta3);
    model y ~ normal(num / den, sigma2);
    random b ~ normal(0, d) subject = Subject;
    ods output Hessian = h;
run;
```

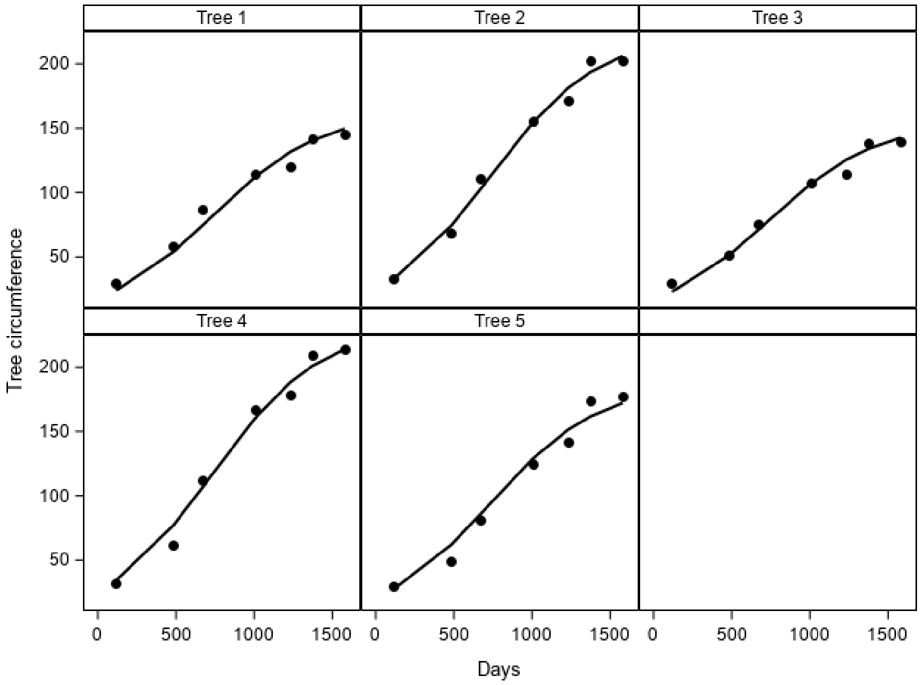


Figure 1: Observed (circles) and fitted profiles (lines) for the orange trees dataset.

The local influence is then obtained using the `IML` procedure and the SAS code to generate the C_i values for each subject is given below. The names of individuals must be given as characters in the dataset so that they are read correctly into variable `S`.

The complete SAS code to generate the C_i on the fixed-effects parameters and on the variance components is available as supplementary materials. Note that it is necessary to inform the number of fixed effect parameters (`f = 3`) and the number of random parameters (`r = 2`) to produce the local influence measures of individuals on the subvectors of $\boldsymbol{\theta}$. Part of the results generated by the function are displayed in Table 1.

```
proc iml;
    use h;
    read all var _NUM_ into L;
    L = -L[1:nrow(L), 2:ncol(L)];
    close h;
    use gradi;
    read all var _NUM_ into Delta;
    Delta = -Delta`;
    read all var _CHAR_ into S[colname = Names];
    close gradi;
    Ci = vecdiag(2 * abs(Delta` * inv(L) * Delta));
    create Local var {"S" "Ci"};
    append;
    close Local;
run;
```

Tree	C_i	$C_i(\theta_1)$	$C_i(\theta_2)$
T1	1.34438	1.33754	0.01007
T2	0.54546	0.50543	0.04019
T3	1.04095	0.79212	0.25279
T4	1.56653	1.39130	0.15089
T5	1.57305	1.29748	0.20742

Table 1: Local influences C_i for each tree and local influence on the fixed effects and on the variance components, $C_i(\theta_1)$ and $C_i(\theta_2)$, respectively, for the orange trees dataset.

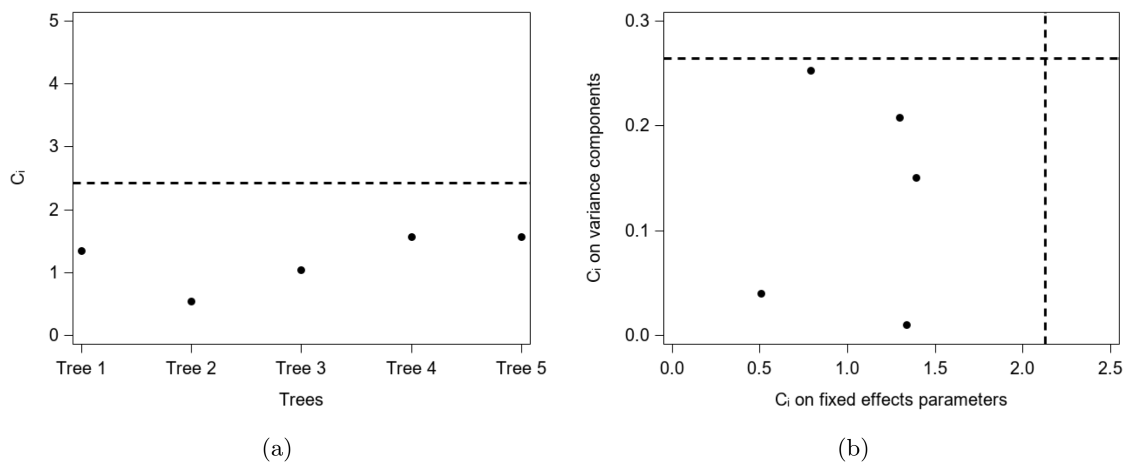


Figure 2: (a) Plot of the local influences C_i for all trees in the orange trees dataset. (b) Plot of C_i on fixed effects parameters versus C_i on variance components for the orange trees dataset.

Figure 2 displays overall C_i and influences for subvectors of fixed effects and variance components under the case-weight perturbation scheme. No individual has a large impact on the parameter estimates, measured by C_i (Figure 2(a)), falling all below the cut-off value of $2\sum_{i=1}^{N} C_i/N \approx 2.43$. Similarly, the cut-off values for $C_i(\theta_1)$ and $C_i(\theta_2)$ are 2.13 and 0.26, respectively (Figure 2(b)). These cut-off values are indicated in Figure 2 by the dashed lines. These results show that the model fit is not excessively influenced by a specific tree.

6. Analysis of the songbird dataset

Van der Linden *et al.* (2002) and Van Meir *et al.* (2004) presented a morphological study in songbirds where they analyzed by magnetic resonance imaging (MRI) the effect of testosterone on the dynamics of manganese accumulation in the nucleus robustus arcopallii (RA) and a less studied area X of 10 female European starlings that had been injected with manganese in their high vocal center. A schematic representation of the song control system in the bird brain was given by the authors.

All birds were studied by MRI before and after the implementation of a subcutaneous capsule in the neck region. The five treated birds (birds 6 to 10) received a capsule of crystalline testosterone and the capsule was left empty for the five control birds (birds 1 to 5).

6.1. MRI signal intensities at nucleus RA

Van der Linden *et al.* (2002) employed the following parametric shape for a bird's profile:

$$SI_{ij}(RA) = \frac{(\phi_0 + \phi_1 G_i) T_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0 + \tau_1 G_i)^{\eta_0+\eta_1 G_i} + T_{ij}^{\eta_0+\eta_1 G_i}} + \gamma_0 + \gamma_1 G_i + \varepsilon_{ij}, \tag{5}$$

where $SI_{ij}(RA)$ is the measurement of the MRI signal intensity for a region of interest RA at occasion j for bird i, G_i is an indicator for group membership (1 for testosterone treated birds and 0 otherwise), and T_{ij} is the measurement time, referring to the before versus after treatment epoch. Under this model, the maximal signal intensity equals $\phi_0 + \phi_1 G_i$, the time required to reach 50% of this maximum (T_{50}) is given by $\tau_0 + \tau_1 G_i$, and the shape of the curve is $\eta_0 + \eta_1 G_i$. In (5), ϕ_0, ϕ_1, τ_0, τ_1, η_0, and η_1 are fixed-effects parameters.

Based on (5), but taking into account the correlation between the repeated measurements in the same bird and after inferences about the fixed and the bird-specific random effects included in the model, Serroyen *et al.* (2005) suggest to fit the MRI signal intensities (SI) at RA at the first period using the following model

$$SI_{ij}(RA) = \frac{(\phi_0 + f_i) T_{ij}^{\eta_0+n_i}}{(\tau_0 + t_i)^{\eta_0+n_i} + T_{ij}^{\eta_0+n_i}} + \varepsilon_{ij}. \tag{6}$$

As expected, no treatment effect is seen in (6) as it fits the period before subcutaneous capsule implementation. So, in this model, ϕ_0, η_0, and τ_0 are fixed effects parameters and the f_i, t_i and n_i are bird-specific random effects, where $(f_i, t_i, n_i)^\top \sim N(\mathbf{0}, \boldsymbol{D})$ with $\text{cov}(f_i, n_i) = (t_i, n_i) = 0$. Parameters estimates for this model can be found in Serroyen *et al.* (2005). The observed and fitted profiles of MRI signal intensities at RA for the first (before the treatment) and second period (after the treatment), whose model is defined later, for each bird separately are presented in Figure 3.

