Stat Papers
DOI 10.1007/s00362-017-0921-9

REGULAR ARTICLE

Elliptical linear mixed models with a covariate subject to measurement error

Joelmir A. Borssoi[1] · **Gilberto A. Paula**[2] · **Manuel Galea**[3]

Received: 3 April 2016 / Revised: 1 June 2017

Abstract In this paper we extend linear mixed models with elliptical errors by adding a covariate subject to measurement error in the linear predictor. The former class is defined appropriately so that the joint marginal distribution of the response and the observed covariate subject to measurement error is also elliptical. Thus, numerical integration methods are not required to obtain the marginal model and the mean and the variance–covariance structures of the hierarchical model are preserved. A kurtosis flexibility is allowed for each joint marginal distribution and since the conditional distributions are also elliptical, the predictions of the random effects as well as of the covariate subject to measurement error may be performed in a similar way of the normal case. A reweighed iterative process based on the maximum likelihood method is derived for obtaining the parameter estimates, which appear to be robust against outlying observations in the sense of the Mahalanobis distance. In order to assess the sample properties of the maximum likelihood parameter estimates as well as their asymptotic standard errors, a simulation study is performed under different parameter settings and error distributions. Goodness-of-fit measures based on the Mahalanobis distance are presented and normal curvatures of local influence are derived under three usual perturbation schemes, which are selected appropriately. Finally, an illustrative

---

✉ Gilberto A. Paula
giapaula@ime.usp.br

Joelmir A. Borssoi
jborssoi@utfpr.edu.br

Manuel Galea
mgalea@mat.puc.cl

[1] Universidade Tecnológica Federal do Paraná, Londrina, Brazil

[2] Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

[3] Pontificia Universidad Católica de Chile, Santiago, Chile

example previously analyzed under normal error models is reanalyzed by considering heavy-tailed error models. The diagnostic procedures are applied for comparing the fitted models.

Keywords Linear mixed models · Local influence · Measurement error models · Robust estimates · Student-t models · Structural models

1 Introduction

Linear mixed models with elliptical errors have become an interesting approach for analyzing data sets with repeated measurements, due to the easy of incorporation of the framework proposed by Laird and Ware (1982). The similarity between the properties of the elliptical class and the normal one, as well as the kurtosis flexibility, may be considered additional attractive to consider elliptical linear models. Some authors have studied elliptical models in the multivariate and repeated measurement contexts. For instance, Lange et al. (1989) presented an approach for modeling multivariate t-distributions with known and unknown degrees of freedom, whereas Welsh and Richardson (1997) described multivariate t-linear mixed models using marginal distributions, Lindsey (1999) discussed the application of the power exponential models in repeated measurement problems and Pinheiro et al. (2001) proposed a robust hierarchical linear mixed model in which the random effects and the within-subject errors follow a multivariate t-distribution. Savalli et al. (2006) introduced the elliptical linear mixed models with emphasis on the assessment of variance components, whereas Staudenmayer et al. (2009) discussed Student-t mixed models under the semiparametric framework. Wu (2010) also discussed mixed-effects models, under the multivariate t-distribution and more recently Lemonte and Patriota (2011) presented a multivariate elliptical model with a general parametrization.

We extend in this paper the structure proposed by Savalli et al. (2006), by adding a covariate subject to measurement error in the linear predictor. From this structure the joint marginal distribution of the response and the observed covariate subject to measurement error is also elliptical. Thus, numerical integration methods are not required to obtain the marginal model, the mean and the variance–covariance structures of the hierarchical model are preserved in the marginal model and a specific value for the kurtosis parameter may be defined for each marginal distribution. In addition, the predictions of the random effects and the covariate subject to measurement error may be performed similarly to the normal case, for instance, by using the empirical Bayes method. Other attractive to consider elliptical linear models is the robustness of the maximum likelihood estimates against outlying observations under heavy-tailed error models.

Some authors have extended the approach proposed by Savalli et al. (2006) for different situations. For instance, Osorio et al. (2007) derived the normal curvatures of local influence under usual perturbation schemes, Russo et al. (2009) defined the elliptical partially nonlinear mixed models by adding a nonlinear function into the linear predictor, whereas Ibacache-Pulgar et al. (2012) added a nonparametric function in the linear predictor for modeling the profiles in longitudinal studies. On the other hand, linear mixed models with measurement error have been well studied under normal errors. For instance, Zhong et al. (2002) proposed a unified method for the parameter estimation based on the corrected score function of Nakamura (1990), whereas Cui et al. (2004) discussed consistency and asymptotic normality in linear mixed-effects models with measurement error in both fixed and random effects. For a discussion of mixed-effects models with measurement error under the assumption of normality see, for instance, Wu (2010) and Buonaccorsi (2010). Recently, Riquelme et al. (2015) extended the functional mixed model proposed by Zhong et al. (2002) by assuming that each measurement error follows an elliptical distribution. A corrected score approach is applied for the parameter estimation; consistency and asymptotic normality are also discussed. Unlike the previous work, we propose in this paper an elliptical structural mixed model, in which the random effects, the unobserved covariate, the measurement errors and the random errors follow jointly a multivariate elliptical distribution.

The work is presented as follows. In Sect. 2 we introduce the class of elliptical linear mixed models with a covariate subject to measurement error. A reweighed iterative process for obtaining the maximum likelihood estimates is derived in Sect. 3 and a methodology for the predictions of the random effects and the covariate subject to measurement error is proposed in Sect. 4. A simulation study to assess the sample properties of the maximum likelihood estimates and their asymptotic standard errors, performed under different parameter settings and error distributions, is described in Sect. 5. Goodness-of-fit measures are presented in Sect. 6 as well as normal curvatures of local influence are derived from the marginal model under three usual perturbation schemes, which are selected appropriately. In Sect. 7 the housing price data, previously analyzed by Longford (1993) and Zhong et al. (2002) under linear mixed models with normal errors, is reanalyzed under heavy-tailed error models. Diagnostic procedures are applied for comparing the fitted models. The final section presents some concluding remarks whereas in the appendices we present the derivations of the score functions, Hessian and Fisher information matrices as well as the perturbation matrices from the local influence methodology.

2 The model

Following a similar notation of that given in Savalli et al. (2006), we will define the elliptical linear mixed models with a covariate subject to measurement error as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \gamma\boldsymbol{u}_i^* + \boldsymbol{\epsilon}_i \tag{1}$$

and

$$\boldsymbol{u}_i = \boldsymbol{u}_i^* + \boldsymbol{e}_i, \tag{2}$$

for $i = 1, \ldots, n$, where $\boldsymbol{y}_i$ is an m_i-dimensional vector of observed responses from the ith cluster, $\boldsymbol{X}_i$ is an $m_i \times p$ matrix which contains the explanatory variable values, $\boldsymbol{\beta}$ is the fixed parameter vector, $\boldsymbol{Z}_i$ is an $m_i \times q$ design matrix of the random effects $\boldsymbol{b}_i$, $\boldsymbol{u}_i^*$ is the m_i-dimensional covariate subject to measurement error with $\boldsymbol{u}_i$ denoting its observed values, γ is a fixed parameter value whereas $\boldsymbol{\epsilon}_i$ and $\boldsymbol{e}_i$ denote m_i-dimensional errors.

We will assume that the random effects $\boldsymbol{b}_i$, the covariate subject to measurement error $\boldsymbol{u}_i^*$ and the model errors $\boldsymbol{\epsilon}_i$ and $\boldsymbol{e}_i$ are not correlated having the following joint distribution:

$$\begin{pmatrix} \boldsymbol{b}_i \\ \boldsymbol{u}_i^* \\ \boldsymbol{\epsilon}_i \\ \boldsymbol{e}_i \end{pmatrix} \overset{\text{ind}}{\sim} \text{El}_{3m_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\mu}_i^* \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_u & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma^2 \boldsymbol{I}_{m_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_e^2 \boldsymbol{I}_{m_i} \end{bmatrix}\right), \tag{3}$$

where $\text{El}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a r-dimensional elliptical distribution with position parameter $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$ (see, for instance, Fang et al. 1990).

Let the observed vectors $\boldsymbol{y}_i$ and $\boldsymbol{u}_i$ jointly expressed as $\boldsymbol{W}_i = [\boldsymbol{y}_i^\top, \boldsymbol{u}_i^\top]^\top$, for $i = 1, \ldots, n$. From Eqs. (1)–(3) we may express

$$\boldsymbol{W}_i = \begin{pmatrix} \boldsymbol{X}_i \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix} + \begin{bmatrix} \boldsymbol{Z}_i & \gamma \boldsymbol{I}_{m_i} & \boldsymbol{I}_{m_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{m_i} & \mathbf{0} & \boldsymbol{I}_{m_i} \end{bmatrix} \begin{pmatrix} \boldsymbol{b}_i \\ \boldsymbol{u}_i^* \\ \boldsymbol{\epsilon}_i \\ \boldsymbol{e}_i \end{pmatrix},$$

for $i = 1, \ldots, n$, that is, $\boldsymbol{W}_i$ is a linear combination of the latent vectors. Then, from statistical properties of the elliptical class, it follows that

$$\boldsymbol{W}_i \overset{\text{ind}}{\sim} \text{El}_{2m_i}\left(\begin{pmatrix} \boldsymbol{X}_i \boldsymbol{\beta} + \gamma \boldsymbol{\mu}_i^* \\ \boldsymbol{\mu}_i^* \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_i & \gamma \boldsymbol{\Sigma}_u \\ \gamma \boldsymbol{\Sigma}_u & \boldsymbol{\Sigma}_u + \sigma_e^2 \boldsymbol{I}_{m_i} \end{bmatrix}\right),$$

where $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^\top + \gamma^2 \boldsymbol{\Sigma}_u + \sigma^2 \boldsymbol{I}_{m_i}$, for $i = 1, \ldots, n$. Typically, the classical inferential analysis is based on this distribution.

In addition, for simplicity, we will assume that $\boldsymbol{\mu}_i^* = \mu \mathbf{1}_{m_i}$ and $\boldsymbol{\Sigma}_u = \sigma_u^2 \boldsymbol{I}_{m_i}$, where $\mathbf{1}_{m_i}$ denotes an $m_i \times 1$ vector of $1's$ and $\boldsymbol{I}_{m_i}$ is the identity matrix of order m_i. The variance–covariance matrix of the random effect vector $\boldsymbol{b}_i$ will be denoted by

$$\boldsymbol{D} = \begin{bmatrix} \tau_1 & \tau_{12} & \cdots & \tau_{1q} \\ \tau_{21} & \tau_2 & \cdots & \tau_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{q1} & \tau_{q2} & \ldots & \tau_q \end{bmatrix},$$

and we will denote the vector $\boldsymbol{\tau} = \text{vech}(\boldsymbol{D})$, which contains q variance parameters $(\tau_1, \ldots, \tau_q)^\top$ and $s = \frac{q(q-1)}{2}$ $(q \geq 2)$ covariance parameters $(\tau_{12}, \ldots, \tau_{(q-1)q})^\top$. Hence, the distribution of the observed data becomes given by

$$\boldsymbol{W}_i \overset{\text{ind}}{\sim} \text{El}_{2m_i}\left(\boldsymbol{\mu}_{iW}, \boldsymbol{V}_i\right), \tag{4}$$

where

$$\boldsymbol{\mu}_{iW} = \begin{pmatrix} \boldsymbol{X}_i \boldsymbol{\beta} + \gamma \mu \mathbf{1}_{m_i} \\ \mu \mathbf{1}_{m_i} \end{pmatrix}$$

and

$$\boldsymbol{V}_i = \begin{bmatrix} \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^\top + (\gamma^2\sigma_u^2 + \sigma^2)\boldsymbol{I}_{m_i} & \gamma\sigma_u^2 \boldsymbol{I}_{m_i} \\ \gamma\sigma_u^2 \boldsymbol{I}_{m_i} & (\sigma_u^2 + \sigma_e^2)\boldsymbol{I}_{m_i} \end{bmatrix},$$

for $i = 1, \ldots, n$.

It can be seen that the parameters of this model are not identifiable. Consequently, restrictions must be placed on the parameters to identify them. In the statistical literature on measurement error models, see for instance (Fuller 2006), there are various suggestions on identifiability restrictions. In this work, for simplicity, we will assume that the scale parameter σ_e^2, associated with the measurement error $\boldsymbol{u}_i^*$, as well as β_0 are known. Thus, with these restrictions on the parameters, the model (4) will be identifiable. In effect, taking into account that the elliptical distributions are fully specified by the position vector and the scale matrix, we have that the conditions of identifiability for the elliptical model (4) are a natural extension of the normal model, when the kurtosis of the distribution is known. Then, using results from Demidenko (2004) and Wang and Heckman (2009), on identifiability in normal mixed models, it follows the identifiability of the elliptical linear mixed models with measurement error (4).

Therefore, we will have $p + q + s + 4$ parameters to be estimated, represented by the vector $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}^\top, \mu, \boldsymbol{\tau}^\top, \sigma_u^2, \sigma^2)^\top$. It follows from the distribution (4) that $\mathrm{Var}(y_{ij}) = \boldsymbol{z}_{ij}^\top \boldsymbol{D} \boldsymbol{z}_{ij} + \gamma^2\sigma_u^2 + \sigma^2$, whereas the intraclass correlation becomes

$$\rho_{jj'} = \mathrm{Corr}(y_{ij}, y_{ij'}) = \frac{\boldsymbol{z}_{ij}^\top \boldsymbol{D} \boldsymbol{z}_{ij}}{\sqrt{\boldsymbol{z}_{ij}^\top \boldsymbol{D} \boldsymbol{z}_{ij} + \gamma^2\sigma_u^2 + \sigma^2}\sqrt{\boldsymbol{z}_{ij'}^\top \boldsymbol{D} \boldsymbol{z}_{ij'} + \gamma^2\sigma_u^2 + \sigma^2}},$$

for $j \neq j'$, where $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijq})^\top$ denotes the jth row of the matrix $\boldsymbol{Z}_i$. One also has that $\mathrm{Cov}(y_{ij}, u_{ij}) = \gamma\sigma_u^2$, $\mathrm{Cov}(y_{ij}, u_{ij'}) = 0$ for $j \neq j'$, $\mathrm{Var}(u_{ij}) = \sigma_u^2 + \sigma_e^2$ and $\mathrm{Cov}(u_{ij}, u_{ij'}) = 0$ for $j \neq j'$. Hypotheses of interest are the assessment of $\mathrm{H}_0 : \boldsymbol{\tau} = \boldsymbol{0}$ against $\mathrm{H}_1 : \boldsymbol{\tau} \neq \boldsymbol{0}$ and $\mathrm{H}_0 : \gamma = 0$ against $\mathrm{H}_1 : \gamma \neq 0$.

