Universidade Estadual de Campinas
Instituto de Matemática, Estatística e
Computação Científica - IMECC
Departamento de Estatística

MODELOS LINEARES E NÃO LINEARES DE EFEITOS MISTOS PARA RESPOSTAS CENSURADAS USANDO AS DISTRIBUIÇÕES NORMAL E T-STUDENT MULTIVARIADAS

LINEAR AND NONLINEAR MIXED-EFFECTS MODELS WITH CENSORED RESPONSE USING THE MULTIVARIATE NORMAL AND STUDENT-T DISTRIBUTIONS

Dissertação de Mestrado

Larissa Avila Matos

Orientador: **Prof. Dr. Víctor Hugo Lachos Dávila**

CAMPINAS

2012

† Este trabalho contou com apoio financeiro da CAPES

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e
Computação Científica - IMECC

Larissa Avila Matos

Modelos lineares e não lineares de efeitos mistos para respostas censuradas usando as distribuições normal e t-Student multivariadas

Linear and nonlinear mixed-effects models with censored response using the multivariate normal and Student-t distributions

Dissertação de Mestrado apresentada ao Instituto de Matemática, Estatística e Computação Científica da UNICAMP para obtenção do título de Mestre em Estatística.

Master's thesis presented to the Institute of Mathematics, Statistics and Computer Science at UNICAMP for the obtaining the title of Master in Statistics.

Orientador: Prof. Dr. Víctor Hugo Lachos Dávila

Este exemplar corresponde à versão final da dissertação defendida pela aluna Larissa Avila Matos, e orientada pelo Prof. Dr. Víctor Hugo Lachos Dávila.

Víctor Hugo Lachos Dávila
Orientador

Campinas, 2012

FICHA CATALOGRÁFICA ELABORADA POR
MARIA FABIANA BEZERRA MULLER - CRB8/6162
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

Matos, Larissa Avila, 1987-
M428m Modelos lineares e não lineares de efeitos mistos para respostas censuradas usando as distribuições normal e t-Student multivariadas / Larissa Avila Matos. – Campinas, SP : [s.n.], 2012.

Orientador: Víctor Hugo Lachos Dávila.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

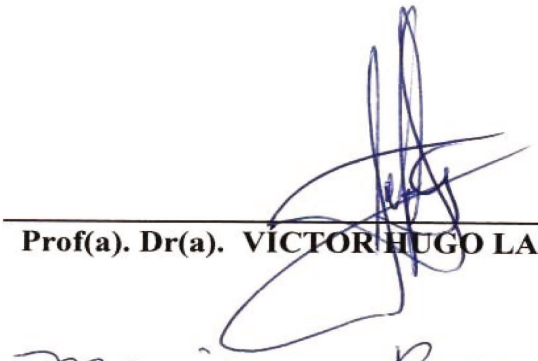
1. Modelos lineares (Estatística). 2. Modelos não lineares (Estatística). 3. Estimativa de parâmetro. I. Lachos Dávila, Víctor Hugo, 1973-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

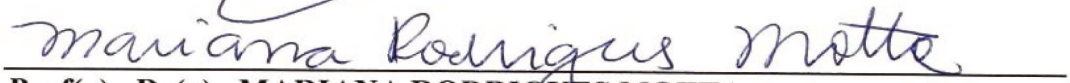
Título em inglês: Linear and nonlinear mixed-effects models with censored response using the multivariate normal and Student-t distributions
Palavras-chave em inglês:
Linear models (Statistics)
Nonlinear models (Statistics)
Parameter estimation
Área de concentração: Estatística
Titulação: Mestre em Estatística
Banca examinadora:
Víctor Hugo Lachos Dávila [Orientador]
Mariana Rodrigues Motta
Celso Rômulo Barbosa Cabral
Data de defesa: 24-02-2012
Programa de Pós-Graduação: Estatística

Dissertação de Mestrado defendida em 24 de fevereiro de 2012 e aprovada

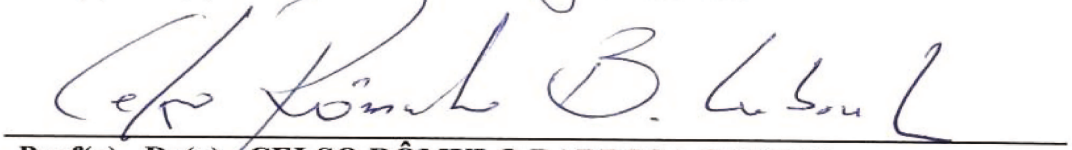
Pela Banca Examinadora composta pelos Profs. Drs.