Figure 4 displays overall C_i and influences for subvectors of $\boldsymbol{\theta}$ under the case-weight perturbation scheme. Individuals with C_i larger than twice the average value (dashed line) are considered as influential and the most influential subjects are indicated by their identification number. The SAS code to obtain these results is available as supplementary material.

Bird 5 has a large impact on the parameter estimates, measured by C_i, falling above the cut-off value of approximately 7.77 in Figure 4(a), indicated by the dashed line. For the fixed-effects parameters, the cut-off value is 1.16 and for the variance components it equals 6.60. Figure 4(b) reveals that both fixed and random effects parameters of the fitted model are affected by the influential bird. Notably, Bird 5 has the highest MRI signal intensities at RA before the treatment implementation (Figure 3). Also, even considering that Bird 5 is influential for both the fixed effects and the random components, the impact is more pronounced in the random effects parameters (Figure 4(b)).

For the second period, after the treatment, the proposed model by Serroyen *et al.* (2005) was

$$SI_{ij}(RA) = \frac{(\phi_0 + f_i) T_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0 + t_i)^{\eta_0+\eta_1 G_i} + T_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij}. \tag{7}$$

Now, ϕ_0, η_0, η_1, and τ_0 are fixed effects parameters and f_i and t_i are bird-specific random effects, where $(f_i, t_i)^\top \sim N(\mathbf{0}, \boldsymbol{D})$ with $\text{cov}(f_i, t_i) = 0$, and a group effect appears in (7). The observed values and the fitted profiles can be seen in Figure 3.

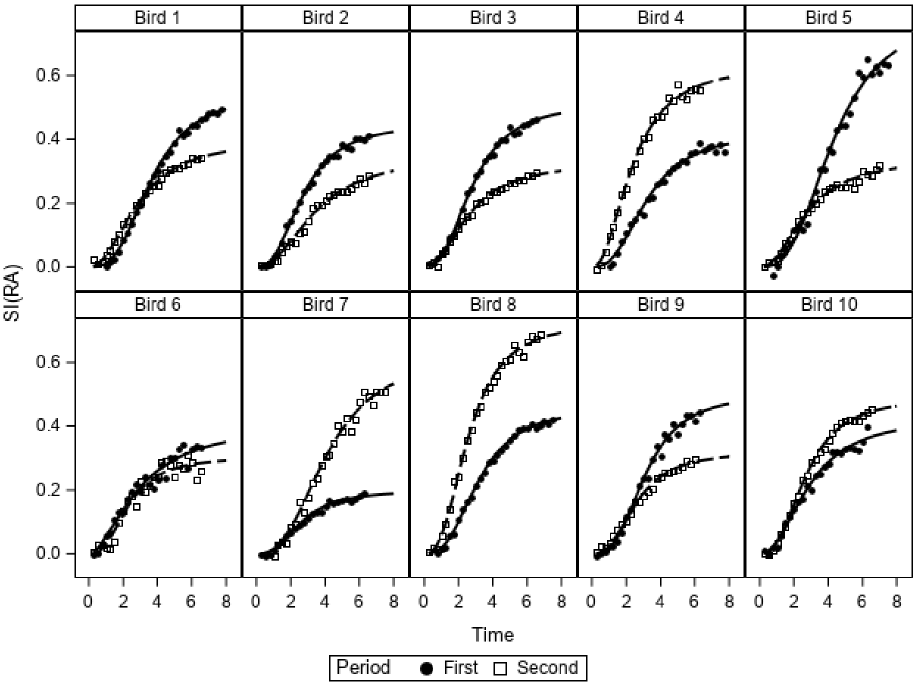
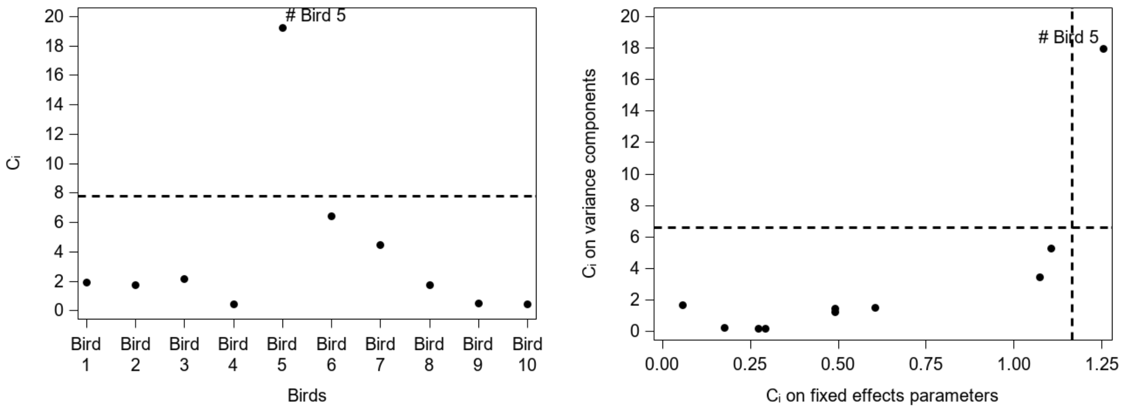


Figure 3: Observed (circles and squares) and fitted (lines) profiles for MRI signal intensities at RA for the first and second periods for the songbird dataset.



(a) (b)

Figure 4: (a) Plot of the local influences C_i for all birds in the songbird dataset before treatment. (b) Plot of C_i on fixed effects parameters versus C_i on variance components for the songbird dataset before treatment. The most influential bird is indicated by its identification number.

Birds 6 and 7, both treated, have a large impact on the parameter estimates, measured by C_i, falling above the cut-off value of approximately 6.75 (Figure 5(a)). For the fixed effects parameters, the cut-off value is 1.85 and for the variance components it equals 4.88. These cut-off values are represented in Figure 5 by the dashed lines. Bird 6 is only influential for the

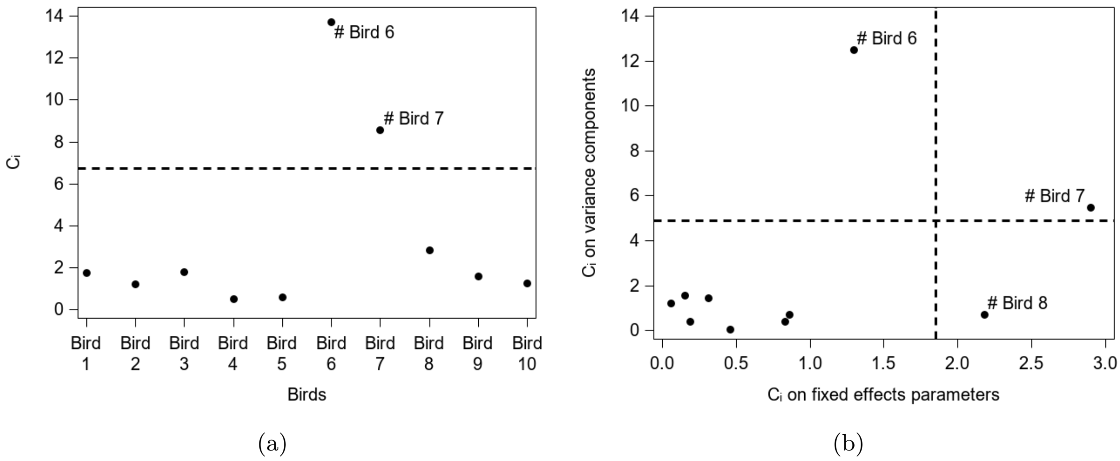


Figure 5: (a) Plot of the local influences C_i for all birds in the songbird dataset after treatment. (b) Plot of C_i on fixed effects parameters versus C_i on variance components for the songbird dataset after treatment. The most influential birds are indicated by their identification number.

estimation of the variance components in the model while Bird 7 was considered influential in both parts of the model (Figure 5(b)). Although Bird 8 is not considered influential overall, it appears to be influential on the fixed effects parameters only.

Another nonlinear mixed model was fitted to the data in order to compare the results of the local influence diagnostic. This model is inspired by a pharmacokinetic two-compartment model. For numerical purposes, the response variable SI at RA was multiplied by 100 and the random effects were considered uncorrelated. The initial model was given by

$$SI_{ij}(RA) = \mathrm{e}^{\beta_0+\beta_1 G_i+b_i} \exp\left[-\mathrm{e}^{\phi_0+f_i} T_{ij}\right] - \mathrm{e}^{\tau_0+\tau_1 G_i+t_i} \exp\left[-\mathrm{e}^{\gamma_0+g_i} T_{ij}\right] + \varepsilon_{ij}. \tag{8}$$

Backward selection was conducted using likelihood ratio tests, resulting in the following model for the first period:

$$SI_{ij}(RA) = \mathrm{e}^{\beta_0+b_i} - \mathrm{e}^{\tau_0+t_i} \exp\left[-\mathrm{e}^{\gamma_0+g_i} T_{ij}\right] + \varepsilon_{ij}$$

and, for the second period,

$$SI_{ij}(RA) = \mathrm{e}^{\beta_0+b_i} \exp\left[-\mathrm{e}^{\phi_0} T_{ij}\right] - \mathrm{e}^{\tau_0+t_i} \exp\left[-\mathrm{e}^{\gamma_0+g_i} T_{ij}\right] + \varepsilon_{ij}.$$

Parameter estimates and standard errors for the final model for both periods are provided in Table 2 and the observed and fitted profiles for each bird separately are presented in Figure 6.