3 Parameter estimation

3.1 Score function

The log-likelihood function for the parameter vector $\boldsymbol{\theta}$ becomes given by

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}),$$

where $L_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{V}_i| + \log\{g(\delta_i)\}$ with $\delta_i = \left(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}\right)^\top \boldsymbol{V}_i^{-1} \left(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}\right)$ being the square of the Mahalanobis distance for the ith observation, $i = 1, \ldots, n$, $g(\cdot)$ is a function from $I\!R : [0, \infty)$ such that $\int_0^\infty u^{\frac{m_i}{2}-1} g(u)du < \infty$. For the score function

$$\boldsymbol{U}(\boldsymbol{\theta}) = (U^{\gamma}(\boldsymbol{\theta}), \boldsymbol{U}^{\beta}(\boldsymbol{\theta})^{\top}, U^{\mu}(\boldsymbol{\theta}), \boldsymbol{U}^{\eta}(\boldsymbol{\theta})^{\top})^{\top} = \sum_{i=1}^{n} \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

we obtain the derivatives

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \gamma} = -\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) + \frac{1}{2}v(\delta_i)\left\{2\boldsymbol{r}_i^{\top}\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^{\top}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right\}, \tag{5}$$

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = v(\delta_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^{\top}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \tag{6}$$

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \mu} = v(\delta_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^{\top}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \text{ and} \tag{7}$$

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \eta_k} = -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \eta_k}\right) + \frac{1}{2}v(\delta_i)\boldsymbol{r}_i^{\top}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \eta_k}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \tag{8}$$

where $\boldsymbol{r}_i = \boldsymbol{W}_i - \boldsymbol{\mu}_{iW}$, η_k is the kth element of the $(q + r + 2) \times 1$ vector $\boldsymbol{\eta} = (\boldsymbol{\tau}^{\top}, \sigma_u^2, \sigma^2)^{\top}$ with the variance–covariance parameters, $v(\delta_i) = -2W_g(\delta_i)$ with $W_g(\delta_i) = \frac{\mathrm{d}}{\mathrm{d}\delta_i}\log\{g(\delta_i)\} = \frac{g'(\delta_i)}{g(\delta_i)}$. The quantity $v(\delta_i)$ that appears in the ith components of the score functions (5)–(8) may be interpreted as a weight, and since in general $g(\delta_i)$ is an increasing function of δ_i, then $v(\delta_i) > 0$. For some heavy-tailed error distributions, such as Student-t with ν degrees of freedom, $v(\delta_i)$ is inversely proportional to the Mahalanobis distance, so that outlying observations in the sense of this distance should receive smaller weights from the estimating equations.

3.2 Fisher information matrix

In order to obtain the Fisher information matrix $\boldsymbol{F}(\boldsymbol{\theta}) = \mathrm{E}\{\boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{U}(\boldsymbol{\theta})^{\top}\}$ for the parameter vector $\boldsymbol{\theta}$ we derive in Appendix A the submatrices of $\boldsymbol{F}(\boldsymbol{\theta})$ according with the subvectors of $\boldsymbol{\theta}$, obtaining

$$\boldsymbol{F}(\boldsymbol{\theta}) = \begin{bmatrix} F_{\gamma\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\gamma\beta}(\boldsymbol{\theta}) & F_{\gamma\mu}(\boldsymbol{\theta}) & F_{\gamma\sigma_u^2}(\boldsymbol{\theta}) & F_{\gamma\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\gamma\tau}(\boldsymbol{\theta}) \\ \boldsymbol{F}_{\beta\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\beta\beta}(\boldsymbol{\theta}) & \boldsymbol{F}_{\beta\mu}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ F_{\mu\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\mu\beta}(\boldsymbol{\theta}) & F_{\mu\mu}(\boldsymbol{\theta}) & 0 & 0 & \mathbf{0} \\ F_{\sigma_u^2\gamma}(\boldsymbol{\theta}) & \mathbf{0} & 0 & F_{\sigma_u^2\sigma_u^2}(\boldsymbol{\theta}) & F_{\sigma_u^2\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\sigma_u^2\tau}(\boldsymbol{\theta}) \\ F_{\sigma^2\gamma}(\boldsymbol{\theta}) & \mathbf{0} & 0 & F_{\sigma^2\sigma_u^2}(\boldsymbol{\theta}) & F_{\sigma^2\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\sigma^2\tau}(\boldsymbol{\theta}) \\ \boldsymbol{F}_{\tau\gamma}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} & \boldsymbol{F}_{\tau\sigma_u^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\tau\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\tau\tau}(\boldsymbol{\theta}) \end{bmatrix}.$$

Note that one has orthogonality between $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ as well as between μ and $\boldsymbol{\eta}$.

3.3 Computation of the maximum likelihood estimates

The maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ will be obtained by maximizing the log-likelihood function $L(\boldsymbol{\theta})$ over the parametric space Θ. Thus, the value of $\boldsymbol{\theta}$ that maximizes $L(\boldsymbol{\theta})$ over Θ should satisfy

$$L(\widehat{\boldsymbol{\theta}}) \geq \operatorname{argmax}_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}).$$

The determination of $\widehat{\boldsymbol{\theta}}$ may be performed by alternating the solution of the estimating equations $U^{\gamma}(\boldsymbol{\theta}) = 0$, $\boldsymbol{U}^{\beta}(\boldsymbol{\theta}) = \mathbf{0}$ and $U^{\mu}(\boldsymbol{\theta}) = 0$ given $\boldsymbol{\eta}$, through a Fisher scoring algorithm, with the maximization of the log-likelihood function in $\boldsymbol{\eta}$ given the parameters $\boldsymbol{\beta}$, γ and μ. This procedure may be represented by the following iterative process:

$$\begin{aligned}
\gamma^{(k+1)} = {} & \left\{ -\sum_{i=1}^{n} \frac{1}{2} v(\delta_i^{(k)}) \boldsymbol{r}_i^{(k)\top} \boldsymbol{A}_i^{(k)} \begin{pmatrix} \mu^{(k)} \mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix} \right\}^{-1} \\
& \times \sum_{i=1}^{n} \left\{ \boldsymbol{c}_i^{(k)} + v(\delta_i^{(k)}) \boldsymbol{r}_i^{(k)\top} \boldsymbol{V}_i^{-1(k)} \begin{pmatrix} \mu^{(k)} \mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix} \right. \\
& \left. + \frac{1}{2} v(\delta_i^{(k)}) \boldsymbol{r}_i^{(k)\top} \boldsymbol{A}_i^{(k)} \begin{pmatrix} \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}^{(k)} \\ \boldsymbol{u}_i - \mu^{(k)} \mathbf{1}_{m_i} \end{pmatrix} \right\},
\end{aligned} \tag{9}$$

$$\begin{aligned}
\boldsymbol{\beta}^{(k+1)} = {} & \left\{ \sum_{i=1}^{n} v(\delta_i^{(k)}) \begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^{\top} \boldsymbol{V}_i^{-1(k)} \begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix} \right\}^{-1} \\
& \times \sum_{i=1}^{n} v(\delta_i^{(k)}) \begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^{\top} \boldsymbol{V}_i^{-1(k)} \begin{pmatrix} \boldsymbol{y}_i - \gamma^{(k)} \mu^{(k)} \mathbf{1}_{m_i} \\ \boldsymbol{u}_i - \mu^{(k)} \mathbf{1}_{m_i} \end{pmatrix},
\end{aligned} \tag{10}$$

$$\begin{aligned}
\mu^{(k+1)} = {} & \left\{ \sum_{i=1}^{n} v(\delta_i^{(k)}) \begin{pmatrix} \gamma^{(k)} \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^{\top} \boldsymbol{V}_i^{-1(k)} \begin{pmatrix} \gamma^{(k)} \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix} \right\}^{-1} \\
& \times \sum_{i=1}^{n} v(\delta_i^{(k)}) \begin{pmatrix} \gamma^{(k)} \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^{\top} \boldsymbol{V}_i^{-1(k)} \begin{pmatrix} \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}^{(k)} \\ \boldsymbol{u}_i \end{pmatrix}
\end{aligned} \tag{11}$$

and

$$\boldsymbol{\eta}^{(k+1)} = \operatorname{argmax}_{\boldsymbol{\eta}} \left\{ L\left(\gamma^{(k)}, \boldsymbol{\beta}^{(k)}, \mu^{(k)}, \boldsymbol{\eta}\right) \right\}, \tag{12}$$

for $k = 0, 1, 2, \ldots$. Starting values should be given for the parameters $\boldsymbol{\beta}$, γ, μ and $\boldsymbol{\eta}$ in the iterative process (9)–(12). In particular, for non Gaussian elliptical models, such values may be assigned as the maximum likelihood estimates from the normal linear mixed model with a covariate subject to measurement error. Due to the similarity between elliptical mixed models and normal mixed models, is reasonable to expect for large n that the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ may be approximated by a $(p+q+s+4)$-dimensional normal distribution of mean $\boldsymbol{\theta}$ and approximate variance–covariance matrix given by $\widehat{\text{Var}}_{\text{approx}}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{F}(\boldsymbol{\theta})^{-1}$. Nevertheless, in Sect. 5, we summarize the results of a simulation study to assess the sample properties of $\widehat{\boldsymbol{\theta}}$ under different parameter settings and model error distributions.

4 Prediction of the random effects and covariate measured with error

The prediction of the random effect $\boldsymbol{b}_i$ as well as of the covariate $\boldsymbol{u}_i^*$ measured with error may be performed by applying the well known properties of the elliptical class, that conditional and marginal distributions of elliptical random variables are also elliptical. Thus, similarly to the normal case (see, for instance, Zhang et al. 1998), one may derive empirical Bayes estimates for $\boldsymbol{b}_i$ and $\boldsymbol{u}_i^*$.

In particular from (3), the marginal distribution of $(\boldsymbol{W}_i^\top, \boldsymbol{b}_i^\top)^\top$ becomes given by

$$\begin{pmatrix} \boldsymbol{W}_i \\ \boldsymbol{b}_i \end{pmatrix} \overset{\text{ind}}{\sim} \text{El}_{2m_i+q}\left(\begin{pmatrix} \boldsymbol{\mu}_{iW} \\ \boldsymbol{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{V}_i & \boldsymbol{C}_{Wb} \\ \boldsymbol{C}_{Wb}^\top & \boldsymbol{D} \end{bmatrix} \right),$$

where $\boldsymbol{C}_{Wb} = \begin{pmatrix} \boldsymbol{Z}_i \boldsymbol{D} \\ \boldsymbol{0} \end{pmatrix}$ denotes the covariance between $\boldsymbol{W}_i$ and $\boldsymbol{b}_i$. Therefore, from the elliptical class properties it follows that

$$\boldsymbol{b}_i | \boldsymbol{W}_i \sim \text{El}_q \left(\boldsymbol{C}_{Wb}^\top \boldsymbol{V}_i^{-1} (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}); \boldsymbol{D} - \boldsymbol{C}_{Wb}^\top \boldsymbol{V}_i^{-1} \boldsymbol{C}_{Wb} \right).$$

Then, from $\text{E}(\boldsymbol{b}_i | \boldsymbol{W}_i)$ we may derive the empirical Bayes estimate of $\boldsymbol{b}_i$ as

$$\begin{aligned} \widehat{\boldsymbol{b}}_i &= \widehat{E(\boldsymbol{b}_i | \boldsymbol{W}_i)} \\ &= \widehat{\boldsymbol{C}}_{Wb}^\top \widehat{\boldsymbol{V}}_i^{-1} (\boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_{iW}). \end{aligned} \tag{13}$$

Similarly, for obtaining the empirical Bayes estimate of $\boldsymbol{u}_i^*$, we may derive the marginal distribution of $(\boldsymbol{W}_i^\top, \boldsymbol{u}_i^{*\top})^\top$ as

$$\begin{pmatrix} \boldsymbol{W}_i \\ \boldsymbol{u}_i^* \end{pmatrix} \overset{\text{ind}}{\sim} \text{El}_{3m_i}\left(\begin{pmatrix} \boldsymbol{\mu}_{iW} \\ \mu \boldsymbol{1}_{m_i} \end{pmatrix}, \begin{bmatrix} \boldsymbol{V}_i & \boldsymbol{C}_{Wu^*} \\ \boldsymbol{C}_{Wu^*}^\top & \sigma_u^2 \boldsymbol{I}_{m_i} \end{bmatrix} \right),$$

where $\boldsymbol{C}_{Wu^*} = \begin{pmatrix} \gamma \sigma_u^2 \boldsymbol{I}_{m_i} \\ \sigma_u^2 \boldsymbol{I}_{m_i} \end{pmatrix}$ denotes the covariance between $\boldsymbol{W}_i$ and $\boldsymbol{u}_i^*$. Therefore, from the elliptical class properties it follows that

$$\boldsymbol{u}_i^* | \boldsymbol{W}_i \sim \text{El}_{m_i} \left(\mu \boldsymbol{1}_{m_i} + \boldsymbol{C}_{Wu^*}^\top \boldsymbol{V}_i^{-1} (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}); \sigma_u^2 \boldsymbol{I}_{m_i} - \boldsymbol{C}_{Wu^*}^\top \boldsymbol{V}_i^{-1} \boldsymbol{C}_{Wu^*} \right).$$

Then, the empirical Bayes estimate of $\boldsymbol{u}_i^*$ takes the form

$$\begin{aligned} \widehat{\boldsymbol{u}^*}_i &= \widehat{E(\boldsymbol{u}_i^* | \boldsymbol{W}_i)} \\ &= \widehat{\mu} \boldsymbol{1}_{m_i} + \widehat{\boldsymbol{C}}_{Wu^*}^\top \widehat{\boldsymbol{V}}_i^{-1} (\boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_{iW}). \end{aligned} \tag{14}$$

5 Simulation study

In this section we summarize the results of a small simulation study performed to assess the empirical behavior of the maximum likelihood estimates and their asymptotic standard errors from the elliptical model

$$\boldsymbol{W}_i \overset{\text{ind}}{\sim} \mathrm{El}_{2\mathrm{m_i}}\left(\boldsymbol{\mu}_{iW}, \boldsymbol{V}_i\right),$$

with

$$\boldsymbol{\mu}_{iW} = \begin{pmatrix} \beta \boldsymbol{X}_i + \gamma\mu\mathbf{1}_{m_i} \\ \mu\mathbf{1}_{m_i} \end{pmatrix}$$

and

$$\boldsymbol{V}_i = \begin{bmatrix} \tau\mathbf{1}_{m_i}\mathbf{1}_{m_i}^\top + (\gamma^2\sigma_u^2 + \sigma^2)\boldsymbol{I}_{m_i} & \gamma\sigma_u^2\boldsymbol{I}_{m_i} \\ \gamma\sigma_u^2\boldsymbol{I}_{m_i} & (\sigma_u^2 + \sigma_e^2)\boldsymbol{I}_{m_i} \end{bmatrix},$$

where $\boldsymbol{X}_i = (x_i, \ldots, x_i)^\top$, for $i = 1, \ldots, n$. The values assigned for the parameters are $\gamma = 0.5, 1, 2, 5$, $\beta = 2, 5, 10$, $\mu = 1$, $\tau = 1, 2, 5, 10$, $\sigma_u^2 = 5, 10, 15$ and $\sigma^2 = 10, 15, 20, 30$. The explanatory variable values x_i's are generated from a uniform distribution in the interval $[-2, 2]$, we assume $\sigma_e^2 = 0.2$, $n = 15, 30, 100$ and $m_i = 5$ and two different error distributions are considered, normal and Student-t with $\nu = 4$ degrees of freedom.