Prof(a). Dr(a). VICTOR HUGO LACHOS DÁVILA



Prof(a). Dr(a). MARIANA RODRIGUES MOTTA



Prof(a). Dr(a). CELSO RÔMULO BARBOSA CABRAL

Para Likinha

Agradecimentos

Primeiramente, agradeço a Deus, por me ofercer saúde, disposição, discernimento e por colocar várias pessoas maravillhosas em minha vida, além de me fornecer inúmeras oportunidades. A Ele só tenho de agradecer por tudo.

Agradeço às pessoas que, com seu tempo, conhecimento, talento, amizade e/ou paciência, me ajudaram a chegar até aqui. Não citarei nomes, pois aqueles que me conhecem sabem que a probabilidade de eu esquecer alguém é de 100%.

Queridos pais, esta vitória dedico a vocês, que me apoiaram, confiaram, amaram e que me deram carinho e esperança nos piores momentos, muito obrigada, amo vocês. À minha irmã, companheira e cúmplice, minha família, meus avós, tios, tias, primos e primas, muito obrigada.

Agradeço ao meu orientador, pelo apoio, suporte e paciência ao longo dessa tese. Aos meus amigos, que deram importantes sugestões para a minha pesquisa e merecem ser lembrados por isso, muito obrigada por serem grandes amigos mesmo, e tolerantes, já que a coisa toda de terminar a tese me tornou distante.

A outros amigos, ainda que não tenham contribuído em nada para esta tese, eu tenho que agradecer, simplesmente por terem feito com que eu saísse de casa, conversasse e tivesse uma vida social.

Aos professores do Departamento de Estatística da Universidade Estadual de Campinas e a outros professores que me ajudaram. Agradeco à CAPES pela bolsa que permitiu que eu me dedicasse à minha pesquisa. Agradeço a todos que me ajudaram e me amaram ao longo dos anos.

Obrigada a todos vocês.

"All models are wrong,
but some are useful."
George E. P. Box

Resumo

Modelos mistos são geralmente usados para representar dados longitudinais ou de medidas repetidas. Uma complicação adicional surge quando a resposta é censurada, por exemplo, devido aos limites de quantificação do ensaio utilizado. Distribuições normais para os efeitos aleatórios e os erros residuais são geralmente assumidas, mas tais pressupostos fazem as inferências vulneráveis, à presença de outliers. Motivados por uma preocupação da sensibilidade para potenciais outliers ou dados com caudas mais pesadas do que a normal, pretendemos desenvolver nessa dissertação, inferência para modelos lineares e não lineares de efeito misto censurados (NLMEC / LMEC) com base na distribuição t- Student multivariada, sendo uma alternativa flexível ao uso da distribuição normal correspondente. Propomos um algoritmo ECM para computar as estimativas de máxima verossimilhança para os NLMEC / LMEC. Este algoritmo utiliza expressões fechadas no passo-E, que se baseia em fórmulas para a média e a variância de uma distribuição t-multivariada truncada. O algoritmo proposto é implementado, pacote tlmec do R. Também propomos aqui um algoritmo ECM exato para os modelos lineares e não lineares de efeito misto censurados, com base na distribuição normal multivariada, que nos permite desenvolver análise de influência local para modelos de efeito misto com base na esperança condicional da função log-verossilhança dos dados completos. Os procedimentos desenvolvidos são ilustrados com a análise longitudinal da carga viral do HIV, apresentada em dois estudos recentes sobre a AIDS.

Abstract

Mixed models are commonly used to represent longitudinal or repeated measures data. An additional complication arises when the response is censored, for example, due to limits of quantification of the assay used. Normal distributions for random effects and residual errors are usually assumed, but such assumptions make inferences vulnerable to the presence of outliers. Motivated by a concern of sensitivity to potential outliers or data with tails longer-than-normal, we aim to develop in this dissertation inference for linear and nonlinear mixed effects models with censored response (NLMEC/LMEC) based on the multivariate Student-t distribution, being a flexible alternative to the use of the corresponding normal distribution. We propose an ECM algorithm for computing the maximum likelihood estimates for NLMEC/LMEC. This algorithm uses closed-form expressions at the E-step, which relies on formulas for the mean and variance of a truncated multivariate-t distribution. The proposed algorithm is implemented in the R package tlmec. We also propose here an exact ECM algorithm for linear and nonlinear mixed effects models with censored response based on the multivariate normal distribution, which enable us to developed local influence analysis for mixed effects models on the basis of the conditional expectation of the complete-data log-likelihood function. The developed procedures are illustrated with two case studies, involving the analysis of longitudinal HIV viral load in two recent AIDS studies.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