The same bird (Bird 5) was influential for the parameter estimates of the two different models fitted for the SI at RA for the first period (Figure 4 and Figures 7(a) and 7(b)). For the second period, after treatment, birds 6 and 7 show the same behavior in Figures 5(b) and 7(d), with bird 6 being influential on the estimates of the variance components of the model and bird 7 on the fixed effect parameters. Bird 8 was not considered influential in the first fitted model, but appeared as influential for the estimation of the fixed effect parameters (Figure 5(b)). In the two-compartment model, this bird was found influential for the complete parameter

Effect	Parameter	Estimate	
		First period	Second period
	β_0	4.1314 (0.1177)	4.2216 (0.1950)
	ϕ_0		-4.1759 (0.7608)
	τ_0	4.2835 (0.1158)	4.3201 (0.1894)
	γ_0	-1.6841 (0.1162)	-1.6647 (0.1496)
var(b_i)	d_{11}	0.1190 (0.0620)	0.0837 (0.0385)
var(t_i)	d_{33}	0.1214 (0.0614)	0.1169 (0.0667)
var(g_i)	d_{44}	0.0871 (0.0544)	0.1073 (0.1167)
var(ε_{ij})	σ^2	4.8995 (0.4600)	4.6653 (0.4400)

Table 2: Parameter estimates and standard errors (values in parentheses) for the two-compartment model fitted to MRI signal intensities at RA for the first and second periods for the songbird dataset. The SI at RA values were multiplied by 100.

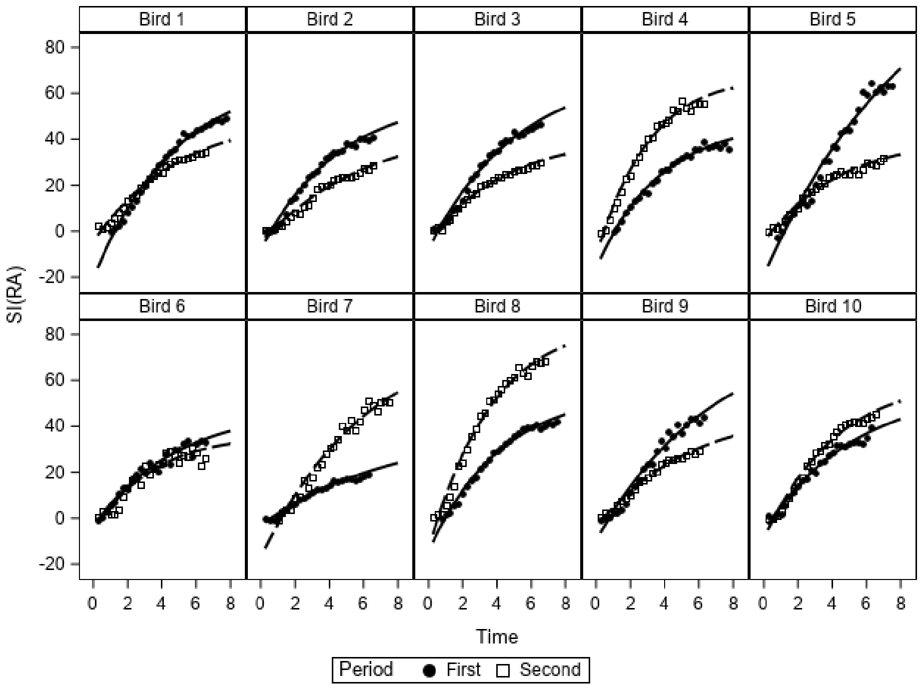


Figure 6: Observed (circles and squares) and fitted (lines) profiles using the two-compartment model for MRI signal intensities at RA for the first and second periods for the songbird dataset. The SI at RA values were multiplied by 100.

vector as well as for the estimation of the fixed effects, but not for the estimation of the variance components (Figures 7(c) and 7(d)). This shows that for the songbird dataset, even with different models, the same animals were identified by the local influence diagnosis.

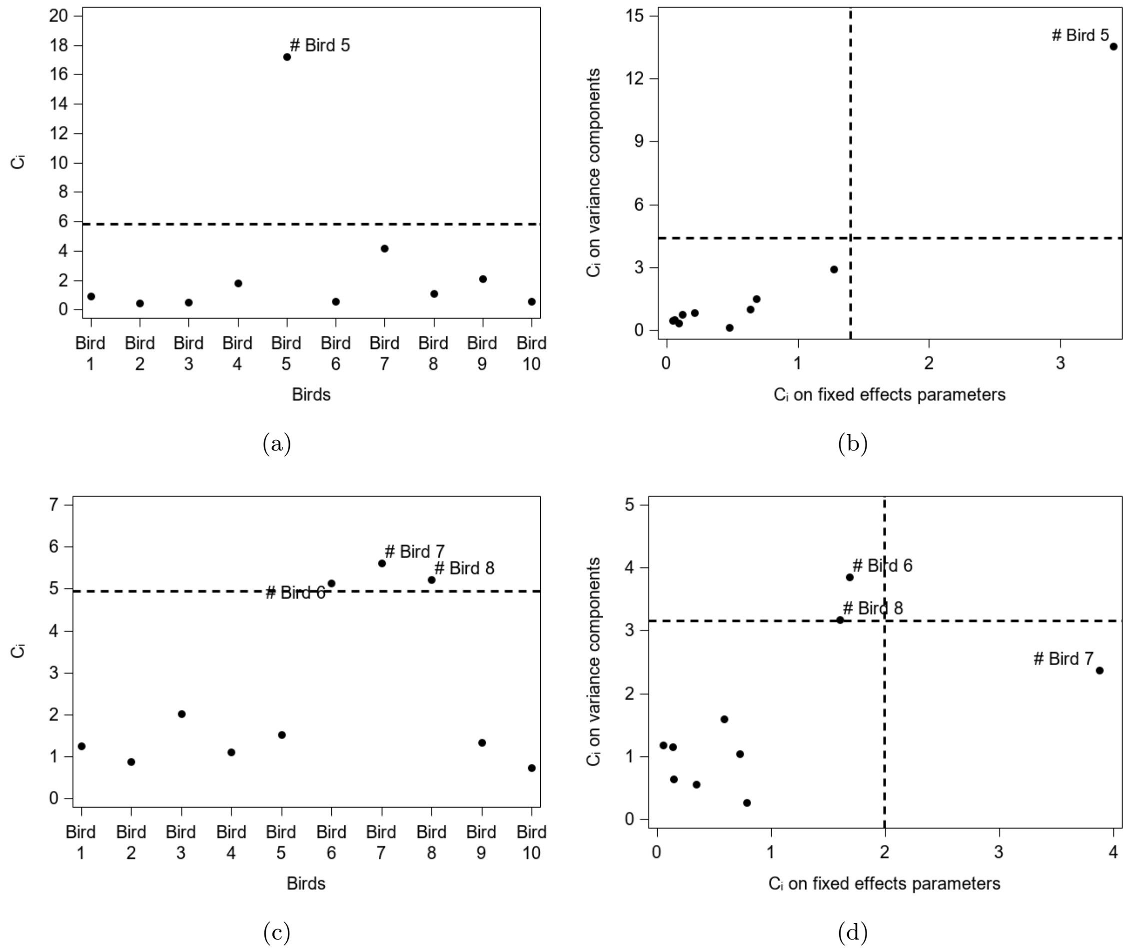


Figure 7: (a) Plot of the local influences C_i using the two-compartment model for the songbird dataset before treatment. (b) Plot of C_i on fixed effects parameters versus C_i on variance components using the two-compartment model for the songbird dataset before treatment. (c) Plot of the local influences C_i using the two-compartment model for the songbird dataset after treatment. (d) Plot of C_i on fixed effects parameters versus C_i on variance components using the two-compartment model for the songbird dataset after treatment. The most influential birds are indicated by their identification number.