We perform the calculation of the relative bias of each estimate and of its approximate variance from the Fisher information matrix, as well as the normal approximation for the sample distribution of each estimate available by the statistic of the Kolmogorov-Smirnov goodness-of-fit test. The relative bias of any estimate from $\boldsymbol{\theta} = (\gamma, \beta, \mu, \tau, \sigma_u^2, \sigma^2)^\top$, namely $\hat{\theta}$, is estimated as $100 \times |\bar{\theta} - \theta|/\theta$, where $\bar{\theta} = R^{-1}\sum_{r=1}^{R}\hat{\theta}^{(r)}$, with $\hat{\theta}^{(r)}$ being the maximum likelihood estimate of θ in the rth replication. The relative bias of $\sqrt{\widehat{\mathrm{Var}}_{\mathrm{approx}}(\hat{\theta})}$ is estimated as $100 \times |\sqrt{\bar{\mathrm{V}}\mathrm{ar}(\hat{\theta})} - \sqrt{\mathrm{Var}(\hat{\theta})}|/\sqrt{\mathrm{Var}(\hat{\theta})}$, where $\sqrt{\bar{\mathrm{V}}\mathrm{ar}(\hat{\theta})} = R^{-1}\sum_{r=1}^{R}\sqrt{\widehat{\mathrm{Var}}_{\mathrm{approx}}^{(r)}(\hat{\theta})}$ and $\mathrm{Var}(\hat{\theta}) = (R-1)^{-1}\sum_{r=1}^{R}[\hat{\theta}^{(r)} - \bar{\theta}]^2$, with $\widehat{\mathrm{Var}}_{\mathrm{approx}}^{(r)}(\hat{\theta})$ being the estimate of $\mathrm{Var}(\hat{\theta})$ in the rth replication. We consider in each scenario $R = 500$ replicates.

Tables 1, 2 and 3 describe the relative biases and we may notice that in general the parameter estimate biases and the standard error biases are less than 5% and the biases seem to decrease as the sample size increases. The only parameter for which the maximum likelihood estimate and its standard error estimate remain with bias greater than 5% in some scenarios is the random effect variance. From the KS statistical test values each parameter estimate seems to follow approximately a normal distribution even for small sample size. We also observe that the biases are in general smaller under the Student-*t* models.

Table 1 Relative bias of the maximum likelihood estimate (Rb-MLE)(in %), relative bias of the asymptotic standard error (Rb-ASE)(in %) and the p-value of the Kolmogorov–Smirnov statistic for assessing the normality of the parameter estimates from the simulation study of linear elliptical models under normal and Student-t ($\nu = 4$) error distributions and $n = 15$

Parameter	Normal			Student-t		
	Rb-MLE	Rb-ASE	p-value	Rb-MLE	Rb-ASE	p-value
$\beta = 10$	0.32	0.08	0.85	0.24	0.08	0.87
$\gamma = 0.5$	0.46	0.04	0.93	0.41	0.06	0.41
$\mu = 1$	6.61	0.09	0.40	5.51	0.12	0.74
$\sigma_u^2 = 5$	1.65	0.56	0.18	1.05	0.71	0.61
$\sigma^2 = 10$	1.22	0.81	0.48	1.17	0.75	0.87
$\tau = 5$	7.89	3.60	0.32	2.94	3.18	0.84
$\beta = 10$	1.16	0.16	0.83	0.89	0.12	0.87
$\gamma = 1$	1.94	0.13	0.78	1.24	0.05	0.40
$\mu = 1$	6.08	0.06	0.46	4.79	0.07	0.69
$\sigma_u^2 = 5$	2.90	0.40	0.86	1.99	0.47	0.51
$\sigma^2 = 20$	2.60	1.97	0.77	1.55	1.91	0.58
$\tau = 5$	11.69	6.24	0.37	7.10	2.91	0.69
$\beta = 10$	1.01	0.17	0.82	0.76	0.07	0.85
$\gamma = 2$	1.17	0.07	0.89	1.15	0.05	0.70
$\mu = 1$	1.37	0.11	0.93	0.74	0.12	0.71
$\sigma_u^2 = 10$	1.78	0.93	0.89	1.09	1.04	0.41
$\sigma^2 = 20$	2.53	2.73	0.66	1.95	0.95	0.95
$\tau = 5$	5.58	6.87	0.68	4.02	2.37	0.79
$\beta = 2$	1.31	0.11	0.95	1.05	0.16	0.66
$\gamma = 1$	3.45	0.17	0.98	1.56	0.10	0.85
$\mu = 1$	1.27	0.04	0.92	1.03	0.06	0.91
$\sigma_u^2 = 5$	1.12	0.25	0.89	0.85	0.39	0.98
$\sigma^2 = 20$	1.87	2.15	0.53	1.42	1.75	0.97
$\tau = 5$	16.97	3.52	0.68	7.04	2.13	0.54
$\beta = 5$	1.73	0.17	0.95	0.95	0.10	0.64
$\gamma = 2$	2.97	0.04	0.67	1.18	0.06	0.78
$\mu = 1$	2.80	0.13	0.52	1.05	0.05	0.70
$\sigma_u^2 = 15$	1.52	1.53	0.99	1.22	0.57	0.88
$\sigma^2 = 15$	6.44	2.62	0.35	5.10	1.62	0.71
$\tau = 2$	13.04	2.17	0.95	8.69	1.63	0.95
$\beta = 2$	5.95	0.21	0.58	3.39	0.33	0.36
$\gamma = 5$	1.37	0.10	0.94	1.10	0.23	0.47
$\mu = 1$	2.26	0.08	0.94	1.66	0.11	0.92
$\sigma_u^2 = 5$	1.14	0.54	0.91	1.09	0.66	0.85
$\sigma^2 = 30$	4.47	4.73	0.58	3.42	1.14	0.63
$\tau = 10$	9.72	5.50	0.31	5.49	2.96	0.62

Table 2 Relative bias of the maximum likelihood estimate (Rb-MLE) (in %), relative bias of the asymptotic standard error (Rb-ASE) (in %) and the p-value of the Kolmogorov–Smirnov statistic for assessing the normality of the parameter estimates from the simulation study of linear elliptical models under normal and Student-t ($\nu = 4$) error distributions and $n = 30$

Parameter	Normal			Student-t		
	Rb-MLE	Rb-ASE	p-value	Rb-MLE	Rb-ASE	p-value
$\beta = 10$	0.25	0.05	0.67	0.17	0.05	0.97
$\gamma = 0.5$	0.26	0.04	0.98	0.18	0.07	0.77
$\mu = 1$	2.83	0.03	0.92	1.40	0.04	0.89
$\sigma_u^2 = 5$	0.99	0.22	0.76	0.36	0.13	0.50
$\sigma^2 = 10$	1.18	0.50	0.89	1.12	0.52	0.67
$\tau = 5$	4.19	1.58	0.10	2.05	1.96	0.73
$\beta = 10$	0.87	0.13	0.54	0.59	0.04	0.68
$\gamma = 1$	1.00	0.07	0.72	0.85	0.08	0.55
$\mu = 1$	4.35	0.05	0.97	3.60	0.06	0.42
$\sigma_u^2 = 5$	1.54	0.30	0.74	1.02	0.37	0.75
$\sigma^2 = 20$	1.39	2.06	0.86	1.03	0.77	0.89
$\tau = 5$	6.28	5.22	0.77	3.50	1.15	0.66
$\beta = 10$	0.89	0.06	0.94	0.46	0.07	0.78
$\gamma = 2$	1.05	0.05	0.66	0.88	0.04	0.94
$\mu = 1$	1.23	0.12	0.87	0.54	0.13	0.15
$\sigma_u^2 = 10$	1.19	1.12	0.84	0.74	0.54	0.83
$\sigma^2 = 20$	1.29	0.85	0.85	1.11	0.22	0.77
$\tau = 5$	3.36	3.28	0.69	2.20	1.57	0.60
$\beta = 2$	0.90	0.09	0.86	0.24	0.08	0.42
$\gamma = 1$	2.13	0.03	0.79	1.17	0.06	0.34
$\mu = 1$	0.66	0.04	0.87	0.55	0.06	0.08
$\sigma_u^2 = 5$	1.04	0.29	0.64	0.80	0.21	0.93
$\sigma^2 = 20$	1.56	1.36	1.00	0.96	0.50	0.91
$\tau = 5$	9.05	1.27	0.54	5.76	1.21	0.68
$\beta = 5$	1.66	0.08	0.78	0.88	0.08	0.84
$\gamma = 2$	1.48	0.04	0.75	0.83	0.02	0.81
$\mu = 1$	2.24	0.05	0.69	0.97	0.05	0.88
$\sigma_u^2 = 15$	1.29	0.62	0.59	1.14	0.54	0.95
$\sigma^2 = 15$	4.31	1.05	0.63	3.59	1.25	0.60
$\tau = 2$	9.59	2.65	0.12	5.42	0.55	0.52
$\beta = 2$	4.33	0.10	0.89	2.88	0.25	0.91
$\gamma = 5$	1.31	0.07	0.85	0.89	0.10	0.73
$\mu = 1$	1.83	0.03	0.49	1.19	0.08	0.77
$\sigma_u^2 = 5$	1.05	0.16	0.92	0.78	0.53	0.89
$\sigma^2 = 30$	2.09	2.07	0.64	1.32	0.88	0.94
$\tau = 10$	5.32	5.21	0.78	2.03	1.88	0.34

Table 3 Relative bias of the maximum likelihood estimate (Rb-MLE) (in %), relative bias of the asymptotic standard error (Rb-ASE) (in %) and the p-value of the Kolmogorov–Smirnov statistic for assessing the normality of the parameter estimates from the simulation study of linear elliptical models under normal and Student-t ($\nu = 4$) error distributions and $n = 100$

Parameter	Normal			Student-t		
	Rb-MLE	Rb-ASE	p-value	Rb-MLE	Rb-ASE	p-value
$\beta = 10$	0.17	0.03	0.81	0.12	0.10	0.91
$\gamma = 0.5$	0.11	0.02	0.64	0.09	0.02	0.57
$\mu = 1$	0.58	0.03	0.42	0.39	0.06	0.48
$\sigma_u^2 = 5$	0.36	0.16	0.93	0.13	0.36	0.76
$\sigma^2 = 10$	0.89	0.38	0.50	0.28	1.11	0.79
$\tau = 5$	3.08	0.31	0.91	1.54	1.05	0.49
$\beta = 10$	0.44	0.07	0.69	0.39	0.01	0.31
$\gamma = 1$	0.40	0.02	0.51	0.18	0.03	0.89
$\mu = 1$	2.50	0.03	0.87	1.05	0.01	0.52
$\sigma_u^2 = 5$	1.03	0.18	0.84	0.79	0.08	0.72
$\sigma^2 = 20$	0.84	1.19	0.90	0.51	0.17	0.35
$\tau = 5$	3.51	1.69	0.85	1.80	0.61	0.38
$\beta = 10$	0.34	0.03	0.97	0.28	0.04	0.94
$\gamma = 2$	0.62	0.04	0.84	0.58	0.02	0.58
$\mu = 1$	1.06	0.07	0.77	0.32	0.04	0.92
$\sigma_u^2 = 10$	0.88	0.67	0.26	0.22	0.38	0.78
$\sigma^2 = 20$	1.07	0.57	0.83	1.01	0.19	0.77
$\tau = 5$	2.25	1.47	0.88	1.61	0.84	0.76
$\beta = 2$	0.64	0.02	0.56	0.16	0.01	0.38
$\gamma = 1$	1.55	0.03	0.54	0.55	0.05	0.70
$\mu = 1$	0.34	0.02	0.48	0.31	0.02	0.82
$\sigma_u^2 = 5$	0.73	0.14	0.28	0.44	0.11	0.82
$\sigma^2 = 20$	0.51	0.56	0.52	0.26	0.38	0.68
$\tau = 5$	3.02	1.12	0.78	1.77	0.50	0.48
$\beta = 5$	1.13	0.03	0.82	0.70	0.02	0.53
$\gamma = 2$	1.02	0.02	0.98	0.69	0.01	0.63
$\mu = 1$	1.18	0.07	0.96	0.65	0.04	0.76
$\sigma_u^2 = 15$	1.11	0.77	0.97	0.98	0.29	0.87
$\sigma^2 = 15$	3.69	0.56	0.72	2.12	0.31	0.66
$\tau = 2$	6.28	1.18	0.85	3.78	0.26	0.91
$\beta = 2$	2.93	0.09	0.65	1.23	0.08	0.55
$\gamma = 5$	0.89	0.05	0.37	0.23	0.09	0.40
$\mu = 1$	1.06	0.03	0.49	0.81	0.04	0.65
$\sigma_u^2 = 5$	0.82	0.14	0.90	0.17	0.25	0.82
$\sigma^2 = 30$	1.09	1.84	0.61	0.76	0.49	0.19
$\tau = 10$	1.37	2.40	0.58	0.98	0.97	0.60

6 Diagnostic methods

6.1 Goodness-of-fit methods

In order to assess the goodness-of-fit of elliptical linear models one may consider, for instance, index plots of the Mahalanobis distance or modified Mahalanobis distances (see, for instance, Osorio et al. 2007). In particular, for normal and Student-t (with ν degrees of freedom) error models one may apply the modified Mahalanobis distance $\vartheta_i = \frac{\delta_i}{2m_i}$ which follows, respectively, under the underling models $\chi^2_{2m_i}$ and $F_{2m_i,\nu}$. Alternatively, one may apply the transformations proposed by Wilson–Hilferty (see, for instance, Johnson et al. 1994) for obtaining appropriate scores that may be compared with the quantiles of the standard normal distribution. Such scores for normal and Student-t models are, respectively, given by

$$d_i^{[N]} = \frac{\left(\frac{\delta_i}{2m_i}\right)^{\frac{1}{3}} - \left(1 - \frac{1}{9m_i}\right)}{\left(\frac{1}{9m_i}\right)^{\frac{1}{2}}}$$

and

$$d_i^{[t]} = \frac{\left(1 - \frac{2}{9\nu}\right)\vartheta_i^{\frac{1}{3}} - \left(1 - \frac{1}{9m_i}\right)}{\left(\frac{2}{9\nu}\vartheta_i^{\frac{2}{3}} + \frac{1}{9m_i}\right)^{\frac{1}{2}}},$$

and, approximately, one has that $d_i^{[N]}, d_i^{[t]} \sim \mathrm{N}(0, 1)$, for $i = 1, \dots, n$.