Modelos lineares e não-lineares mistos (LME/NLME) são frequentemente usados para análise de dados agrupados, pois oferecem uma flexibilidade em modelar a correlação entre e intra unidades amostrais, usualmente presente nesses tipos de dados (Pinheiro and Bates, 2000). Exemplos de dados agrupados incluem dados de medidas repetidas, dados multiníveis e dados longitudinais (entre outros). Entretanto, em vários estudos longitudinais, como estudos sobre a poluição ambiental e doenças infecciosas, a medição de algumas variáveis pode ser sujeita a um limite de quantificação, isto é, um certo limite abaixo ou acima em que a medição não é quantificada. Por exemplo, a carga viral mede a quantidade de atividade de reprodução dos vírus e, dependendo do ensaio do diagnóstico usado, essas medidas podem ser subjetivas a um limite de detecção superior ou inferior (por isso, censurados à direita e à esquerda), valores acima ou abaixo em que eles não são quantificados. A proporção de dados censurados nestes estudos pode não ser trivial e, considerando métodos ad-hoc, isto é, substituindo o valor limite ou algum ponto arbitrário como o ponto médio entre zero e o corte para detecção (Vaida and Liu, 2009), pode conduzir a estimativas tendenciosas para os efeitos fixos e para os componentes da variância (Wu, 2010). Como alternativa para métodos ad-hoc, Hughes (1999) propôs uma verossimilhança baseada no algoritmo EM de Monte Carlo

(MCEM) para LME com respostas censuradas (LMEC). Vaida et al. (2007) propôs um algoritmo EM híbrido (HEM) para modelos lineares e não-lineares mistos com respostas censuradas (LMEC/NLMEC) usando uma implementação mais eficiente do algoritmo de Hughes, baseado num eficiente esquema de amostragem em blocos. Vaida and Liu (2009) propôs um algoritmo EM exato para LMEC/ NLMEC que usa expressões em forma fechada no passo E, em oposição a simulação de Monte Carlo, levando a um avanço na velocidade de compilação.

No contexto de LMEC/ NLMEC, os efeitos aleatórios e os erros entre unidades amostrais, por conveniência matemática, são frequentemente considerados ter uma distribuição normal. Contudo, tais suposições de normalidade podem nem sempre ser realísticas porque são vulneráveis à presença de uma observação atípica. Para lidar com o problema de uma observação atípica no LME com respostas completas, algumas proposições foram feitas na literatura como substituindo a suposição de normalidade por uma classe de distribuições mais flexíveis. Por exemplo, Pinheiro et al. (2001) propôs um modelo linear com efeito misto considerando a distribuição t multivariada (t-LME) e demonstrou a robustez contra outliers através de simulações e uma aplicação com dados ortodônticos. Lin and Lee (2007) desenvolveu algumas ferramentas adicionais para t-LME através de uma perspectiva Bayesiana. Rosa et al. (2003) defende o uso de subclasses de distribuições elípticas, chamada distribuição normal/independente (NI) (Liu, 1996) e adotou o contexto Bayesiano para análises posteriores para LMEC/NLMEC de cauda pesada. Mais elaborações no t-LME foram estudadas por Song et al. (2007) e Wang and Fan (2011). Mais recentemente, no contexto de LMEC/NLMEC de cauda pesada, Lachos et al. (2011) defende o uso de classe de distribuição NI e adotou o contexto Bayesiano para análises posteriores. Apesar de alguns trabalhos com distribuição elíptica terem aparecido recentemente na literatura, não há estudos em LMEC/NLMEC sob a distribuição t de Student na perspectiva frequentista.

Motivados por isso, neste trabalho, primeiramente propomos uma modificação no algoritmo EM proposto por Vaida and Liu (2009), em que todos os parâmetros são atualizados (passo M) considerando os efeitos aleatórios e as observações censuradas como dados perdidos. Depois, propomos uma modelagem paramétrica robusta nos LMEC/NLMEC baseada na distribuição t-multivariada para que a t-LMEC/t-NLMEC seja definida e uma abordagem totalmente baseada na verossimilhança seja considerada, incluindo a implementação de um algoritmo ECM exato para as estimativas de máxima

verossimilhança (ML). Tendo ainda por base a obra de Vaida and Liu (2009), neste trabalho também desenvolvemos e apresentamos uma análise de diagnóstico em modelos lineares e não lineares mistos para resposta censuradas, usando a distribuição normal.