6.2. MRI signal intensities at area X

To fit the MRI signal intensities at area X, Van der Linden *et al.* (2002) employed the same nonlinear model presented in (5). After model simplification, we obtained the final model for the first period as

$$SI_{ij}(\text{area } X) = \frac{(\phi_0 + f_i)T_{ij}^{\eta_0+\eta_1 G_i}}{\tau_0^{\eta_0+\eta_1 G_i} + T_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij}. \quad (9)$$

Now, we have that the model for the first period has a treatment effect at the shape of the curve, which was not expected, since the first period is before the implementation of a subcutaneous capsule at the treated birds. This indicates a significant difference between the

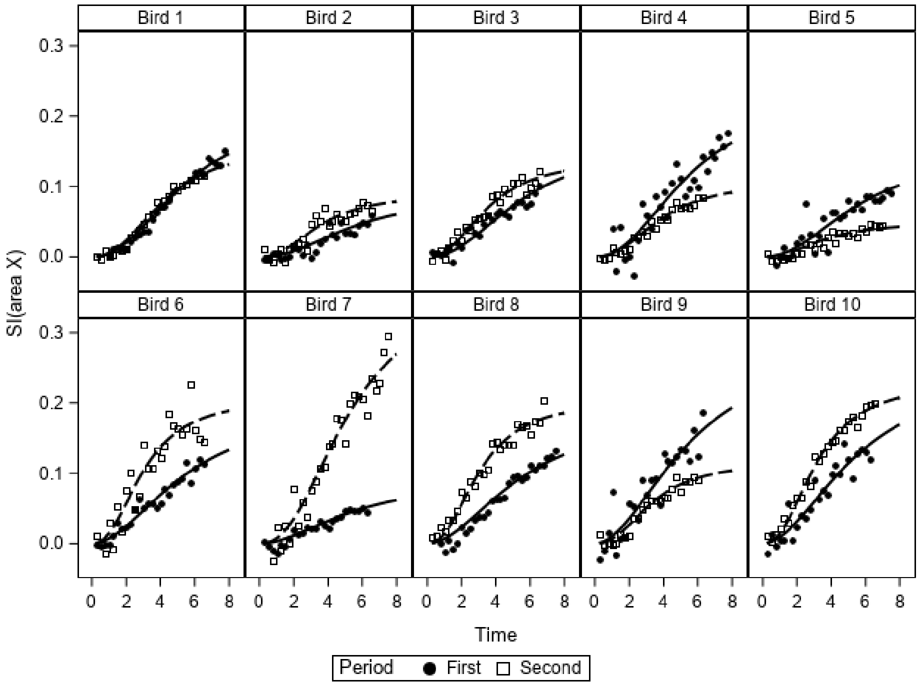


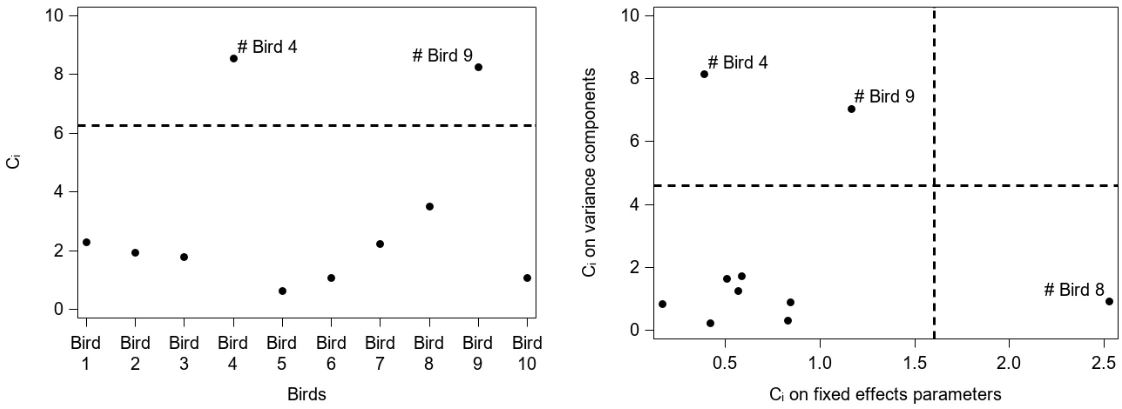
Figure 8: Observed (circle and squares) and fitted (lines) profiles for MRI signal intensities at area X for the first and second periods for the songbird dataset.

Effect	Parameter	Estimate	
		First period	Second period
	ϕ_0	0.1864 (0.0346)	0.1032 (0.0260)
	ϕ_1		0.1330 (0.0310)
	η_0	2.2168 (0.2579)	2.3492 (0.1497)
	η_1	-0.2925 (0.1566)	
	τ_0	5.5324 (0.8421)	3.7190 (0.3222)
$\mathrm{var}(f_i)$	d_{11}	0.0039 (0.0021)	0.0042 (0.0022)
$\mathrm{var}(t_i)$	d_{22}		0.4923 (0.2780)
$\mathrm{cov}(f_i, t_i)$	d_{12}		0.0339 (0.0225)
$\mathrm{var}(\varepsilon_{ij})$	σ^2	0.0002 (0.00002)	0.0002 (0.00002)

Table 3: Parameter estimates and standard errors (values in parentheses) for the model fitted to MRI signal intensities at area X for the first and second periods for the songbird dataset.

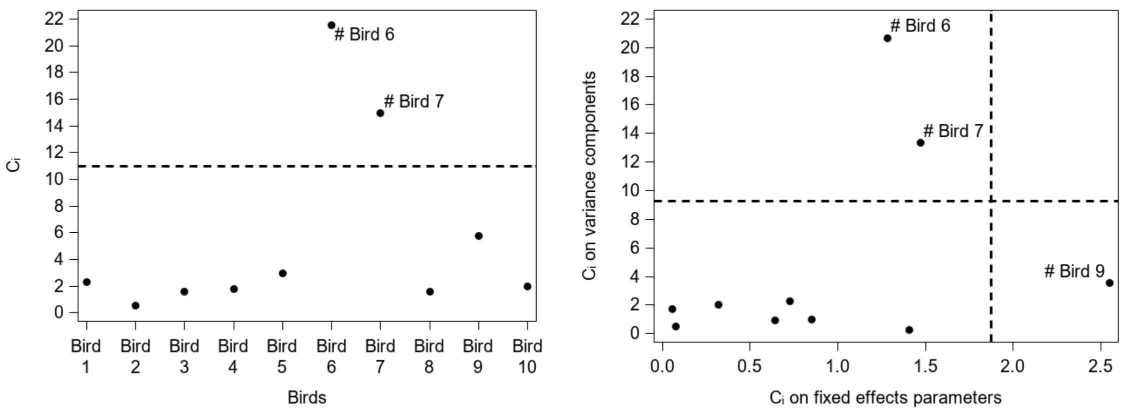
treated and control groups before the treatment was administered. Such a difference should be interpreted with caution and is likely be due to chance. The observed and fitted profiles for MRI signal intensities at area X in each bird are presented in Figure 8 and the parameter estimates were presented in Table 3.

Through the C_i value of local influence, birds 4 and 9 were considered influential subjects in the first period, falling above the cut-off value of 6.27 in Figure 9(a). Bird 4 was a control bird and bird 9, a treated one, but both had the highest observed values of SI at area X in the first period (Figure 8). The two influential birds were considered influential for the estimation



(a) (b)

Figure 9: (a) Plot of the local influences C_i for all birds in the songbird dataset before treatment. (b) Plot of C_i on fixed effects parameters versus C_i on random effects parameters for the songbird dataset before treatment. The most influential birds are indicated by their identification number.



(a) (b)

Figure 10: (a) Plot of the local influences C_i for all birds in the songbird dataset after treatment. (b) Plot of C_i on fixed effects parameters versus C_i on random effects parameters for the songbird dataset after treatment. The most influential birds are indicated by their identification number.

of the variance components in the model (Figure 9(b)), considering the cut-off value of 4.60. In (9), only two parameters correspond to the variance component elements, which are σ^2 and the variance of the bird-specific random effect f_i, which is related to the maximal signal intensity at area X.

To fit the MRI signal intensities at area X for the second period, we use the following model

$$SI_{ij}(\text{area } X) = \frac{(\phi_0 + \phi_1 G_i + f_i) T_{ij}^{\eta_0}}{(\tau_0 + t_i)^{\eta_0} + T_{ij}^{\eta_0}} + \varepsilon_{ij},$$

Effect	Parameter	Estimate	
		First period	Second period
	β_0	3.2895 (0.2109)	2.8826 (0.1805)
	β_1		0.7353 (0.1882)
	τ_0	3.3772 (0.1922)	2.9768 (0.1700)
	τ_1		0.7117 (0.1865)
	γ_0	-2.4537 (0.2238)	-2.2021 (0.1558)
var(b_i)	d_{11}	0.1140 (0.0552)	0.0847 (0.0407)
var(t_i)	d_{33}	0.1092 (0.0544)	0.0808 (0.0399)
var(ε_{ij})	σ^2	2.0841 (0.1927)	1.9228 (0.1767)

Table 4: Parameter estimates and standard errors (values in parentheses) for the two-compartment model fitted to MRI signal intensities at area X for the first and second periods for the songbird dataset.