6.2 Local influence

The local influence method proposed by Cook (1986) has been largely applied in the statistical modeling area to assess the impact of minor perturbations in the model or data on the parameter estimates. In this sense, the influence measure

$$LD(\boldsymbol{\omega}) = 2\left[L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}_\omega|\boldsymbol{\omega})\right],$$

named likelihood displacement, has been proposed to compare the perturbed and fitted models, where $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ denotes the perturbed log-likelihood function with $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ being the perturbation vector and $\widehat{\boldsymbol{\theta}}_\omega$ is the maximum likelihood estimate under the perturbed model. In addition, there is a no perturbation vector $\boldsymbol{\omega}_0$ such that $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$. Therefore, one has $LD(\boldsymbol{\omega}) \geq 0$. The idea of the methodology is to study the behavior of the $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$ which consists in selecting a unit direction $\boldsymbol{\ell}$ ($||\boldsymbol{\ell}|| = 1$) and to consider the plot of $LD(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ versus a, for $a \in \Re$. This plot is called lifted line, and may be characterized by considering the normal curvature $C_\ell(\boldsymbol{\theta})$ around $a = 0$. Large values of $C_\ell(\boldsymbol{\theta})$ indicate sensitivity to the perturbation scheme in the direction $\boldsymbol{\ell}$. Cook (1986) showed that the normal curvature in the direction $\boldsymbol{\ell}$ may be expressed as

$$C_{\ell}(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^{\top}\boldsymbol{\Delta}^{\top}\{-\ddot{\boldsymbol{L}}(\hat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta}\boldsymbol{\ell}|,$$

where $\boldsymbol{\ell} \in \boldsymbol{\Omega}$, $||\boldsymbol{\ell}|| = 1$, with

$$\ddot{\boldsymbol{L}}(\boldsymbol{\theta}) = \left.\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\right|_{\theta=\widehat{\theta}} \quad \text{and} \quad \boldsymbol{\Delta} = \left.\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^{\top}}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0},$$

$-\ddot{\boldsymbol{L}}(\hat{\boldsymbol{\theta}})$ denotes the observed information matrix of $\boldsymbol{\theta}$ whose expressions are derived in Appendix B, whereas $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\gamma}, \boldsymbol{\Delta}_{\beta}^{\top}, \boldsymbol{\Delta}_{\mu}, \boldsymbol{\Delta}_{\sigma_u^2}, \boldsymbol{\Delta}_{\sigma^2}, \boldsymbol{\Delta}_{\tau}^{\top})^{\top}$ is the perturbation matrix, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. The direction $\boldsymbol{\ell}_{max}$ from the maximum curvature, C_{max}, corresponds to the eigenvector related with the largest eigenvalue (in absolute value) of the matrix $\boldsymbol{B} = \boldsymbol{\Delta}^{\top}\{-\ddot{\boldsymbol{L}}(\hat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta}$, that is, the direction related with the largest variation. The index plot of $|\boldsymbol{\ell}_{max}|$ may reveal those observations that contribute most in such direction.

Poon and Poon (1999) proposed the conformal normal curvature, that is invariant under uniform changes of scale, and is defined as

$$B_{\ell}(\boldsymbol{\theta}) = \frac{C_{\ell}(\boldsymbol{\theta})}{2||\boldsymbol{\ell}^{\top}\boldsymbol{\Delta}^{\top}\{-\ddot{\boldsymbol{L}}(\hat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta}\boldsymbol{\ell}||_F},$$

where $||\cdot||_F$ denotes the Frobenius norm, $||\boldsymbol{A}||_F = \{\mathrm{tr}(\boldsymbol{A}^{\top}\boldsymbol{A})\}^{\frac{1}{2}}$ for the matrix $\boldsymbol{A}$. A feature of this curvature is that $0 \leq B_{\ell}(\boldsymbol{\theta}) \leq 1$. For example, the index plot of $B_{\ell_i}(\boldsymbol{\theta})$, where $\boldsymbol{\ell}_i$ is an $n \times 1$ vector of zeros with 1 at the ith position, may be performed to assess the influence of each observation on $B_{\ell}(\boldsymbol{\theta})$. Other influence graphs based on normal conformal curvature were proposed by Poon and Poon (1999).

Various works have been developed on the derivation of local influence curvatures in elliptical linear models. For example, Galea et al. (2003) derived such curvatures in univariate elliptical linear models, whereas Osorio et al. (2007) and Ibacache-Pulgar et al. (2012) derived the local influence curvatures in parametric and semiparametric elliptical linear mixed models, respectively. Similarly, Cao et al. (2010) and Relvas and Paula (2016) obtained the local influence graphs for univariate AR(1) elliptical models.

6.2.1 Selecting the perturbation scheme

Zhu et al. (2007) proposed a method for selecting appropriate perturbation schemes for the perturbed model whose probability density function, denoted by $f(\boldsymbol{W}; \boldsymbol{\theta}, \boldsymbol{\omega})$, leads to the perturbed log-likelihood function $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}|\omega_i)$. The methodology is based on the Fisher information matrix with respect to the perturbation vector $\boldsymbol{\omega}$, under the perturbed model. This matrix is denoted by $\boldsymbol{G}(\boldsymbol{\omega}) = \{g_{ij}(\boldsymbol{\omega})\}$, with

$$g_{ij}(\boldsymbol{\omega}) = \mathrm{E}_{\omega}\left\{\frac{\partial}{\partial \omega_i} L(\boldsymbol{\theta}|\boldsymbol{\omega}) \frac{\partial}{\partial \omega_j} L(\boldsymbol{\theta}|\boldsymbol{\omega})\right\}, \quad i, j = 1, \ldots, n,$$

where E_{ω} denotes the expectation taken with respect to the probability density function of the perturbed model. The elements $g_{ii}(\boldsymbol{\omega})$ of $\boldsymbol{G}(\boldsymbol{\omega})$ are the variances of the scores with respect to the components of $\boldsymbol{\omega}$, and indicate the amount of perturbation introduced by the ith component of $\boldsymbol{\omega}$. The off-diagonal elements of $\boldsymbol{G}(\boldsymbol{\omega})$ are the covariances among the scores with respect to different components of $\boldsymbol{\omega}$ and represent the linear correlation between different components of $\boldsymbol{\omega}$. Thus, if $\boldsymbol{G}(\boldsymbol{\omega})$ is a diagonal matrix, the components of $\boldsymbol{\omega}$ are orthogonal in the sense of Cox and Reid (1987). Therefore, a perturbation scheme will be considered appropriate if $\boldsymbol{G}(\boldsymbol{\omega}_0) = c\boldsymbol{I}_n$, where $c > 0$. However, if a perturbation scheme satisfies

$$\boldsymbol{G}(\boldsymbol{\omega}_0) = \operatorname{diag}\{g_{11}(\boldsymbol{\omega}_0), \ldots, g_{nn}(\boldsymbol{\omega}_0)\},$$

we can choose an appropriate perturbation scheme $\tilde{\boldsymbol{\omega}}$, by making

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}_0 + c^{-\frac{1}{2}} \boldsymbol{G}(\boldsymbol{\omega}_0)^{\frac{1}{2}} (\boldsymbol{\omega} - \boldsymbol{\omega}_0),$$

so that $\boldsymbol{G}(\tilde{\boldsymbol{\omega}})$ evaluated at $\boldsymbol{\omega}_0$ equals $c\boldsymbol{I}_n$.

Nevertheless, few works have been developed under the approach of selecting the perturbation scheme of local influence. A recent and interesting derivation in capital asset pricing model under the multivariate normal distribution is described in Galea and Giménez (2016).

Case-weight perturbation

Under this perturbation scheme we assume that

$$L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i L_i(\boldsymbol{\theta}),$$

where $L_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{V}_i| + \log\{\mathrm{g}(\delta_i)\}$, $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^{\top}$. For this perturbation scheme $\boldsymbol{G}(\boldsymbol{\omega}) = \operatorname{diag}[\operatorname{Var}_{\omega}\{L_1(\boldsymbol{\theta})\}, \ldots, \operatorname{Var}_{\omega}\{L_n(\boldsymbol{\theta})\}]$, and the form $\boldsymbol{G}(\boldsymbol{\omega}_0) = \mathrm{c}\boldsymbol{I}_n$ is only attained when $m_1 = m_2 = \cdots = m_n$. In general, one has that

$$\tilde{\omega_i} = 1 + (\omega_i - 1)\sqrt{\operatorname{Var}_{\omega_0}\{L_i(\boldsymbol{\theta})\}},$$

so that the perturbed log-likelihood function becomes given by

$$L(\boldsymbol{\theta}|\tilde{\boldsymbol{\omega}}) = \sum_{i=1}^{n} \left[1 + (\omega_i - 1)\sqrt{\operatorname{Var}_{\omega_0}\{L_i(\boldsymbol{\theta})\}}\right] L_i(\boldsymbol{\theta}).$$

Below we derive $\tilde{\omega_i}$ under the particular cases of normal and Student-t errors.

Normal case

In this case one has that

$$\begin{aligned}\mathrm{Var}_{\omega_0}\{L_i(\boldsymbol{\theta})\} &= \mathrm{Var}\left(-\frac{1}{2}\boldsymbol{r}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right)\\ &= \frac{1}{4}\mathrm{Var}(\delta_i) = m_i,\end{aligned}$$

since $\delta_i \sim \chi^2_{2m_i}$. Then, we have that $\tilde{\omega_i} = 1 + (\omega_i - 1)\sqrt{m_i}$.

Student-t case

Here, we obtain

$$\mathrm{Var}_{\omega_0}\{L_i(\boldsymbol{\theta})\} = \left(\frac{2m_i + \nu}{2}\right)^2 \mathrm{Var}\left\{\log(1 + \nu^{-1}\delta_i)\right\},$$

where

$$\mathrm{Var}\left\{\log(1 + \nu^{-1}\delta_i)\right\} = \Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu + 2m_i}{2}\right),$$

with $\Psi'(\cdot)$ denoting the trigamma function (see, for instance Arellano-Valle 2010). Then, the perturbation scheme yields

$$\tilde{\omega_i} = 1 + \frac{2m_i + \nu}{2}\sqrt{\Psi'\left(\frac{\nu}{2}\right) - (\omega_i - 1)\Psi'\left(\frac{\nu + 2m_i}{2}\right)}, \quad i = 1, \ldots, n.$$

The elements of the matrix $\boldsymbol{\Delta}$ for this perturbation scheme are given in Appendix C.

Scale matrix perturbation

In this case we will consider the perturbation scheme $\omega_i^{-1}\boldsymbol{V}_i$, with $\omega_i > 0$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. The perturbed log-likelihood function becomes given by

$$L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}|\omega_i),$$

where

$$\begin{aligned}L_i(\boldsymbol{\theta}|\omega_i) &= -\frac{1}{2}\log|\omega_i^{-1}\boldsymbol{V}_i| + \log\{g(\delta_{\omega_i})\}\\ &= m_i \log\omega_i - \frac{1}{2}\log|\boldsymbol{V}_i| + \log\{g(\delta_{\omega_i})\},\end{aligned}$$

with $\delta_{\omega_i} = \omega_i\delta_i = (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^\top \omega_i \boldsymbol{V}_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})$.

The score function for ω_i becomes given by

$$\begin{aligned} U_{\omega_i} &= \frac{\partial L_i(\boldsymbol{\theta}|\omega_i)}{\partial \omega_i} \\ &= \frac{m_i}{\omega_i} + W_g(\delta_{\omega_i})\delta_i, \end{aligned}$$

and

$$\frac{\partial^2 L_i(\boldsymbol{\theta}|\omega_i)}{\partial \omega_i \partial \omega_j} = 0, \quad \forall\, i \neq j.$$

Hence

$$\mathrm{g}_{ij}(\boldsymbol{\omega}) = \begin{cases} \mathrm{E}_\omega \left[\left\{ \frac{m_i}{\omega_i} + W_g(\delta_{\omega_i})\delta_i \right\}^2 \right], & i = j, \\ 0, \quad i \neq j \end{cases}$$

and thus $\boldsymbol{G}(\boldsymbol{\omega}) = \mathrm{diag}\{\mathrm{g}_{11}(\boldsymbol{\omega}), \ldots, \mathrm{g}_{nn}(\boldsymbol{\omega})\}$.

Then, we have

$$\begin{aligned} \mathrm{g}_{ii}(\boldsymbol{\omega}) &= \mathrm{E}_\omega \left\{ \left(\frac{m_i}{\omega_i} \right)^2 + \frac{2m_i}{\omega_i} W_g(\delta_{\omega_i})\delta_i + W_g^2(\delta_{\omega_i})\delta_i^2 \right\} \\ &= \left(\frac{m_i}{\omega_i} \right)^2 + \frac{2m_i}{\omega_i} \mathrm{E}_\omega \left\{ W_g(\delta_{\omega_i})\delta_i \right\} + \mathrm{E}_\omega \left\{ W_g^2(\delta_{\omega_i})\delta_i^2 \right\} \\ &= \left(\frac{m_i}{\omega_i} \right)^2 + \frac{2m_i}{\omega_i} \left(-\frac{m_i}{2\omega_i} \right) + \frac{f_{g_i}}{\omega_i^2} \\ &= \frac{f_{g_i}}{\omega_i^2}. \end{aligned}$$

Since

$$\begin{aligned} \mathrm{E}_\omega \left\{ W_g(\delta_{\omega_i})\delta_i \right\} &= \mathrm{E}_\omega \left\{ \frac{1}{\omega_i} W_g(\delta_{\omega_i})\delta_{\omega_i} \right\} \\ &= \frac{1}{\omega_i} \left(-\frac{m_i}{2} \right) \\ &= -\frac{m_i}{2\omega_i} \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}_\omega \left\{ W_g^2(\delta_{\omega_i})\delta_{\omega_i}^2 \frac{1}{\omega_i^2} \right\} &= \frac{1}{\omega_i^2} \mathrm{E}_\omega \left\{ W_g^2(\delta_{\omega_i})\delta_{\omega_i}^2 \right\} \\ &= \frac{1}{\omega_i^2} f_{g_i}, \end{aligned}$$

we obtain

$$\begin{aligned}\boldsymbol{G}(\boldsymbol{\omega}_0) &= \operatorname{diag}\left(\frac{f_{g_1}}{\omega_{10}^2}, \ldots, \frac{f_{g_n}}{\omega_{n0}^2}\right)\\ &= \operatorname{diag}\left(f_{g_1}, \ldots, f_{g_n}\right).\end{aligned}$$

Note that $\boldsymbol{G}(\boldsymbol{\omega}_0) \neq c\boldsymbol{I}_n$. Some algebraic manipulation leads to $\tilde{\omega}_i = 1 + (\omega_i - 1)\sqrt{f_{g_i}}$, for $i = 1, \ldots, n$. Expressions for f_{g_i} are derived in Appendix A.