Baseados no que foi discutido aqui, mostramos uma breve descrição do algoritmo EM, que será usado para encontrar as estimativas de máxima verossimilhança dos parâmetros nos LMEC/NLMEC para as distribuições normal e t de Student. Também apresentamos uma breve descrição de análise de influência aplicada nos LMEC/NLMEC para distribuição normal e, finalmente, descrevemos os objetivos e a organização deste trabalho.

1.2 O algoritmo EM

O algoritmo EM (Dempster et al., 1977) é um procedimento iterativo eficiente para calcular as estimativas de máxima verossimilhança (ML) na presença de dados faltantes. Na estimação pelo método de máxima verossimilhança, desejamos estimar os parâmetros do modelo para o qual os dados observados sejam mais prováveis. Esse algoritmo é aplicado em problemas de estimação para dados incompletos, aumentando o vetor de dados observados ($\mathbf{y}_{obs}$) com a inclusão de variáveis latentes ($\mathbf{y}_{nobs}$), que não são diretamente observadas, obtendo-se, assim, o vetor de dados completos $\mathbf{y}_c = (\mathbf{y}_{obs}, \mathbf{y}_{nobs})$. A função de log-verossimilhança é representada por $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \log(f(\mathbf{y}_c|\boldsymbol{\theta}))$ e cada iteração do algoritmo EM consiste em dois passos:

- **Passo E (Esperança):**
  Este passo consiste em calcular a esperança da log-verossimilhança completa, denotada por $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(i-1)})$, condicionada no vetor de dados observados. Isto é, para a i-ésima iteração temos que, dado $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}^{(i-1)}$,

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(i-1)}) = E\{\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_{obs}, \widehat{\boldsymbol{\theta}}^{(i-1)}\};$$

- **Passo M (Maximização):**
  Consiste em maximizar a log-verossimilhança completa em relação aos parâmetros do modelo, substituindo os dados latentes por seus valores esperados condicionais obtidos no passo E. Para a i-ésima iteração, obtemos $\widehat{\boldsymbol{\theta}}^{(i)}$ que maximiza $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(i-1)})$, tal que,

$$Q(\widehat{\boldsymbol{\theta}}^{(i)}|\widehat{\boldsymbol{\theta}}^{(i-1)}) > Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(i-1)}), \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Esse procedimento é repetido até que uma certa margem envolvendo duas avaliações sucessivas da log-verossimilhança $\ell(\boldsymbol{\theta}|\mathbf{y})$, como $|\ell(\widehat{\boldsymbol{\theta}}^{(i)}) - \ell(\widehat{\boldsymbol{\theta}}^{(i-1)})|$ ou $|\ell(\widehat{\boldsymbol{\theta}}^{(i)})/\ell(\widehat{\boldsymbol{\theta}}^{(i-1)}) - 1|$, seja pequena o suficiente.

Quando o passo M do algoritmo é difícil de implementar, é comum substituir este com uma sequência de passos de maximização restrita (CM), cada uma das quais maximiza $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(i-1)})$ sob $\boldsymbol{\theta}$ com alguma função de $\boldsymbol{\theta}$ mantida fixa. Isto leva a uma simples extensão do algoritmo EM, chamado de algoritmo ECM (Meng and Rubin, 1993).

1.3 Análise de diagnóstico

Os modelos estatísticos são ferramentas importantes para extrair e compreender características essenciais de um conjunto de dados. Uma etapa importante na análise é a verificação de possíveis afastamentos das suposições feitas no modelo, como, por exemplo, a existência de observações extremas com alguma interferência nos resultados do ajuste. Os elementos do conjunto de dados, que efetivamente controlam aspectos da análise, são ditos *influentes* se eles produzem alterações no resultado da análise quando excluídos ou submetidos a algum tipo de perturbação.

Existem duas abordagens principais para a detecção de observações influentes. A primeira abordagem é o método de eliminação de casos (Cook (1977)), é um método intuitivamente atraente (ver também Cook and Weisberg (1982)), onde o impacto de se excluir uma observação na previsão é diretamente avaliada por medidas, tais como afastamento pela verossimilhança e a distância de Cook. A segunda abordagem, que é uma técnica geral estatística, utilizada para avaliar a estabilidade das estimativas com relação ao modelo teórico, é a abordagem de influência local Cook (1986). Após o trabalho pioneiro de Cook (1986), esse método tem, recentemente, recebido considerável

atenção na literatura estatística de modelos de efeito misto (LME/NLME); veja, por exemplo, Lesaffre and Verbeke (1998), Zhu and Lee (2001), Lee and Xu (2004), Osorio et al. (2007), Russo et al. (2009), entre outros.