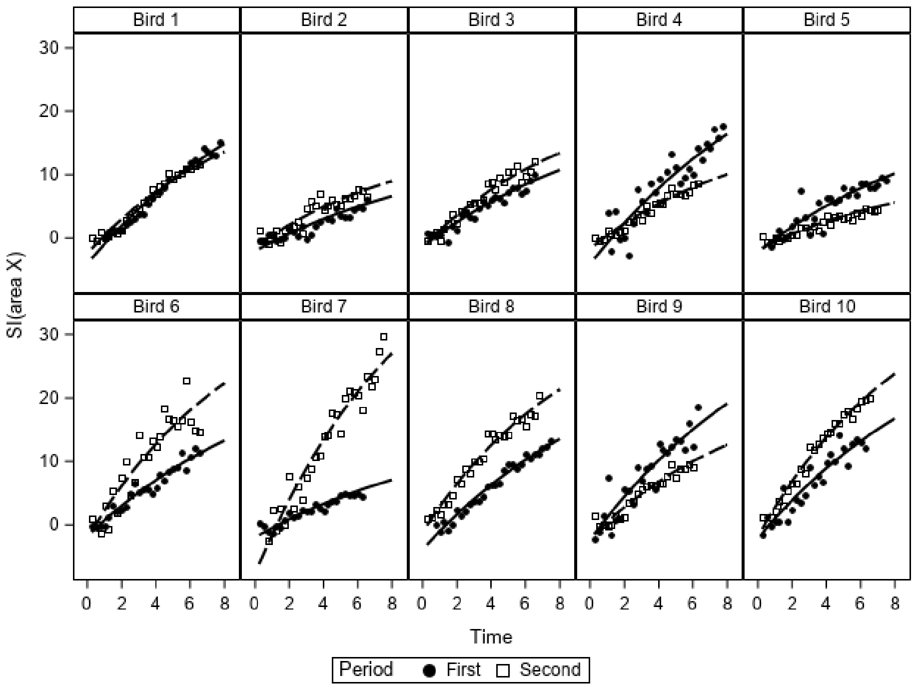


Figure 11: Observed (circles and squares) and fitted (lines) profiles using the two-compartment model for MRI signal intensities at area X for the first and second periods for each bird separately for the songbird dataset. The SI at area X values were multiplied by 100.

in which a treatment effect is included only for the maximal signal intensity (ϕ_1), and random effects parameters related to the maximal signal intensity and the time required to reach 50% of this maximum. The model estimates are presented in Table 3 and the fitted profiles can be seen in Figure 8.

Now, for the second period (after treatment), two treated birds were considered influential (birds 6 and 7), falling above the cut-off value of 10.98 (Figure 10(a)). Also, both of them

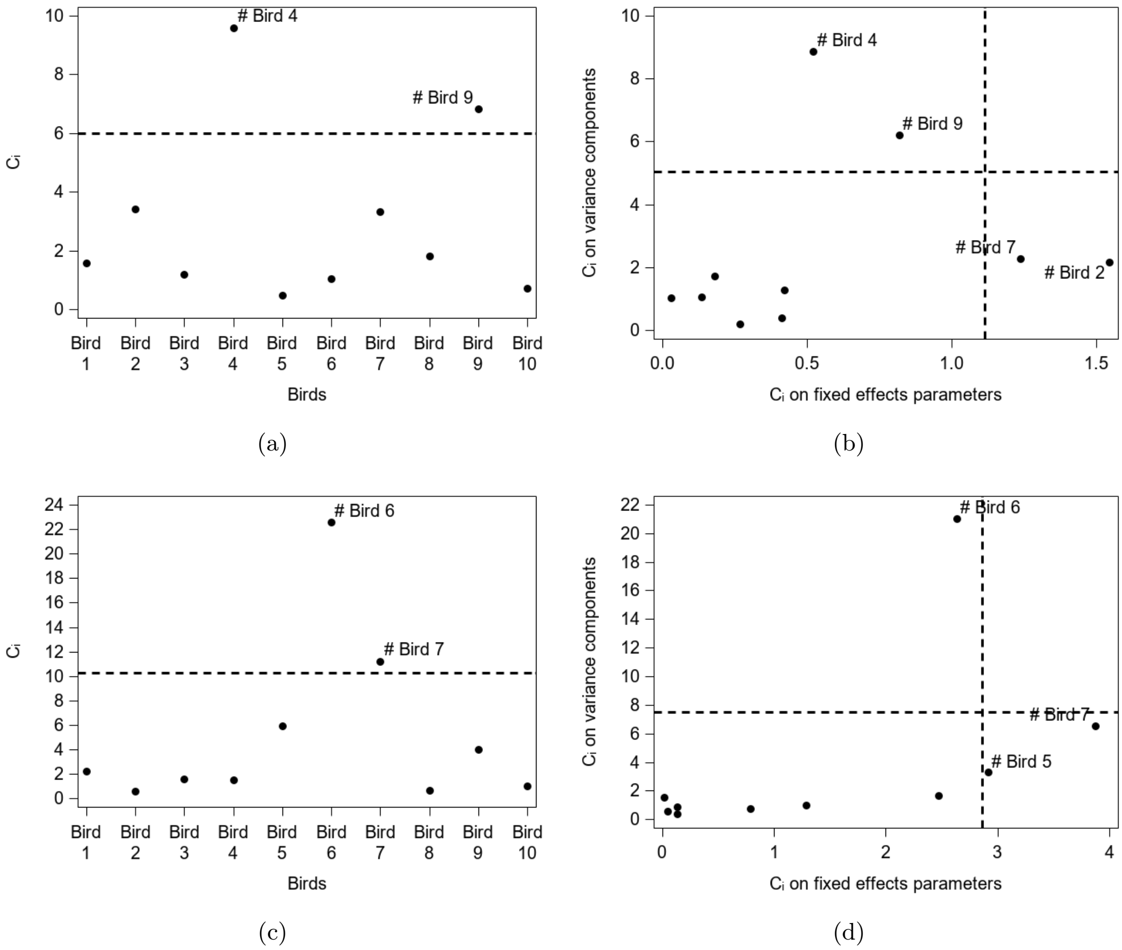


Figure 12: (a) Plot of the local influences C_i using the two-compartment model for the songbird dataset before treatment. (b) Plot of C_i on fixed effects parameters versus C_i on variance components using the two-compartment model for the songbird dataset before treatment. (c) Plot of the local influences C_i using the two-compartment model for the songbird dataset after treatment. (d) Plot of C_i on fixed effects parameters versus C_i on variance components using the two-compartment model for the songbird dataset after treatment. The most influential birds are indicated by their identification number.

were considered influential for the estimation of the variance components in the model (Figure 10(b)). Bird 9 was not considered influential overall but it was shown to be influential for the estimation of the fixed-effect parameters. This bird is the only treated bird that shows a lower SI at area X value after the testosterone treatment was administered. The SAS code for obtaining local influence for both periods can be found in the supplementary material.

As was done for the fit of the SI at RA, another nonlinear mixed model was fitted to the data. Again, the response variable was multiplied by 100 and the initial model was given by (8) and after the inferences about the fixed and the bird-specific random effects included in the model, we fitted the following model:

$$SI_{ij}(\text{area } X) = \mathrm{e}^{\beta_0 + b_i} - \mathrm{e}^{\tau_0 + t_i} \exp\left[-\mathrm{e}^{\gamma_0} T_{ij}\right] + \varepsilon_{ij}$$

for the first period and

$$SI_{ij}(\text{area } X) = \mathrm{e}^{\beta_0+\beta_1 G_i+b_i} - \mathrm{e}^{\tau_0+\tau_1 G_i+t_i} \exp\left[-\mathrm{e}^{\gamma_0} T_{ij}\right] + \varepsilon_{ij}$$

for the second period. Parameter estimates and standard errors for the final two-compartment model for both periods are provided in Table 4 and the observed and fitted profiles for each bird separately are presented in Figure 11.

The same birds were considered influential for the parameter estimates of the two fitted models. Birds 4 and 9 are considered influential in Figures 9(a) and 12(a) and birds 6 and 7 are influential in Figures 10(a) and 12(c). Some small differences were observed in the plots of the C_i on the fixed effects parameters versus C_i on variance components, in which different birds appear above the cut-off values. Also, while in Figures 9(b) and 12(b) the two influential birds appear to be influential on the estimates of the variance components, in Figures 10(b) and 12(d), bird 7 appears to be influential in different parts of the model.

7. Concluding remarks

In this paper, we have considered the local influence methodology for the nonlinear mixed models under the case-weight perturbation scheme and we have presented a set of analyses in order to illustrate the methodology in two different datasets. The illustrative analysis aimed to show how the local influence diagnostic can easily be applied to nonlinear mixed models through the `PROC NLMIXED` statement in SAS software as a means of identifying influential individuals.

For the case-deletion approach, the model needs to be fitted $N + 1$ times, once for the entire dataset, and once for each subject deleted. Knowing how computationally demanding nonlinear mixed models are, this is a major disadvantage. For the local influence approach, instead, obtaining the values of the influence measure is computationally inexpensive once the mixed model is fitted, as the model only needs to be fitted once. Also, this influence measure makes it possible to distinguish between the influence on the model's fixed-effects parameters and on the variance components.

After identifying influential individuals, each case should be carefully considered. We do not recommend simply removing these cases without further investigation, as they may well represent legitimate draws from the population under investigation and hence should remain in the dataset. However, once the influential cases have been identified, several strategies can be applied, including collecting additional data or additional measurements on existing cases, checking data consistency, adapting model specification or ultimately deleting the influential cases from the analysis if they represent inappropriate data (Verbeke and Molenberghs 2000; Van der Meer, Te Grotenhuis, and Pelzer 2010; Nieuwenhuis, Te Grotenhuis, and Pelzer 2012).