The elements of the matrix $\boldsymbol{\Delta}$ for this perturbation scheme are given in Appendix C.

Observed data perturbation

We will consider the perturbation scheme

$$\boldsymbol{W}_{\omega_i} = \boldsymbol{W}_i + \boldsymbol{V}_i^{\frac{1}{2}}\boldsymbol{\omega}_i,$$

with $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{i2m_i})^\top$, $\boldsymbol{\omega}_i \in \Re^{2m_i}$, $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_n^\top)^\top$ and $\boldsymbol{\omega}_0$ is a $2m_i \times 1$ vector of zeros. Thus, the perturbed log-likelihood function takes the form

$$L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}|\boldsymbol{\omega}_i),$$

where $L_i(\boldsymbol{\theta}|\boldsymbol{\omega}_i) = -\frac{1}{2}\log|\boldsymbol{V}_i| + \log\{g(\delta_{\omega_i})\}$, with $\delta_{\omega_i} = (\boldsymbol{W}_{\omega_i} - \boldsymbol{\mu}_{iW})^\top \boldsymbol{V}_i^{-1}(\boldsymbol{W}_{\omega_i} - \boldsymbol{\mu}_{iW})$. The score function for $\boldsymbol{\omega}_i$ becomes given by

$$\begin{aligned}\boldsymbol{U}_{\omega_i} &= \frac{\partial L_i(\boldsymbol{\theta}|\boldsymbol{\omega}_i)}{\partial \boldsymbol{\omega}_i}\\ &= W_g(\delta_{\omega_i})\frac{\partial \delta_{\omega_i}}{\partial \boldsymbol{\omega}_i}\\ &= 2W_g(\delta_{\omega_i})\frac{\partial (\boldsymbol{V}_{\omega_i}^{\frac{1}{2}})^\top}{\partial \boldsymbol{\omega}_i}\boldsymbol{V}_i^{-1}(\boldsymbol{W}_{\omega_i} - \boldsymbol{\mu}_{iW})\\ &= 2W_g(\delta_{\omega_i})\boldsymbol{V}_i^{-\frac{1}{2}}(\boldsymbol{W}_{\omega_i} - \boldsymbol{\mu}_{iW}).\end{aligned}$$

One has that $\boldsymbol{G}(\boldsymbol{\omega}) = \operatorname{diag}\{\boldsymbol{G}_{11}(\boldsymbol{\omega}), \ldots, \boldsymbol{G}_{nn}(\boldsymbol{\omega})\}$, where

$$\begin{aligned}\boldsymbol{G}_{ii}(\boldsymbol{\omega}) &= \mathrm{E}_\omega\{\boldsymbol{U}_{\omega_i}\boldsymbol{U}_{\omega_i}^\top\}\\ &= 4\mathrm{E}_\omega\left\{W_g^2(\delta_{\omega_i})\boldsymbol{V}_i^{-\frac{1}{2}}(\boldsymbol{W}_{\omega_i} - \boldsymbol{\mu}_{iW})(\boldsymbol{W}_{\omega_i} - \boldsymbol{\mu}_{iW})^\top \boldsymbol{V}_i^{-\frac{1}{2}}\right\}\\ &= \frac{4d_{g_i}}{2m_i}\boldsymbol{V}_i^{-\frac{1}{2}}\boldsymbol{V}_i\boldsymbol{V}_i^{-\frac{1}{2}}\\ &= c_i\boldsymbol{I}_{2m_i},\end{aligned}$$

with $c_i = \frac{2d_{g_i}}{m_i}$, for $i = 1, \ldots, n$.

Hence $\boldsymbol{G}_{ii}(\boldsymbol{\omega}) \neq c\boldsymbol{I}_{2m_i}$ and consequently $\boldsymbol{G}(\boldsymbol{\omega}) \neq cq$. We may apply a simple transformation to obtain

$$\begin{aligned}\tilde{\boldsymbol{\omega}} &= \boldsymbol{\omega}_0 + \boldsymbol{G}^{\frac{1}{2}}(\boldsymbol{\omega}_0)(\boldsymbol{\omega} - \boldsymbol{\omega}_0)\\ &= \boldsymbol{G}^{\frac{1}{2}}(\boldsymbol{\omega}_0)\boldsymbol{\omega}\\ &= \text{diag}(\sqrt{c_1}\boldsymbol{I}_{2m_1}, \ldots, \sqrt{c_n}\boldsymbol{I}_{2m_n})\boldsymbol{\omega}.\end{aligned}$$

Therefore, the perturbation scheme becomes given by $\tilde{\boldsymbol{\omega}}_i = \sqrt{\frac{m_i}{2d_{g_i}}}\boldsymbol{V}_i^{\frac{1}{2}}\boldsymbol{\omega}_i$, for $i = 1, \ldots, n$. Expressions for d_{g_i} are derived in Appendix A.

The elements of the matrix $\boldsymbol{\Delta}$ for this perturbation scheme are given in Appendix C.

7 Application

As a motivating example we will consider the housing price data that has been analyzed by various authors (see, for instance, Harrison and Rubinfeld 1978; Belsley et al. 1980; Zare and Rasekh 2011; Härdle and Simar 2012). The aim of the study is to assess the association of house prices with the air quality of the neighborhood by using regression models. The outcome variable LMV (logarithm of the median house price in USD 1000) is related with 13 explanatory variables, 6 of them are defined from census tracts and the remaining variables are defined for districts. Altogether there are 506 observations. Longford (1993) applied a linear mixed model with normal errors to analyze the complete data by assuming the 13 explanatory variables with fixed values and the parameter estimation performed by the maximum likelihood procedure. Some residual graphs available in the Longford's analysis point out extreme observations indicating that a heavy-tailed error distribution could be assumed in the model. On the other hand, Zhong et al. (2002) considered only the data of $n = 132$ census tracts within the 15 districts of Boston, but by adding in the model a covariate subject to measurement error and with the parameter estimation based upon the corrected score function of Nakamura (1990). The number of census tracts in the 15 districts of Boston are $\{8, 6, 3, 2, 7, 11, 13, 8, 19, 23, 11, 6, 7, 4, 4\}$.

Thus, motivated by the works addressed above, we will reanalyze the subset considered by Zhong et al. (2002) assuming the model (1)–(2) under normal and heavy-tailed error models. The census tracts explanatory variables considered in the analysis are: CRIME (per capita crime rate), ROOM (average number of rooms squared), AGE (percentage of owner-occupied units prior to 1940), DIST (logarithm of the weighted mean of distances to five employment centers in the Boston region), BLACK ($(B_k - 0.63)^2$, where B_k denotes the proportion of Black American in the census tract) and LSTAT (logarithm of the proportion of the population that is lower status). In addition, similarly to Zhong et al. (2002), the dummy variable CHAS (=1 if the census tract borders the Charles River, =0 otherwise) and the pollution covariate NOXSQ (annual average nitric oxide concentration, square of parts per 10^8), assumed subject to measurement error, will also be included in the model.

Proposed model

In order to reanalyze the subset of the house prices data we will assume the following linear mixed model:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + b_i \mathbf{1}_{m_i} + \gamma \boldsymbol{u}_i^* + \boldsymbol{\epsilon}_i, \tag{15}$$

$$\boldsymbol{u}_i = \boldsymbol{u}_i^* + \boldsymbol{e}_i, \tag{16}$$

where $\boldsymbol{y}_i$ denotes an $m_i \times 1$ vector with the values of LMV for the m_i census tracts of the ith district, $\boldsymbol{X}_i$ is an $m_i \times 8$ matrix whose jth row $\mathbf{x}_{ij}^\top$ contains a constant and the values of the explanatory variables CRIME, ROOM, AGE, DIST, BLACK, LSTAT and CHAS for the jth census tract in the ith district, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_7)^\top$ are the coefficients related with the fixed effects, b_i is the random-effect, $\mathbf{1}_{m_i}$ denotes an $m_i \times 1$ vector of $1's$, $\boldsymbol{u}_i^*$ denotes the covariate NOXSQ subject to measurement error for the ith district whereas $\boldsymbol{u}_i$ contains the values of NOXSQ for the m_i census tracts in the ith district, $i = 1, \ldots, 15$. In addition, we will assume for $\boldsymbol{W}_i = [\boldsymbol{y}_i^\top, \boldsymbol{u}_i^\top]^\top$ the following marginal distribution:

$$\boldsymbol{W}_i = \begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{u}_i \end{pmatrix} \overset{\text{ind}}{\sim} \text{El}_{2m_i}\left(\boldsymbol{\mu}_{iW}, \boldsymbol{V}_i\right), \tag{17}$$

where

$$\boldsymbol{\mu}_{iW} = \begin{pmatrix} \boldsymbol{X}_i \boldsymbol{\beta} + \gamma \mu \mathbf{1}_{m_i} \\ \mu \mathbf{1}_{m_i} \end{pmatrix}$$

and

$$\boldsymbol{V}_i = \begin{bmatrix} \tau \mathbf{1}_{m_i} \mathbf{1}_{m_i}^\top + (\gamma^2 \sigma_u^2 + \sigma^2) \boldsymbol{I}_{m_i} & \gamma \sigma_u^2 \boldsymbol{I}_{m_i} \\ \gamma \sigma_u^2 \boldsymbol{I}_{m_i} & (\sigma_u^2 + \sigma_e^2) \boldsymbol{I}_{m_i} \end{bmatrix}.$$

The estimation and inference we will be based on the marginal model. To avoid identification problems, the parameter β_0 and σ_e^2 will be assumed known, namely $\beta_0 = 9.0$ (obtained from the linear mixed model with normal error and no measurement errors in covariates) and $\sigma_e^2 = 0.2$.

Fitting the models

Table 5 presents the maximum likelihood estimates from the fitted models under normal and Student-t errors. We estimated the degrees of freedom for the Student$-t$ model by applying the Akaike criterion (Akaike 1974; Lange et al. 1989) and we found $\nu = 5$, see Table 4. We may notice from Table 5 that the explanatory variables LSTAT and NOXSQ are marginally significant for both models. This is an important difference from the results obtained by Zhong et al. (2002), that used the corrected score approach ($\hat{\gamma} = -0.0118(0.0084)$ and z-value$= -1.4$), where the covariate

Table 4 AIC values under some Student-t models with ν degrees of freedom

ν	AIC
1	890.73
2	883.31
3	880.88
4	880.03
5	**879.83**
6	879.95
∞	891.10

Bold value indicate the selected degrees of freedom

Table 5 Maximum likelihood estimates, approximate standard errors and some statistics from the elliptical model (15)–(17) under normal and Student-t errors (with $\nu = 5$ degrees of freedom) fitted to the housing price data

Parameter (effect)	Normal			Student-t		
	Estimate	SE	z-value	Estimate	SE	z-value
β_1 (CRIME)	−0.0072	0.0053	−1.3602	−0.0065	0.0054	−1.2061
β_2 (ROOM)	−0.0010	0.0099	−0.1023	−0.0021	0.0101	−0.2105
β_3 (AGE)	0.0009	0.0043	0.2113	0.0013	0.0043	0.2988
β_4 (DIST)	0.0786	0.4647	0.1692	0.1712	0.4517	0.3791
β_5 (BLACK)	0.4503	0.5947	0.7572	0.4513	0.5993	0.7531
β_6 (LSTAT)	−0.5427	0.2211	−2.4547	−0.5668	0.2229	−2.5432
β_7 (CHAS)	−0.0352	0.3469	−0.1015	−0.0225	0.3572	−0.0631
γ (NOXSQ)	−0.0097	0.0017	−5.7188	−0.0124	0.0014	−8.6419
μ	45.5838	2.6904	16.9430	47.4301	2.4409	19.4315
σ_u^2	63.4978	30.3667		48.4985	41.4285	
σ^2	0.0290	0.0147		0.0284	0.0249	
τ	0.0464	0.0723		0.0319	0.0580	
Statistics	Value			Value		
$L(\widehat{\boldsymbol{\theta}})$	−434.1991			−427.9173		
AIC	892.3982			879.8346		

measured with error NOXSQ was not significant. In addition, from Table 5 we may notice that the approximate standard errors of the MLE of γ are lower than the one obtained by Zhong et al. (2002).

The interpretations are similar, as the LSTAT (or NOXSQ) increases (by keeping the remaining explanatory variable values fixed) is expected a decreasing of the response LMV.

For assessing the hypotheses $\mathrm{H}_0 : \tau = 0$ against $\mathrm{H}_1 : \tau > 0$ we may apply one-sided tests discussed, for instance, by Savalli et al. (2006). In particular, for large n, the asymptotic null distribution of the likelihood ratio statistics follows a mixture of chi-squared distributions, namely $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, where χ_0^2 denotes the degenerated distribution at the origin. By applying such test we obtain, respectively, the statistical values $\xi_{LR} = 6.1922$ ($p < 0.001$) and $\xi_{LR} = 4.5342$ ($p < 0.001$) under the fitted models with normal and Student-t errors. Then, one may estimate the intraclass correlation as

$$\hat{\rho} = \frac{\hat{\tau}}{\hat{\tau} + \hat{\gamma}^2\hat{\sigma}_u^2 + \hat{\sigma}^2}$$

obtaining, respectively, $\hat{\rho} = 0.570$ and $\hat{\rho} = 0.471$ for the fitted models.

Diagnostic analysis

In order to assess the goodness-of-fit of the elliptical proposed models, Figure 1 describes the normal probability plots of the estimated scores $\hat{d}_i^{[N]}$ and $\hat{d}_i^{[t]}$. We may notice from the graphs that (15)–(17) under the Student-t distribution with $\nu = 5$ degrees of freedom seems to be more appropriate to fit the housing price data then under the normal distribution.

In the sequel we will present the influence graphs from the two fitted models and under the three perturbation schemes discussed in Sect. 6 by considering the methodology of selecting the perturbation proposed by Zhu et al. (2007).

Figure 2 describes the index plots of $|\ell_{max}|$ under the case-weight perturbation scheme from the normal and Student-t elliptical models fitted to the housing price data. We may notice more sensitivity under the normal model for which district #10 appears as the most influential. Under the Student-t model district #6 appears with some influence. Similarly, from Fig. 3, the index plots of $|\ell_{max}|$ under the scale perturbation scheme point out district #10 as the most influential under the normal model and districts #4 and #6 with smaller influence under the Student-t model. Elimination of each district above does not change the inference. Particularly, district #10 that appears as the most influential has the largest number of census tracts.

Finally, in Fig. 4 one has the index plots of $|\ell_{max}|$ under the observed data perturbation scheme from the normal and Student-t elliptical models fitted to the housing

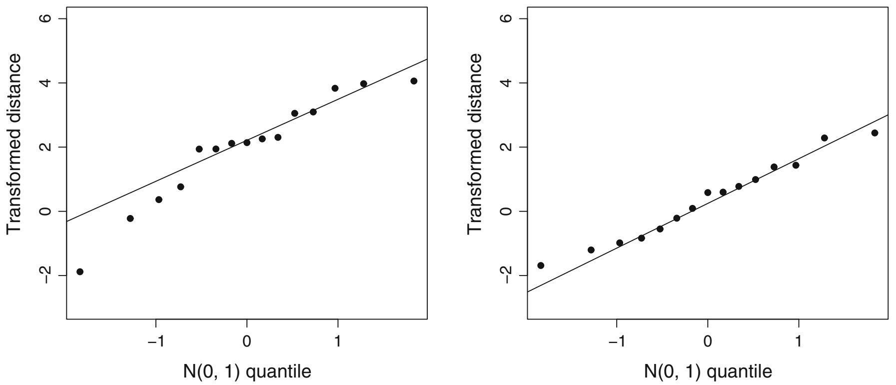

Fig. 1 Normal probability plot for the transformed distance from the normal (*left*) and Student-t (with $\nu = 5$ degrees of freedom) (*right*) elliptical models fitted to the housing price data

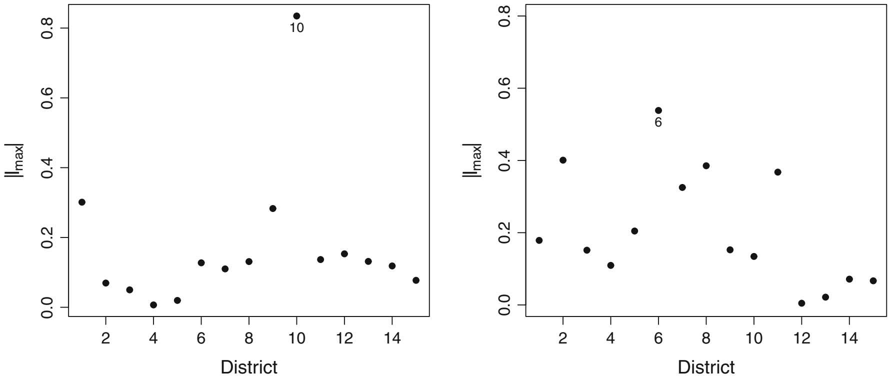

Fig. 2 Index plots of $|\ell_{max}|$ under the case-weight perturbation scheme from the normal (*left*) and Student-*t* (with $\nu = 5$ degrees of freedom) (*right*) elliptical models fitted to the housing price data

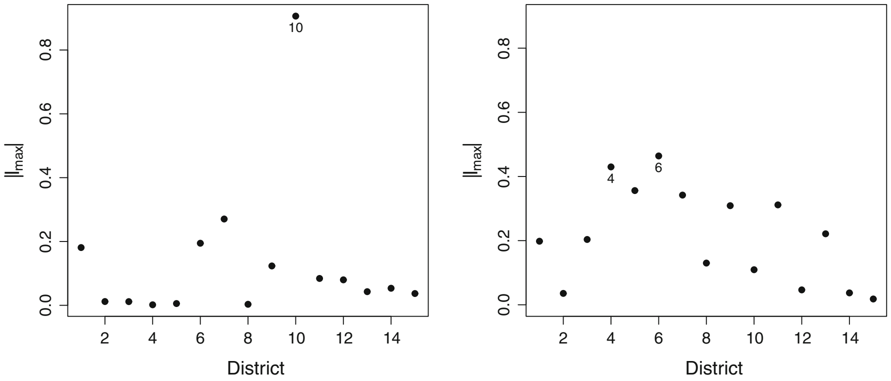

Fig. 3 Index plots of $|\ell_{max}|$ under the scale perturbation scheme from the normal (*left*) and Student-*t* (with $\nu = 5$ degrees of freedom) (*right*) elliptical models fitted to the housing price data

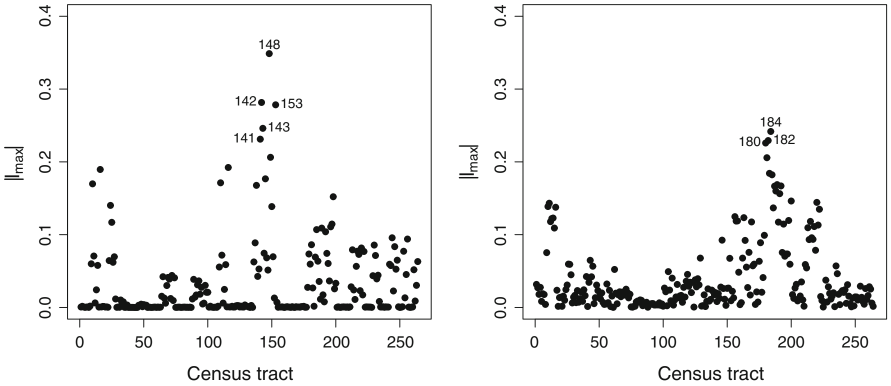

Fig. 4 Index plots of $|\ell_{max}|$ under the observed data perturbation scheme from the normal (*left*) and Student-*t* (with $\nu = 5$ degrees of freedom) (*right*) elliptical models fitted to the housing price data

price data. The census tracts #141, #142, #143, #148 and #153 are pointed out from the normal model, whereas the census tracts #180, #182 and #184 appear with smaller influence from the Student-t model.

8 Concluding remarks

In this paper we propose a linear mixed models with elliptical errors in which a covariate subject to measurement error is considered in the linear predictor. The parameter estimation and inference are performed by the marginal model obtained from the joint distribution of the observed response and observed covariate subject to measurement error. The proposed elliptical model has some interesting features, such as (i) closed-form expression for the marginal model, that is also an elliptical model, (ii) preservation in the marginal model of the mean and variance–covariance structures of the hierarchical model, (iii) possibility of kurtosis flexibility for the marginal distribution, (iv) robust aspects of the maximum likelihood estimates by down-weighting outlying observations in the sense of the Mahalanobis distance and (v) possibility of predicting the random-effects as well as the covariate subject to measurement error in a closed-form by applying the empirical Bayes estimate. We derive in closed-form expression the Fisher information matrix for the full parameter vector and a reweighed iterative process based on the Fisher scoring algorithm is developed for the parameter estimation. Some goodness-of-fit statistics for elliptical models are presented and the normal curvatures of local influence are derived under three perturbation schemes which are chosen by the methodology proposed (Zhu et al. 2007). A simulation study is performed under various parameter settings and indicates for consistency and asymptotic normality, even for small and moderate sample sizes, for the parameter estimators. Finally, a reanalyze of the housing price data set by considering a pollution covariate subject to measurement error and the methodology developed in this paper reveals that a Student-t elliptical model (with $\nu = 5$ degrees of freedom) seems to be more appropriate than the normal model.

Acknowledgements This work was supported by CNPq and FAPESP, Brazil and FONDECYT-Chile 1150325. The authors are grateful to the reviewers for their helpful comments and suggestions.

Appendix A: Derivation of the Fisher information matrix

Consider the Mahalanobis distance expressed in the following form:

$$\begin{aligned}\delta_i &= (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^\top \boldsymbol{V}_i^{-\frac{1}{2}} \boldsymbol{V}_i^{-\frac{1}{2}} (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})\\ &= \boldsymbol{r}_i^\top \boldsymbol{V}_i^{-\frac{1}{2}} \boldsymbol{V}_i^{-\frac{1}{2}} \boldsymbol{r}_i\\ &= \boldsymbol{z}_i^\top \boldsymbol{z}_i\\ &= \|\boldsymbol{z}_i\|^2,\end{aligned}$$

where $\|\boldsymbol{z}_i\|$ denotes the norm of $\boldsymbol{z}_i = \boldsymbol{\Sigma}_i^{-\frac{1}{2}} \boldsymbol{r}_i$ and $\boldsymbol{z}_i \overset{\text{ind}}{\sim} \text{El}_{m_i}(\boldsymbol{0}, \boldsymbol{I}_{m_i})$, $i = 1, \ldots, n$.

From Fang et al. (1990) one has

$$\begin{aligned}
\mathrm{E}\left\{W_g(\boldsymbol{\delta}_i)\|\boldsymbol{z}_i\|^2\right\} &= -\frac{m_i}{2},\\
\mathrm{E}\left\{W_g^2(\boldsymbol{\delta}_i)\|\boldsymbol{z}_i\|^2\right\} &= d_{g_i} \quad \text{and}\\
\mathrm{E}\left\{W_g^2(\boldsymbol{\delta}_i)\|\boldsymbol{z}_i\|^4\right\} &= f_{g_i}.
\end{aligned}$$

In particular, for the normal distribution $d_{g_i} = \frac{m_i}{2}$ and $f_{g_i} = m_i(m_i+1)$, whereas for the Student-t distribution with ν degrees of freedom these quantities become $d_{g_i} = \frac{m_i}{2}\frac{\nu+2m_i}{\nu+2m_i+2}$ and $f_{g_i} = m_i(m_i+1)\frac{\nu+2m_i}{\nu+2m_i+2}$. Expression for d_{g_i} and f_{g_i} for other elliptical distributions may be found, for instance, in Osorio et al. (2007).

Fisher information for $\boldsymbol{\gamma}$

The Fisher information for the parameter γ is defined by

$$F_{\gamma\gamma}(\boldsymbol{\theta}) = \mathrm{E}\left\{U^{\gamma}(\boldsymbol{\theta})U^{\gamma}(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n}\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\gamma}(\boldsymbol{\theta})\right\},$$

where

$$\begin{aligned}
\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\gamma}(\boldsymbol{\theta})\right\} = \frac{\boldsymbol{a}_i^2}{4}\left(\frac{f_{g_i}}{m_i(m_i+1)} - 1\right) + \frac{f_{g_i}}{2m_i(m_i+1)}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\right)\\
+2d_{g_i}\begin{pmatrix}\mu\boldsymbol{1}_{m_i}\\ \boldsymbol{0}\end{pmatrix}^{\top}\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\boldsymbol{1}_{m_i}\\ \boldsymbol{0}\end{pmatrix}, \quad i = 1,\ldots,n,
\end{aligned}$$

with $\boldsymbol{0}$ being an $m_i \times 1$ vector of zeros.

Fisher information submatrix for $\boldsymbol{\beta}$

The Fisher information submatrix for the parameter vector $\boldsymbol{\beta}$ is defined by

$$\boldsymbol{F}_{\beta\beta}(\boldsymbol{\theta}) = \mathrm{E}\left\{\boldsymbol{U}^{\beta}(\boldsymbol{\theta})\boldsymbol{U}^{\beta}(\boldsymbol{\theta})^{\top}\right\} = \sum_{i=1}^{n}\mathrm{E}\left\{\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta})\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta})^{\top}\right\},$$

where

$$\mathrm{E}\left\{\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta})\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta})^{\top}\right\} = \frac{2d_{g_i}}{m_i}\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^{\top}\boldsymbol{V}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix},$$

and $\boldsymbol{X}_i$ is an $m_i \times p$ matrix with the explanatory variable values for the ith observation, $i = 1,\ldots,n$.

Fisher information for $\boldsymbol{\mu}$

The Fisher information for the parameter μ is defined by

$$F_{\mu\mu}(\boldsymbol{\theta}) = \mathrm{E}\left\{U^{\mu}(\boldsymbol{\theta})U^{\mu}(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\mu}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})\right\},$$

where

$$\mathrm{E}\left\{U_i^{\mu}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})\right\} = \frac{2d_{g_i}}{m_i}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^{\top}\boldsymbol{V}_i^{-1}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix},\ i=1,\ldots,n.$$

Fisher information for $\boldsymbol{\sigma_u^2}$

The Fisher information for the parameter σ_u^2 is defined by

$$F_{\sigma_u^2\sigma_u^2}(\boldsymbol{\theta}) = \mathrm{E}\left\{U^{\sigma_u^2}(\boldsymbol{\theta})U^{\sigma_u^2}(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})\right\},$$

where

$$\begin{aligned}\mathrm{E}\left\{U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})\right\} &= \frac{1}{4}\,\mathrm{tr}^2\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\right)\left\{\frac{f_{g_i}}{m_i(m_i+1)}-1\right\}\\ &\quad+\frac{f_{g_i}}{2m_i(m_i+1)}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\right),\quad i=1,\ldots,n.\end{aligned}$$

Fisher information for $\boldsymbol{\sigma^2}$

The Fisher information for the parameter σ^2 is defined by

$$F_{\sigma^2\sigma^2}(\boldsymbol{\theta}) = \mathrm{E}\left\{U^{\sigma^2}(\boldsymbol{\theta})U^{\sigma^2}(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})\right\},$$

where

$$\begin{aligned}\mathrm{E}\left\{U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})\right\} &= \frac{1}{4}\,\mathrm{tr}^2\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\right)\left\{\frac{f_{g_i}}{m_i(m_i+1)}-1\right\}\\ &\quad+\frac{f_{g_i}}{2m_i(m_i+1)}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\right),\quad i=1,\ldots,n.\end{aligned}$$