For the orange tree dataset, in which no trees were found influential on the parameter estimates of the model, we described in detail the entire process for obtaining the local influence diagnostic for the case-weight perturbation scheme, describing the derivatives analytically (see Appendix B). Also, the SAS code to generate the same results was described.

For the songbird dataset, the results of the local influence diagnostic were compared for two different nonlinear mixed models, showing that the method was very stable when changing the model. Another advantage is that using the `PROC NLMIXED` statement, the process of obtaining the results does not change for different models.

The methodology described here has been implemented in the SAS software system. The replication codes to generate the results presented in the paper are available in the supplementary materials.

Acknowledgments

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) under Grant 001.

References

Beckman RJ, Nachtsheim CJ, Cook RD (1987). "Diagnostics for Mixed-Model Analysis of Variance." *Technometrics*, **29**(4), 413–426. doi:10.2307/1269452.

Cook RD (1977). "Detection of Influential Observation in Linear Regression." *Technometrics*, **19**(1), 15–18. doi:10.1080/00401706.1977.10489493.

Cook RD (1986). "Assessment of Local Influence." *Journal of the Royal Statistical Society B*, **48**(2), 133–169. doi:10.1111/j.2517-6161.1986.tb01398.x.

Davidian M, Giltinan DM (2003). "Nonlinear Models for Repeated Measurement Data: An Overview and Update." *Journal of Agricultural, Biological, and Environmental Statistics*, **8**(4), 387–419. doi:10.1198/1085711032697.

Draper NR, Smith H (1981). *Applied Regression Analysis*. 2nd edition. John Wiley & Sons, New York.

Lesaffre E, Verbeke G (1998). "Local Influence in Linear Mixed Models." *Biometrics*, **54**(2), 570–582. doi:10.2307/3109764.

Lindstrom MJ, Bates DM (1990). "Nonlinear Mixed Effects Models for Repeated Measures Data." *Biometrics*, **46**(3), 673–687. doi:10.2307/2532087.

Molenberghs G, Verbeke G (2005). *Models for Discrete Longitudinal Data*. 1st edition. Springer-Verlag, New York. doi:10.1007/0-387-28980-1.

Nieuwenhuis R, Te Grotenhuis M, Pelzer B (2012). "**influence.ME**: Tools for Detecting Influential Data in Mixed Effects Models." *The R Journal*, **4**(2), 38–47. doi:10.32614/rj-2012-011.

Pinheiro JC, Bates DM (1995). "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model." *Journal of Computational and Graphical Statistics*, **4**(1), 12–35. doi:10.2307/1390625.

Rakhmawati TW, Molenberghs G, Verbeke G, Faes C (2017). "Local Influence Diagnostics for Hierarchical Finite-Mixture Random-Effects Models." *Biometrical Journal*, pp. 1–12. doi:10.1002/bimj.201600203.

SAS Institute Inc (2023). *SAS/STAT 9.4 Software*. Cary. URL https://www.sas.com/.

Serroyen J, Molenberghs G, Verhoye M, Van Meir V, Van der Linden A (2005). "Dynamic Manganese-Enhanced MRI Signal Intensity Processing Based on Nonlinear Mixed Modeling to Study Changes in Neuronal Activity." *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 170–183. doi:10.1198/108571105x46426.

Van der Linden A, Verhoye M, Van Meir V, Tindemans I, Eens M, Absil P, Balthazart J (2002). "In Vivo Manganese-Enhanced Magnetic Resonance Imaging Reveals Connections and Functional Properties of the Songbird Vocal Control System." *Neuroscience*, **112**(2), 467–474. doi:10.1016/s0306-4522(02)00070-2.

Van der Meer T, Te Grotenhuis M, Pelzer B (2010). "Influential Cases in Multilevel Modeling: A Methodological Comment." *American Sociological Review*, **75**(1), 173–178. doi:10.1177/0003122409359166.

Van Meir V, Verhoye M, Absil P, Eens M, Balthazart J, Van der Linden A (2004). "Differential Effects of Testosterone on Neuronal Populations and Their Connections in a Sensorimotor Brain Nucleus Controlling Song Production in Songbirds: A Manganese Enhanced-Magnetic Resonance Imaging Study." *NeuroImage*, **21**(3), 914–923. doi:10.1016/j.neuroimage.2003.10.007.

Verbeke G, Molenberghs G (eds.) (1997). *Linear Mixed Models in Practice: A SAS-Oriented Approach*. 1st edition. Springer-Verlag, New York. doi:10.1007/978-1-4612-2294-1.

Verbeke G, Molenberghs G (2000). *Linear Mixed Models for Longitudinal Data*. 1st edition. Springer-Verlag, New York. doi:10.1007/978-1-4419-0300-6.

A. Individual log-likelihood $\ell_i(\boldsymbol{\theta})$ for orange tree dataset

The given model is nonlinear in the fixed-effect parameters, but linear in the random effect b_i, simplifying the calculation of the marginal mean over the random-effects distribution, so the conditional mean is

$$E(Y_{ij} \mid b_i) = \frac{\beta_1 + b_i}{1 + \exp\left[-(t_{ij} - \beta_2)/\beta_3\right]}.$$

Defining $\boldsymbol{\mu}_i = \boldsymbol{\eta}_i + \boldsymbol{\lambda}_i b_i$, where

$$\boldsymbol{\eta}_i = \begin{bmatrix} \frac{\beta_1}{1+\exp\left[-(t_{i1}-\beta_2)/\beta_3\right]} \\ \cdots \\ \frac{\beta_1}{1+\exp\left[-(t_{in_i}-\beta_2)/\beta_3\right]} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\lambda}_i = \begin{bmatrix} \frac{1}{1+\exp\left[-(t_{i1}-\beta_2)/\beta_3\right]} \\ \cdots \\ \frac{1}{1+\exp\left[-(t_{in_i}-\beta_2)/\beta_3\right]} \end{bmatrix},$$

and using (1), the marginal density of $\boldsymbol{Y}_i$ is given by

$$\begin{aligned} f_i(\boldsymbol{y}_i \mid \beta_1, \beta_2, \beta_3, d, \sigma^2) &= \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}\left[(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)\right]}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}} \frac{\mathrm{e}^{-\frac{1}{2}\frac{b_i^2}{d}}}{\sqrt{2\pi d}} \, db_i \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}\left[(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)+d^{-1}b_i^2\right]}}{(2\pi)^{(n_i+1)/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \, db_i. \end{aligned} \tag{10}$$

The exponent of (10) can be rewritten as

$$(\boldsymbol{Y}_i - \boldsymbol{\eta}_i - \boldsymbol{\lambda}_i b_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\eta}_i - \boldsymbol{\lambda}_i b_i) + d^{-1}b_i^2. \tag{11}$$

Let $\boldsymbol{\gamma}_i = \boldsymbol{Y}_i - \boldsymbol{\eta}_i$, then (11) became

$$\begin{aligned} &= (\boldsymbol{\gamma}_i - \boldsymbol{\lambda}_i b_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\gamma}_i - \boldsymbol{\lambda}_i b_i) + d^{-1}b_i^2 \\ &= \boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i - 2\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i b_i + \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i b_i^2 + d^{-1}b_i^2 \\ &= (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})b_i^2 - 2\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i b_i + \boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i. \end{aligned} \tag{12}$$

Let $V_i^{-1} = \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1}$, we can now write (12) as

$$\begin{aligned} V_i^{-1}b_i^2 - 2\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i b_i + \boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i &= (\theta_i - b_i)^\top V_i^{-1}(\theta_i - b_i) + \zeta \\ V_i^{-1}b_i^2 - 2\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i b_i + \boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i &= \theta_i^2 V_i^{-1} - 2\theta_i V_i^{-1} b_i + V_i^{-1} b_i^2 + \zeta. \end{aligned}$$

Then

$$\begin{aligned} 2\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i b_i &= 2\theta_i V_i^{-1} b_i \\ \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i (V_i) &= \theta_i V_i^{-1}(V_i) \\ \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i V_i &= \theta_i \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i &= \theta_i^2 V_i^{-1} + \zeta \\ \boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i &= (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2 (V_i)^2 V_i^{-1} + \zeta \\ \boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2 V_i &= \zeta. \end{aligned}$$

Using this results, we can rewrite the exponent of (10) as $(\theta_i - b_i)^\top V_i^{-1}(\theta_i - b_i) + \zeta$, resulting in

$$f_i(\boldsymbol{y}_i \mid \beta_1, \beta_2, \beta_3, d, \sigma^2) = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}\left[(\theta_i - b_i)^\top V_i^{-1}(\theta_i - b_i) + \zeta\right]}}{(2\pi)^{(n_i+1)/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \; db_i$$

$$= \frac{\mathrm{e}^{-\frac{1}{2}\zeta}}{(2\pi)^{(n_i+1)/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\left[(\theta_i - b_i)^\top V_i^{-1}(\theta_i - b_i)\right]} \; db_i \qquad (13)$$

because ζ does not depend on b_i. Multiplying and dividing (13) by $(2\pi)^{1/2}V_i^{1/2}$, we obtain the density function of a normal distribution

$$f_i(\boldsymbol{y}_i \mid \boldsymbol{\beta}, d, \sigma^2) = \frac{\mathrm{e}^{-\frac{1}{2}\zeta}V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi V_i}} \mathrm{e}^{-\frac{1}{2}\left[(\theta_i - b_i)^\top V_i^{-1}(\theta_i - b_i)\right]} \; db_i$$

$$= \frac{V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \mathrm{e}^{-\frac{1}{2}\left[\boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2 V_i\right]}.$$

And the individual log-likelihood is given by

$$\ell_i(\boldsymbol{\theta}) = \log\left\{ \frac{V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \mathrm{e}^{-\frac{1}{2}\left[\boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2 V_i\right]} \right\},$$

where $V_i = [\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1}]^{-1}$ and $\boldsymbol{\gamma}_i = \boldsymbol{Y}_i - \boldsymbol{\eta}_i$.