Fisher information submatrix for $\boldsymbol{\tau}$

The Fisher information submatrix for the parameter vector $\boldsymbol{\tau}$ is defined by

$$\boldsymbol{F}_{\tau\tau}(\boldsymbol{\theta}) = \mathrm{E}\left\{\boldsymbol{U}^{\tau}(\boldsymbol{\theta})\boldsymbol{U}^{\tau}(\boldsymbol{\theta})^{\top}\right\},$$

with the (j, l)th element being expressed as

$$\mathrm{E}\left\{U^{\tau_j}(\boldsymbol{\theta})U^{\tau_l}(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\tau_j}(\boldsymbol{\theta})U_i^{\tau_l}(\boldsymbol{\theta})\right\},$$

where

$$\mathrm{E}\left\{U_i^{\tau_j}(\boldsymbol{\theta})U_i^{\tau_l}(\boldsymbol{\theta})\right\} = \frac{1}{4}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right)\left\{\frac{f_{g_i}}{m_i(m_i+1)} - 1\right\} + \frac{f_{g_i}}{2m_i(m_i+1)}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right), \; i = 1, \ldots, n.$$

Fisher information submatrix for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$

The Fisher information submatrix for the parameter γ and the vector $\boldsymbol{\beta}$ is defined by

$$\boldsymbol{F}_{\gamma\beta}(\boldsymbol{\theta}) = \mathrm{E}\left\{U^{\gamma}(\boldsymbol{\theta})\boldsymbol{U}^{\beta}(\boldsymbol{\theta})^{\top}\right\} = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta})^{\top}\right\},$$

where

$$\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta})^{\top}\right\} = \frac{2d_{g_i}}{m_i}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^{\top}\boldsymbol{V}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}, \; i = 1, \ldots, n.$$

Fisher information for $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$

The Fisher information for the parameters γ and μ is defined by

$$F_{\gamma\mu}(\boldsymbol{\theta}) = \mathrm{E}\left\{U^{\gamma}(\boldsymbol{\theta})U^{\mu}(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})\right\},$$

where

$$\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})\right\} = \frac{2d_{g_i}}{m_i}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^{\top}\boldsymbol{V}_i^{-1}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}, \quad i = 1, \ldots, n.$$

Fisher information for γ and σ_u^2

The Fisher information for the parameters γ and σ_u^2 is defined by

$$F_{\gamma\sigma_u^2}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})\right\},$$

where

$$\begin{aligned}\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})\right\} &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)\left\{\frac{f_{g_i}}{m_i(m_i+1)}-1\right\}\\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right), \quad i=1,\ldots,n.\end{aligned}$$

Fisher information for γ and σ^2

The Fisher information for the parameters γ and σ^2 is defined by

$$F_{\gamma\sigma^2}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})\right\},$$

where

$$\begin{aligned}\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})\right\} &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right)\left\{\frac{f_{g_i}}{m_i(m_i+1)}-1\right\}\\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right), \quad i=1,\ldots,n.\end{aligned}$$

Fisher information submatrix for γ and $\boldsymbol{\tau}$

The Fisher information submatrix for the parameters γ and $\boldsymbol{\tau}$ is defined by

$$\boldsymbol{F}_{\gamma\tau}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})\boldsymbol{U}_i^{\tau}(\boldsymbol{\theta})^{\top}\right\},$$

where $\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})\boldsymbol{U}_i^{\tau}(\boldsymbol{\theta})^{\top}\right\}$ is an $1 \times (q+s)$ vector whose jth element is given by

$$\begin{aligned}\mathrm{E}\left\{U_i^{\gamma}(\boldsymbol{\theta})U_i^{\tau_j}(\boldsymbol{\theta})\right\} &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\left\{\frac{f_{g_i}}{m_i(m_i+1)}-1\right\}\\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right), \quad i=1,\ldots,n.\end{aligned}$$

Fisher information submatrix for $\boldsymbol{\beta}$ and μ

The Fisher information submatrix for the parameters $\boldsymbol{\beta}$ and μ is defined by

$$\boldsymbol{F}_{\beta\mu}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta}) U_i^{\mu}(\boldsymbol{\theta})\right\},$$

where

$$\mathrm{E}\left\{\boldsymbol{U}_i^{\beta}(\boldsymbol{\theta}) U_i^{\mu}(\boldsymbol{\theta})\right\} = \frac{2d_{g_i}}{m_i} \begin{pmatrix} \boldsymbol{X}_i \\ \boldsymbol{0} \end{pmatrix}^{\top} \boldsymbol{V}_i^{-1} \begin{pmatrix} \gamma \boldsymbol{1}_{m_i} \\ \boldsymbol{1}_{m_i} \end{pmatrix}, \quad i = 1, \ldots, n.$$

Fisher information for σ_u^2 and σ^2

The Fisher information for the parameters σ_u^2 and σ^2 is defined by

$$F_{\sigma_u^2 \sigma^2}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\sigma_u^2}(\boldsymbol{\theta}) U_i^{\sigma^2}(\boldsymbol{\theta})\right\},$$

where

$$\begin{aligned} \mathrm{E}\left\{U_i^{\sigma_u^2}(\boldsymbol{\theta}) U_i^{\sigma^2}(\boldsymbol{\theta})\right\} &= \frac{1}{4} \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) \left\{\frac{f_{g_i}}{m_i(m_i+1)} - 1\right\} \\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)} \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2} \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right), \quad i = 1, \ldots, n. \end{aligned}$$

Fisher information submatrix for σ_u^2 and $\boldsymbol{\tau}$

The Fisher information submatrix for the parameter σ_u^2 and the vector $\boldsymbol{\tau}$ is defined by

$$\boldsymbol{F}_{\sigma_u^2 \tau}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\sigma_u^2}(\boldsymbol{\theta}) \boldsymbol{U}_i^{\tau}(\boldsymbol{\theta})^{\top}\right\},$$

where $\mathrm{E}\{U_i^{\sigma_u^2}(\boldsymbol{\theta}) \boldsymbol{U}_i^{\tau}(\boldsymbol{\theta})^{\top}\}$ is an $1 \times (q+s)$ vector whose jth element is given by

$$\begin{aligned} \mathrm{E}\left\{U_i^{\sigma_u^2}(\boldsymbol{\theta}) U_i^{\tau_j}(\boldsymbol{\theta})\right\} &= \frac{1}{4} \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right) \left\{\frac{f_{g_i}}{m_i(m_i+1)} - 1\right\} \\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)} \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2} \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right), \quad i = 1, \ldots, n. \end{aligned}$$

Fisher information submatrix for $\boldsymbol{\sigma}^2$ and $\boldsymbol{\tau}$

The Fisher information submatrix for the parameter σ_u^2 and the vector $\boldsymbol{\tau}$ is defined by

$$\boldsymbol{F}_{\sigma^2\tau}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{U_i^{\sigma^2}(\boldsymbol{\theta})\boldsymbol{U}_i^{\tau}(\boldsymbol{\theta})^\top\right\},$$

where $\mathrm{E}\left\{U_i^{\sigma^2}(\boldsymbol{\theta})\boldsymbol{U}_i^{\tau}(\boldsymbol{\theta})^\top\right\}$ is an $1 \times (q+s)$ vector whose jth element is given by

$$\begin{aligned}\mathrm{E}\left\{U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\tau_j}(\boldsymbol{\theta})\right\} &= \frac{1}{4}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right)\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\left\{\frac{f_{g_i}}{m_i(m_i+1)} - 1\right\}\\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right), \quad i = 1, \ldots, n.\end{aligned}$$

Applying similar calculation we may show the orthogonality between $\boldsymbol{\beta}$ and $(\boldsymbol{\tau}^\top, \sigma_u^2, \sigma^2)^\top$ as well as between μ and $(\boldsymbol{\tau}^\top, \sigma_u^2, \sigma^2)^\top$.

Appendix B: Derivation of the Hessian matrix

The Hessian matrix for $\boldsymbol{\theta}$ may be expressed as

$$\ddot{\boldsymbol{L}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ddot{\boldsymbol{L}}_i(\boldsymbol{\theta}),$$

where

$$\begin{aligned}\ddot{\boldsymbol{L}}_i(\boldsymbol{\theta}) &= \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\\ &= \begin{pmatrix} \ddot{L}_i^{\gamma\gamma} & \ddot{\boldsymbol{L}}_i^{\gamma\beta} & \ddot{L}_i^{\gamma\mu} & \ddot{L}_i^{\gamma\sigma_u^2} & \ddot{L}_i^{\gamma\sigma^2} & \ddot{\boldsymbol{L}}_i^{\gamma\tau}\\ \ddot{\boldsymbol{L}}_i^{\beta\gamma} & \ddot{\boldsymbol{L}}_i^{\beta\beta} & \ddot{\boldsymbol{L}}_i^{\beta\mu} & \ddot{\boldsymbol{L}}_i^{\beta\sigma_u^2} & \ddot{\boldsymbol{L}}_i^{\beta\sigma^2} & \ddot{\boldsymbol{L}}_i^{\beta\tau}\\ \ddot{L}_i^{\mu\gamma} & \ddot{\boldsymbol{L}}_i^{\mu\beta} & \ddot{L}_i^{\mu\mu} & \ddot{L}_i^{\mu\sigma_u^2} & \ddot{L}_i^{\mu\sigma^2} & \ddot{\boldsymbol{L}}_i^{\mu\tau}\\ \ddot{L}_i^{\sigma_u^2\gamma} & \ddot{\boldsymbol{L}}_i^{\sigma_u^2\beta} & \ddot{L}_i^{\sigma_u^2\mu} & \ddot{L}_i^{\sigma_u^2\sigma_u^2} & \ddot{L}_i^{\sigma_u^2\sigma^2} & \ddot{\boldsymbol{L}}_i^{\sigma_u^2\tau}\\ \ddot{L}_i^{\sigma^2\gamma} & \ddot{\boldsymbol{L}}_i^{\sigma^2\beta} & \ddot{L}_i^{\sigma^2\mu} & \ddot{L}_i^{\sigma^2\sigma_u^2} & \ddot{L}_i^{\sigma^2\sigma^2} & \ddot{\boldsymbol{L}}_i^{\sigma^2\tau}\\ \ddot{\boldsymbol{L}}_i^{\tau\gamma} & \ddot{\boldsymbol{L}}_i^{\tau\beta} & \ddot{\boldsymbol{L}}_i^{\tau\mu} & \ddot{\boldsymbol{L}}_i^{\tau\sigma_u^2} & \ddot{\boldsymbol{L}}_i^{\tau\sigma^2} & \ddot{\boldsymbol{L}}_i^{\tau\tau} \end{pmatrix}.\end{aligned}$$

Below we derive the submatrices above evaluated at $\hat{\boldsymbol{\theta}}$. One has

$$\begin{aligned}\ddot{L}_i^{\gamma\gamma}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\gamma}\right|_{\theta=\widehat{\theta}}\\ &= \frac{1}{2}\,\mathrm{tr}\left\{\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\right\} - \frac{1}{2}\,\mathrm{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial^2 \boldsymbol{V}_i}{\partial\gamma^2}\right)\end{aligned}$$

$$
\begin{aligned}
&+ W_g'(\widehat{\delta}_i)\left\{-2\widehat{\boldsymbol{r}}_i^\top \boldsymbol{V}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \widehat{\boldsymbol{r}}_i^\top \frac{\partial \boldsymbol{V}_i^{-1}}{\partial \gamma}\widehat{\boldsymbol{r}}_i\right\}^2\\
&+ W_g(\widehat{\delta}_i)\left\{2\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^\top \widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + 4\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right\}\\
&+ W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\left\{2\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma} - \frac{\partial^2 \boldsymbol{V}_i}{\partial \gamma^2}\right\}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i,
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\boldsymbol{L}}_i^{\beta\beta}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 Ł_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}\right|_{\theta=\widehat{\theta}}\\
&= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^\top \widehat{\boldsymbol{V}}_i^{-1}\left\{2W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix} + W_g(\widehat{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\ddot{L}_i^{\mu\mu}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \mu\partial \mu}\right|_{\theta=\widehat{\theta}}\\
&= 2\begin{pmatrix}\widehat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^\top \widehat{\boldsymbol{V}}_i^{-1}\left\{2W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\widehat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix} + W_g(\widehat{\delta}_i)\begin{pmatrix}\widehat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\ddot{L}_i^{\sigma_u^2\sigma_u^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \sigma_u^2\partial \sigma_u^2}\right|_{\theta=\widehat{\theta}}\\
&= \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) + W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\\
&\quad + 2W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i,
\end{aligned}
$$

$$
\begin{aligned}
\ddot{L}_i^{\sigma^2\sigma^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \sigma^2\partial \sigma^2}\right|_{\theta=\widehat{\theta}}\\
&= \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) + W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\\
&\quad + 2W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i,
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\boldsymbol{L}}_i^{\tau\tau}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}\partial \boldsymbol{\tau}^\top}\right|_{\theta=\widehat{\theta}}\\
&= \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}^\top}\right) + W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}^\top}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\\
&\quad + 2W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^\top \widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}^\top}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i,
\end{aligned}
$$

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\gamma\beta}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\,\partial\boldsymbol{\beta}^\top}\right|_{\theta=\hat{\theta}} \\ &= \left\{2W_g'(\widehat{\boldsymbol{\delta}}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^\top \widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^\top + W_g(\widehat{\boldsymbol{\delta}}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^\top\right\}\widehat{\boldsymbol{V}}_i^{-1}\left\{2\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right. \\ &\quad\left.+\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\right\},\end{aligned}$$

$$\begin{aligned}\ddot{L}_i^{\gamma\mu}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\,\partial\mu}\right|_{\theta=\hat{\theta}} \\ &= 2W_g'(\hat{\boldsymbol{\delta}}_i)\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^\top \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\left\{2\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\} \\ &\quad +2W_g(\hat{\boldsymbol{\delta}}_i)\left\{\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^\top\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right. \\ &\quad\left.+\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},\end{aligned}$$

$$\begin{aligned}\ddot{L}_i^{\gamma\sigma_u^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\,\partial\sigma_u^2}\right|_{\theta=\hat{\theta}} \\ &= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)-\frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma\,\partial\sigma_u^2}\right) \\ &\quad -W_g'(\hat{\delta}_i)\left\{\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}\left\{2\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\} \\ &\quad -W_g(\hat{\delta}_i)\left\{2\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{G}}_i^{\sigma_u^2}\hat{\boldsymbol{r}}_i\right\},\end{aligned}$$

where $\hat{\boldsymbol{G}}_i^{\sigma_u^2} = \frac{\partial^2\boldsymbol{V}_i^{-1}}{\partial\gamma\,\partial\sigma_u^2} = \hat{\boldsymbol{V}}_i^{-1}\left(\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}-\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma\,\partial\sigma_u^2}+\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)\hat{\boldsymbol{V}}_i^{-1}$,

$$\begin{aligned}\ddot{L}_i^{\gamma\sigma^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\,\partial\sigma^2}\right|_{\theta=\hat{\theta}} \\ &= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)-W_g(\hat{\delta}_i)\left\{2\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{G}}_i^{\sigma^2}\hat{\boldsymbol{r}}_i\right\} \\ &\quad -W_g'(\hat{\delta}_i)\left\{\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}\left\{2\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},\end{aligned}$$

where $\hat{\boldsymbol{G}}_i^{\sigma^2} = \frac{\partial^2\boldsymbol{V}_i^{-1}}{\partial\gamma\,\partial\sigma^2} = \hat{\boldsymbol{V}}_i^{-1}\left(\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}+\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right)\hat{\boldsymbol{V}}_i^{-1}$,

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\gamma\tau}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\boldsymbol{\tau}}\right|_{\theta=\hat{\theta}}\\ &= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right) - W_g(\hat{\delta}_i)\left\{2\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{G}}_i^{\tau}\hat{\boldsymbol{r}}_i\right\}\\ &\quad - W_g'(\hat{\delta}_i)\left\{\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}\left\{2\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},\end{aligned}$$

where $\hat{\boldsymbol{G}}_i^{\tau} = \frac{\partial^2\boldsymbol{V}_i^{-1}}{\partial\gamma\partial\boldsymbol{\tau}} = \hat{\boldsymbol{V}}_i^{-1}\left(\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma} + \frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\right)\hat{\boldsymbol{V}}_i^{-1}$,

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\beta\mu}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\mu}\right|_{\theta=\hat{\theta}}\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^{\top}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g(\hat{\boldsymbol{\delta}}_i)\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix} + 2W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},\end{aligned}$$

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\beta\sigma_u^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma_u^2}\right|_{\theta=\hat{\theta}}\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^{\top}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,\end{aligned}$$

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\beta\sigma^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma^2}\right|_{\theta=\hat{\theta}}\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^{\top}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,\end{aligned}$$

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\beta\tau}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\tau}^{\top}}\right|_{\theta=\hat{\theta}}\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^{\top}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^{\top}}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,\end{aligned}$$

$$\begin{aligned}\ddot{L}_i^{\mu\sigma_u^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\mu\partial\sigma_u^2}\right|_{\theta=\hat{\theta}}\\ &= 2\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^{\top}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,\end{aligned}$$

$$\begin{aligned}\ddot{L}_i^{\mu\sigma^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\mu\partial\sigma^2}\right|_{\theta=\hat{\theta}}\\ &= 2\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^{\top}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,\end{aligned}$$

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\mu\tau}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\mu\partial\boldsymbol{\tau}^\top}\right|_{\theta=\hat{\theta}}\\ &= 2\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^\top \hat{\boldsymbol{V}}_i^{-1}\left\{W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^\top}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,\end{aligned}$$

$$\begin{aligned}\ddot{L}_i^{\sigma_u^2\sigma^2}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\sigma^2}\right|_{\theta=\hat{\theta}}\\ &= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right)\\ &\quad+\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1} + 2W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,\end{aligned}$$

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\sigma_u^2\tau}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\boldsymbol{\tau}^\top}\right|_{\theta=\hat{\theta}}\\ &= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^\top}\right)\\ &\quad+\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1} + 2W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^\top}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\end{aligned}$$

and

$$\begin{aligned}\ddot{\boldsymbol{L}}_i^{\sigma^2\tau}(\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma^2\partial\boldsymbol{\tau}^\top}\right|_{\theta=\hat{\theta}}\\ &= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^\top}\right)\\ &\quad+\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\left\{W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1} + 2W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}\right\}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^\top}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.\end{aligned}$$

Appendix C: Derivation of $\boldsymbol{\Delta}$ matrix

Case-weight perturbation

In particular for the normal case we obtain

$$\begin{aligned}\boldsymbol{\Delta}_\gamma &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\gamma\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0}\\ &= -\frac{\sqrt{m_i}}{2}\left[\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right) - \hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\left\{2\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}\right],\end{aligned}$$

$$\boldsymbol{\Delta}_\beta = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = \sqrt{m_i}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^\top\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

$$\boldsymbol{\Delta}_{\mu} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \mu \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = \sqrt{m_i}\begin{pmatrix} \hat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^{\top} \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

$$\boldsymbol{\Delta}_{\sigma_u^2} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \sigma_u^2 \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\frac{\sqrt{m_i}}{2}\left\{\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) - \hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},$$

$$\boldsymbol{\Delta}_{\sigma^2} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \sigma^2 \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\frac{\sqrt{m_i}}{2}\left\{\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) - \hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}$$

and

$$\boldsymbol{\Delta}_{\tau_j} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_j \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\frac{\sqrt{m_i}}{2}\left\{\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right) - \hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},$$

where $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ and $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$, for $j = 1, \ldots, q+s$.

For the Student-t case we obtain

$$\begin{aligned}\boldsymbol{\Delta}_{\gamma} &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \gamma \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} \\ &= -\frac{p_i}{2}\left[\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) - v(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\left\{2\begin{pmatrix} \hat{\mu}\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix} + \frac{\partial \boldsymbol{V}_i}{\partial \gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}\right],\end{aligned}$$

$$\boldsymbol{\Delta}_{\beta} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = p_i v(\hat{\delta}_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^{\top} \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

$$\boldsymbol{\Delta}_{\mu} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \mu \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = p_i v(\hat{\delta}_i)\begin{pmatrix} \hat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^{\top} \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

$$\boldsymbol{\Delta}_{\sigma^2} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \sigma_u^2 \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\frac{p_i}{2}\left\{\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) - v(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},$$

$$\boldsymbol{\Delta}_{\sigma^2} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \sigma^2 \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\frac{p_i}{2}\left\{\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) - v(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}$$

and

$$\boldsymbol{\Delta}_{\tau_j} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_j \partial \omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\frac{p_i}{2}\left\{\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right) - v(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\},$$

where $p_i = \frac{2m_i+\nu}{2}\sqrt{\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+2m_i}{2}\right)}$, $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^\top \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ and $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$, fo $j = 1, \ldots, q+s$.

Scale matrix perturbation

In general we obtain

$$\begin{aligned}
\boldsymbol{\Delta}_\gamma &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\gamma\,\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} \\
&= \sqrt{f_{g_i}}\left\{W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right\}\left\{2\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^\top - \hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right\}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i, \\
\boldsymbol{\Delta}_\beta &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\,\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -2\sqrt{f_{g_i}}\left\{W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right\}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^\top\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i, \\
\boldsymbol{\Delta}_\mu &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\mu\,\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -2\sqrt{f_{g_i}}\left\{W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right\}\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^\top\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i, \\
\boldsymbol{\Delta}_{\sigma_u^2} &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma_u^2\,\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\sqrt{f_{g_i}}\left\{W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right\}\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i; \\
\boldsymbol{\Delta}_{\sigma^2} &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma^2\,\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\sqrt{f_{g_i}}\left\{W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right\}\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i
\end{aligned}$$

and

$$\boldsymbol{\Delta}_{\tau_j} = \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\tau_j\,\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\sqrt{f_{g_i}}\left\{W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right\}\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\tau_j}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

where $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ and $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$, for $j = 1, \ldots, q+s$.

Observed response perturbation

In general we obtain

$$\begin{aligned}
\boldsymbol{\Delta}_\gamma &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\gamma\,\partial\boldsymbol{\omega}_i^\top}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} \\
&= -\frac{2}{\sqrt{2d_{g_i}/m_i}}\Bigg[W_g(\hat{\delta}_i)\left\{\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^\top\hat{\boldsymbol{V}}_i^{-\frac{1}{2}} + \hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\left(\frac{\partial\boldsymbol{V}_i^{\frac{1}{2}}}{\partial\gamma} - \frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\right)\hat{\boldsymbol{V}}_i^{\frac{1}{2}}\right\} \\
&\qquad + W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-1}\times\left\{\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - \frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right\}\hat{\boldsymbol{r}}_i^\top\hat{\boldsymbol{V}}_i^{-\frac{1}{2}}\Bigg],
\end{aligned}$$

$$
\begin{aligned}
\mathbf{\Delta}_{\beta} &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}_i^{\top}}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0}\\
&= -\frac{2}{\sqrt{2d_{g_i}/m_i}}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^{\top}\left\{W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}+2W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\right\}\hat{\boldsymbol{V}}_i^{\frac{1}{2}}\hat{\boldsymbol{V}}_i^{-1},\\
\mathbf{\Delta}_{\mu} &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\mu\partial\boldsymbol{\omega}_i^{\top}}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0}\\
&= -\frac{2}{\sqrt{2d_{g_i}/m_i}}\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^{\top}\left\{W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}+2W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\right\}\hat{\boldsymbol{V}}_i^{\frac{1}{2}}\hat{\boldsymbol{V}}_i^{-1},\\
\mathbf{\Delta}_{\sigma_u^2} &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma_u^2\partial\boldsymbol{\omega}_i^{\top}}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0}\\
&= \frac{2}{\sqrt{2d_{g_i}/m_i}}\left\{W_g(\hat{\delta}_i)\left(\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i^{\frac{1}{2}}}{\partial\sigma_u^2}-\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-\frac{1}{2}}\right)\right.\\
&\quad\left.-W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-\frac{1}{2}}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{\Delta}_{\sigma 2} - \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma^2\partial\boldsymbol{\omega}_i^{\top}}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0}&\\
= \frac{2}{\sqrt{2d_{g_i}/m_i}}\left\{W_g(\hat{\delta}_i)\left(\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i^{1/2}}{\partial\sigma^2}-\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-\frac{1}{2}}\right)\right.&\\
\left.-W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-\frac{1}{2}}\right\}&
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{\Delta}_{\tau_j} &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\tau_j\partial\boldsymbol{\omega}_i^{\top}}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0}\\
&= \frac{2}{\sqrt{2d_{g_i}/m_i}}\left\{W_g(\hat{\delta}_i)\left(\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i^{\frac{1}{2}}}{\partial\tau_j}-\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\tau_j}\hat{\boldsymbol{V}}_i^{-\frac{1}{2}}\right)\right.\\
&\quad\left.-W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\tau_j}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-\frac{1}{2}}\right\},
\end{aligned}
$$

where $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^{\top}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ and $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$, for $j = 1, \ldots, q + s$.

References

Akaike H (1974) A new look at statistical models identification. IEEE Trans Autom Control 19:716–722

Arellano-Valle RB (2010) On the information matrix of the multivariate skew-t model. METRON Int J Stat 68:371–386

Belsley DA, Kuh E, Welsch RE (1980) Regression diagnostics: identifying influence data and sources of collinearity. Wiley, New York

Buonaccorsi JP (2010) Measurement error: models, methods, and applications. Chapman and Hall, London

Cao CZ, Lin JG, Zhu LX (2010) Heteroscedasticity and/or autocorrelation diagnostics in nonlinear models with AR(1) and symmetrical errors. Stat Pap 51:813–836

Cook RD (1986) Assessment of local influence (with discussion). J R Stat Soc B 48:133–169

Cox D, Reid N (1987) Parameter orthogonality and approximate conditional inference. J R Stat Soc B 49:1–39

Cui H, Ng KW, Zhu L (2004) Estimation in mixed effects model with errors in variables. J Multivar Anal 91:53–73

Demidenko E (2004) Mixed models: theory and applications. Wiley, Hoboken

Fang KT, Kotz S, Ng KW (1990) Symmetric multivariate and related distributions. Chapman & Hall, London

Fuller WA (2006) Measurement error models. Wiley, New York

Galea M, Giménez P (2016) Local influence diagnostics for the test of man-variance efficiency and systematic risks in the capital asset pricing model. Stat Pap. doi:10.1007/s00362-016-0838-8

Galea M, Paula GA, Uribe-Opazo M (2003) On influence diagnostic in univariate elliptical linear regression models. Stat Pap 44:23–45

Härdle WK, Simar L (2012) Applied multivariate statistical analysis, 3rd edn. Springer, New York

Harrison D, Rubinfeld DL (1978) Hedonic prices and the demand for clean air. J Environ Econ Manag 5:81–102

Ibacache-Pulgar G, Paula GA, Galea M (2012) Influence diagnostics for elliptical semiparametric mixed models. Stat Model 12:165–170

Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, 2nd edn. Wiley, New York

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963–974

Lange KL, Little RJ, Taylor JMG (1989) Robust statistical modeling using the t distribution. J Am Stat Assoc 84:881–896

Lemonte AJ, Patriota AG (2011) Multivariate elliptical models with general parametrization. Stat Methodol 8:389–400

Lindsey JK (1999) Multivariate elliptically contoured distributions for repeated measurements. Biometrics 55:1277–1280

Longford NT (1993) Random coefficient models. Oxford University Press, New York

Nakamura T (1990) Corrected score function for errors-in-variables models: methodology and application to generalized linear models. Biometrika 77:127–137

Osorio F, Paula GA, Galea M (2007) Assessment of local influence in elliptical linear models with longitudinal structure. Comput Stat Data Anal 51:4354–4368

Pinheiro JC, Liu C, Wu YN (2001) Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t-distribution. J Comput Graph Stat 10:249–276

Poon W, Poon YS (1999) Conformal normal curvature and assessment of local influence. J R Stat Soc B 61:51–61

Relvas CEM, Paula GA (2016) Partially linear models with first-order autoregressive symmetric errors. Stat Pap 57:795–825

Riquelme M, Bolfarine H, Galea M (2015) Robust linear functional mixed models. J Multivar Anal 132:82–98

Russo C, Paula GA, Aoki R (2009) Influence diagnostics in nonlinear mixed-effects elliptical models. Comput Stat Data Anal 53:4143–4156

Savalli C, Paula GA, Cysneiros FJA (2006) Assessment of variance components in elliptical linear mixed models. Stat Model 6:59–76

Staudenmayer J, Lake EE, Wand MP (2009) Robustness for general design mixed models using the t-distribution. Stat Model 9:235–255

Wang N, Heckman N (2009) Identifiability in linear mixed models. Technical report. Department of Statistics, University of British Columbia, Vancouver

Welsh AH, Richardson AM (1997) Approaches of the robust estimation of mixed models. In: Maddala GS, Rao CR (eds) Handbook of statistics, vol 15. Elsevier, Amsterdam, pp 343–384

Wu L (2010) Mixed effects models for complex data. Chapman and Hall, New York

Zare K, Rasekh A (2011) Diagnostic measures for linear mixed measurement error models. SORT 35:125–144

Zhang D, Lin X, Raz J, Sowers M (1998) Semiparametric stochastic mixed models for longitudinal data. J Am Stat Assoc 93:710–719

Zhong XP, Fung WK, Wei BC (2002) Estimation in linear models with random effects and errors-in-variables. Ann Inst Stat Math 54:595–606

Zhu H, Ibrahim JG, Lee S, Zhang H (2007) Perturbation selection and influence measures in local influence analysis. Ann Stat 35:2565–2588