B. Derivatives of $\ell_i(\boldsymbol{\theta})$ for the orange tree dataset

B.1. Derivative with respect to β_1

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \; \log\left\{ \frac{V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \mathrm{e}^{-\frac{1}{2}\left[\boldsymbol{\gamma}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2 V_i\right]} \right\}.$$

For convenience, let us call $c = \frac{V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}}$, and knowing that $\boldsymbol{\gamma}_i = \boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i$, we now have

$$\begin{aligned}
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_1} &= \frac{\partial}{\partial \beta_1} \; \log\left(c\ \mathrm{e}^{-\frac{1}{2}\left\{(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i) - [\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i)]^2 V_i\right\}} \right) \\
&= \frac{\frac{\partial}{\partial \beta_1}\left(c\ \mathrm{e}^{-\frac{1}{2}\left\{(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i) - [\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i)]^2 V_i\right\}} \right)}{c\ \mathrm{e}^{-\frac{1}{2}\left\{(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i) - [\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \beta_1\boldsymbol{\lambda}_i)]^2 V_i\right\}}} \\
&= \frac{\partial}{\partial \beta_1}\left\{ -\frac{1}{2}\left[\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i - 2\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i\beta_1 + \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i\beta_1^2 \right.\right. \\
&\quad \left.\left. -(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i - \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i\beta_1)^2 V_i \right]\right\} \\
&= \frac{\partial}{\partial \beta_1}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i\beta_1 \right] - \frac{\partial}{\partial \beta_1}\left[\frac{\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i\beta_1^2}{2} \right. \\
&\quad \left. + V_i(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i - \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i\beta_1) \left[\frac{\partial}{\partial \beta_1}\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i - \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i\beta_1 \right) \right] \right] \\
&= \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i - \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i\beta_1 - V_i(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i - \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i\beta_1)(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i)
\end{aligned}$$

which can be simplified as

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_1} = [(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i)V_i - 1]\left[(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i)\beta_1 - \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i\right].$$

B.2. Derivative with respect to β_2

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_2} = \frac{\partial}{\partial \beta_2} \quad \log\left\{\frac{V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}} \mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2 V_i\right]}\right\}.$$

For convenience, let us call $c = (2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}$ and $\boldsymbol{\Sigma}^{-1} = \sigma^{-2}\boldsymbol{I} = s^{-1}\boldsymbol{I}$, and knowing that $V_i = (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{-1}$, then

$$\begin{aligned}
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_2} &= \frac{\partial}{\partial \beta_2} \quad \log\left(\frac{\mathrm{e}^{-\frac{1}{2}\left\{\gamma_i^\top (s^{-1}\boldsymbol{I})\gamma_i - (\boldsymbol{\lambda}_i^\top (s^{-1}\boldsymbol{I})\gamma_i)^2\left[\boldsymbol{\lambda}_i^\top (s^{-1}\boldsymbol{I})\boldsymbol{\lambda}_i + d^{-1}\right]^{-1}\right\}}}{c\left[\boldsymbol{\lambda}_i^\top (s^{-1}\boldsymbol{I})\boldsymbol{\lambda}_i + d^{-1}\right]^{1/2}}\right) \\
&= \frac{\partial}{\partial \beta_2} \quad \log\left\{\frac{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top\gamma_i s^{-1} - (\boldsymbol{\lambda}_i^\top\gamma_i)^2 s^{-2}\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^{-1}\right]}}{c\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^{1/2}}\right\} \\
&= \frac{c\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^{1/2}}{\mathrm{e}^{-\frac{1}{2}\left[\frac{\gamma_i^\top\gamma_i}{s} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right]}} \times \frac{\partial}{\partial \beta_2}\left\{\frac{\mathrm{e}^{-\frac{1}{2}\left[\frac{\gamma_i^\top\gamma_i}{s} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right]}}{c\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^{1/2}}\right\} \\
&= \frac{\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^{1/2}}{\mathrm{e}^{-\frac{1}{2}\left[\frac{\gamma_i^\top\gamma_i}{s} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right]}}\left(\frac{\frac{\partial}{\partial \beta_2}\left\{\mathrm{e}^{-\frac{1}{2}\left[\frac{\gamma_i^\top\gamma_i}{s} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right]}\right\}}{\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^{1/2}}\right. \\
&\qquad \left. - \frac{\mathrm{e}^{-\frac{1}{2}\left[\frac{\gamma_i^\top\gamma_i}{s} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right]}\frac{\partial}{\partial \beta_2}\left[\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^{1/2}\right]}{\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right) \\
&= -\frac{1}{2}\left\{\frac{\partial}{\partial \beta_2}\left[\frac{\boldsymbol{\gamma}_i^\top\boldsymbol{\gamma}_i}{s} - \frac{(\boldsymbol{\lambda}_i^\top\boldsymbol{\gamma}_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right] + \frac{\frac{\partial}{\partial \beta_2}\left[\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)\right]}{\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right\} \\
&= -\frac{1}{2}\left\{\frac{\partial}{\partial \beta_2}\left[\frac{\boldsymbol{\gamma}_i^\top\boldsymbol{\gamma}_i}{s}\right] - \frac{\partial}{\partial \beta_2}\left[\frac{(\boldsymbol{\lambda}_i^\top\boldsymbol{\gamma}_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right] + \frac{\frac{\partial}{\partial \beta_2}\left[\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)\right]}{\left(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right\}.
\end{aligned} \tag{14}$$

Solving each one of the derivatives in (14), first we have

$$\frac{\partial}{\partial \beta_2}\left[\frac{\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i}{s}\right] = \frac{1}{s}\ \frac{\partial}{\partial \beta_2}\left[\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i\right],$$

but $\boldsymbol{\gamma}_i = \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i$, so

$$\begin{aligned}
\frac{1}{s}\ \frac{\partial}{\partial \beta_2}\left[\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i\right] &= \frac{1}{s}\frac{\partial}{\partial \beta_2}\left[(\boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i)^\top (\boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i)\right] \\
&= \frac{1}{s}\frac{\partial}{\partial \beta_2}\left[\boldsymbol{Y}_i^\top \boldsymbol{Y}_i - 2\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i \beta_1 + \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i \beta_1^2\right] \\
&= \frac{-2\beta_1}{s}\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i\right] + \frac{\beta_1^2}{s}\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right],
\end{aligned}$$

Also, in (14), we have

$$\begin{aligned}
&\frac{\partial}{\partial \beta_2}\left[\frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right] = \frac{1}{s^2}\ \frac{\partial}{\partial \beta_2}\left[\frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right] \\
&= \frac{1}{s^2}\frac{\frac{\partial}{\partial \beta_2}\left[(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2\right]\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right) - \frac{\partial}{\partial \beta_2}\left[\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)\right](\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^2} \\
&= \frac{\frac{\partial}{\partial \beta_2}\left\{\left[\boldsymbol{\lambda}_i^\top (\boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i)\right]^2\right\}}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)} - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2 \frac{\partial}{\partial \beta_2}\left[\frac{\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i}{s}\right]}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^2} \\
&= \frac{\frac{\partial}{\partial \beta_2}\left[(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)^2\right]}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)} - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2 \frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right]}{s^3\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^2} \\
&= \frac{2(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right]}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)} - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2 \frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right]}{s^3\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^2}.
\end{aligned}$$

The last part in (14) is

$$\frac{\frac{\partial}{\partial \beta_2}\left[\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)\right]}{\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)} = \frac{\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right]}{s\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}.$$

So, returning to (14), we have that

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_2} = -\frac{1}{2}\left\{ \frac{\partial}{\partial \beta_2}\left[\frac{\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i}{s}\right] - \frac{\partial}{\partial \beta_2}\left[\frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right] + \frac{\frac{\partial}{\partial \beta_2}\left[\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)\right]}{\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)} \right\}$$

$$= -\frac{1}{2}\left\{ \frac{-2\beta_1}{s}\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i\right] + \frac{\beta_1^2}{s}\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right] - \frac{2(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i\right]}{s^2(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1})} \right.$$

$$\left. + \frac{2\beta_1(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right]}{s^2(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1})} + \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right]}{s^3\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^2} + \frac{\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right]}{s\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)} \right\}$$

$$= \frac{\partial}{\partial \beta_2}\left(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i\right)\left[\frac{\beta_1}{s} + \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right] - \frac{\partial}{\partial \beta_2}\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right)$$

$$\times \left[\frac{\beta_1^2}{2s} + \frac{\beta_1(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)}{s^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)} + \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{2s^3\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)^2} + \frac{1}{2s\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1} + d^{-1}\right)}\right],$$

or, knowing that $s^{-1} = \sigma^{-2}$,

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\left(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i\right)\left[\frac{\beta_1}{\sigma^2} + \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)}{\sigma^4\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i \sigma^{-2} + d^{-1}\right)}\right] - \frac{\partial}{\partial \beta_2}\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right) \times$$

$$\left[\frac{\beta_1^2}{2\sigma^2} + \frac{\beta_1(\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i - \beta_1 \boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)}{\sigma^4\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i \sigma^{-2} + d^{-1}\right)} + \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{2\sigma^6\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i \sigma^{-2} + d^{-1}\right)^2} + \frac{1}{2\sigma^2\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i \sigma^{-2} + d^{-1}\right)}\right] \quad (15)$$

where

$$\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i\right] = \frac{\partial}{\partial \beta_2}\left[\sum_{j=1}^{n_i} \frac{Y_{ij}}{1 + \mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}}\right] = \sum_{j=1}^{n_i} \frac{-Y_{ij}\mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}}{\beta_3\left[1 + \mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}\right]^2} \quad (16)$$

$$\frac{\partial}{\partial \beta_2}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right] = \frac{\partial}{\partial \beta_2}\left[\sum_{j=1}^{n_i} \frac{1}{\left[1 + \mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}\right]^2}\right] = \sum_{j=1}^{n_i} \frac{-2\mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}}{\beta_3\left[1 + \mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}\right]^3}. \quad (17)$$

B.3. Derivative with respect to β_3

The calculation of the derivative of the individual log-likelihood with respect to β_3 follows the same logic as Section B.2, with the exception of the results in (16) and (17), where:

$$\frac{\partial}{\partial \beta_3}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{Y}_i\right] = \frac{\partial}{\partial \beta_3}\left[\sum_{j=1}^{n_i} \frac{Y_{ij}}{1 + \mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}}\right] = \sum_{j=1}^{n_i} \frac{-Y_{ij}(t_{ij} - \beta_2)\mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}}{\beta_3^2\left[1 + \mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}\right]^2}.$$

$$\frac{\partial}{\partial \beta_3}\left[\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i\right] = \frac{\partial}{\partial \beta_3}\left[\sum_{j=1}^{n_i} \frac{1}{\left[1+\mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}\right]^2}\right] = \sum_{j=1}^{n_i} \frac{-2(t_{ij}-\beta_2)\mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}}{\beta_3^2\left[1+\mathrm{e}^{\frac{-(t_{ij}-\beta_2)}{\beta_3}}\right]^3}.$$

B.4. Derivative with respect to σ^2

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2} \quad \log\left\{\frac{V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}d^{1/2}}\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2 V_i\right]}\right\}.$$

For convenience, let us call $c = (2\pi)^{n_i/2}d^{1/2}$ and $\sigma^2 = s$. We know that $V_i = (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{-1}$ and $\boldsymbol{\Sigma} = s\boldsymbol{I}$, so $|\boldsymbol{\Sigma}| = s^{n_i}$ and $\boldsymbol{\Sigma}^{-1} = s^{-1}\boldsymbol{I}$. Then

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial s} = \frac{\partial}{\partial s} \quad \log\left\{\frac{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top\gamma_i s^{-1} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i s^{-1})^2}{(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right]}}{cs^{n_i/2}(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}}\right\}$$

$$= \frac{cs^{n_i/2}(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}}{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top\gamma_i s^{-1} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i s^{-1})^2}{(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right]}} \times \frac{\partial}{\partial s}\left\{\frac{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top\gamma_i s^{-1} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i s^{-1})^2}{(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right]}}{cs^{n_i/2}(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}}\right\}$$

$$= \frac{s^{n_i/2}(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}}{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top\gamma_i s^{-1} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i s^{-1})^2}{(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right]}}\left(\frac{\frac{\partial}{\partial s}\left\{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top\gamma_i s^{-1} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i s^{-1})^2}{(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right]}\right\}}{s^{n_i/2}(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}}\right.$$

$$\left.-\frac{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top\gamma_i s^{-1} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i s^{-1})^2}{(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right]}\frac{\partial}{\partial s}\left[s^{n_i/2}(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}\right]}{s^{n_i}(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right)$$

$$= -\frac{\gamma_i^\top\gamma_i}{2}\frac{\partial}{\partial s}\left[\frac{1}{s}\right] + \frac{(\boldsymbol{\lambda}_i^\top\gamma_i)^2}{2}\frac{\partial}{\partial s}\left[\frac{1}{s^2(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right]$$

$$-\frac{\frac{\partial}{\partial s}\left[(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}\right]}{(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^{1/2}} - \frac{\frac{\partial}{\partial s}\left[s^{n_i/2}\right]}{s^{n_i/2}}$$

$$= \frac{\gamma_i^\top\gamma_i}{2s^2} - \frac{(\boldsymbol{\lambda}_i^\top\gamma_i)^2}{2}\left[\frac{s^2\frac{\partial}{\partial s}\left[\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1}\right] + (\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})\frac{\partial}{\partial s}\left[s^2\right]}{s^4(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})^2}\right]$$

$$-\frac{\frac{\partial}{\partial s}\left[(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})\right]}{2(\boldsymbol{\lambda}_i^\top\boldsymbol{\lambda}_i s^{-1}+d^{-1})} - \frac{n_i}{2s}$$

$$
\begin{aligned}
&= \frac{\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i}{2s^2} - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{2}\left[-\frac{\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i}{s^4(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1}+d^{-1})^2} + \frac{2}{s^3(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1}+d^{-1})}\right] \\
&\quad + \frac{\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i}{2s^2(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1}+d^{-1})} - \frac{n_i}{2s} \\
&= \frac{\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i - n_i s}{2s^2} + \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)}{2s^4(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1}+d^{-1})^2} + \frac{s(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i) - 2(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{2s^3(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i s^{-1}+d^{-1})},
\end{aligned}
$$

or, considering that $s = \sigma^2$,

$$
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial(\sigma^2)} = \frac{\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i - n_i\sigma^2}{2\sigma^4} + \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i)}{2\sigma^8(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i \sigma^{-2}+d^{-1})^2} + \frac{\sigma^2(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i) - 2(\boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_i)^2}{2\sigma^6(\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i \sigma^{-2}+d^{-1})}.
$$

B.5. Derivative with respect to d

$$
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial d} = \frac{\partial}{\partial d} \quad \log\left\{\frac{V_i^{1/2}}{(2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2} d^{1/2}} \mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2 V_i\right]}\right\}.
$$

For convenience, let us call $c = (2\pi)^{n_i/2}|\boldsymbol{\Sigma}|^{1/2}$. We know that $V_i = (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{-1}$, then

$$
\begin{aligned}
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial d} &= \frac{\partial}{\partial d} \quad \log\left\{\frac{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right]}}{c\ d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}}\right\} \\
&= \frac{c\ d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}}{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right]}} \times \frac{\partial}{\partial d}\left\{\frac{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right]}}{c\ d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}}\right\} \\
&= \frac{d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}}{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right]}} \left(\frac{\frac{\partial}{\partial d}\left\{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right]}\right\}}{d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}}\right. \\
&\quad \left. - \frac{\mathrm{e}^{-\frac{1}{2}\left[\gamma_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i - \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\gamma_i)^2}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right]} \frac{\partial}{\partial d}\left[d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}\right]}{d(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2}\frac{\partial}{\partial d}\left[\frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}\right] - \frac{\frac{\partial}{\partial d}\left[d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}\right]}{d^{1/2}\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}} \\
&= -\frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2}{2}\left[\frac{\frac{\partial}{\partial d}\left(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1}\right)}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^2}\right] - \frac{\frac{\partial}{\partial d}\left(d^{1/2}\right)}{d^{1/2}} \\
&\quad - \frac{\frac{\partial}{\partial d}\left[(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}\right]}{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^{1/2}} \\
&= \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2}{2\ d^2\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^2} - \frac{1}{2d} + \frac{1}{2\ d^2\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})} \\
&= \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2 - d\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^2 + (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})}{2\ d^2\ (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i + d^{-1})^2},
\end{aligned}
$$

which can be simplified as

$$
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial d} = \frac{(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i)^2 - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i^2)d - (\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i)}{2(\boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_i d + 1)^2}.
$$

Affiliation:

Jhessica Leticia Kirch
Department of Exact Sciences
Luiz de Queiroz College of Agriculture
University of São Paulo
Av. Pádua Dias, 11
Piracicaba, Brazil
E-mail: jhessicakirch@usp.br

Journal of Statistical Software https://www.jstatsoft.org/
published by the Foundation for Open Access Statistics https://www.foastat.org/
December 2025, Volume 115, Code Snippet 1 *Submitted:* 2024-04-12
doi:10.18637/jss.v115.c01 *Accepted:* 2025-03-10