Embora vários estudos de diagnósticos em LME/NLME tenham aparecido na literatura, nenhum estudo foi feito para diagnóstico de influência em NLMEC/LMEC e na análise de influência local. A principal dificuldade deve-se ao fato de que função log-verossimilhança observada dos NLMEC/LMEC envolve integrais intratáveis (por exemplo, a função de densidade de probabilidade da distribuição multinormal truncada), tornando a aplicação direta da abordagem de Cook (Cook, 1986) para este modelo muito difícil, já que as medidas envolvem as derivadas parciais de primeira e segunda ordem dessa função. Zhu and Lee (2001) desenvolveu uma abordagem para a realização da análise de influência local em modelos estatísticos gerais com dados ausentes. Este é baseado na função Q-afastamento, que está relacionada com a esperança condicional da log-verossimilhança dos dados completos no passo E do algoritmo EM. Essa abordagem produz resultados muito semelhantes aos obtidos a partir do método de Cook. Além disso, o método de eliminação de casos pode ser estudado pela função Q-afastamento seguindo a abordagem proposta por Zhu and Lee (2001).

1.4 Objetivos e organização da pesquisa

Neste trabalho, pretendemos fazer um estudo de inferência estatística em modelos lineares e não lineares de efeito misto para dados censurados usando as distribuições normal e t de Student. Além disso, temos a intenção de fazer diagnóstico de influência em LMEC/NLMEC usando a distribuição normal. Para o processo de estimação de máxima verossimilhança, usamos o algoritmo EM. Aproveitando-se da esperança condicional da função de verossimilhança completa, derivamos medidas de diagnóstico. Assim, os objetivos específicos deste trabalho podem ser resumidos como se segue:

1. Apresentar um estudo de estimação e diagnóstico, em modelos lineares e não lineares de efeito misto para dados censurados, usando a distribuição normal.

2. Apresentar um estudo de estimação, em modelos lineares e não lineares de efeito misto para dados censurados, usando a distribuição t de Student.

O trabalho contido nesta dissertação é organizado em quatro capítulos. No Capítulo 2, apresentaremos os LMEC/NLMEC usando a distribuição normal. Depois, iremos desenvolver o algoritmo EM para estimar os parâmetros do modelo e faremos uma análise de diagnóstico de influência local e global, baseado na metodologia proposta por (Cook, 1986) e Poon and Poon (1999). Concluiremos este capítulo com a aplicação de dois conjuntos de dados usados por Vaida and Liu (2009) e um estudo de simulação para ilustrar a metodologia proposta.

No Capítulo 3, faremos uma apresentação e uma descrição dos LMEC/NLMEC utilizando a distribuição t de Student propondo o algoritmo EM para estimar os parâmetros nesse modelo. Vamos ilustrar a metodologia proposta com a aplicação de dois conjuntos de dados utilizados no capítulo 2. Finalmente, apresentaremos um conjunto de dados simulados para ilustrar como os procedimentos podem ser usados para avaliar suposições do modelo, identificar outliers, e obter estimativas robustas dos parâmetros.

Por fim, o Capítulo 4 é dedicado a comentários finais e direções para trabalhos futuros.

Capítulo 2

The normal linear and nonlinear mixed-effect models with censored data

2.1 Introduction

Studies of HIV viral dynamics, often considered to be the a key issue in AIDS research, considers repeated/longitudinal measures over a period of treatment routinely analyzed using linear and non-linear mixed effects models (LME/NLME) to assess rates of changes in HIV-1 RNA level or viral load (Wu, 2005, 2010). Viral load measures the amount of actively replicating virus and its reduction is frequently used as a primary endpoint in clinical trials of anti-retroviral (ARV) therapy. However, depending on the diagnostic assays used, its measurement may be subjected to some upper and lower detection limits, below or above which they are not quantifiable (resulting in left or right censoring). The proportion of censored data in these studies may not be small (Hughes, 1999) and so the use of crude/adhoc methods viz., substituting threshold value or some arbitrary point such as mid-point between zero and cut-off for detection (Vaida and Liu, 2009) might lead to biased estimates of fixed effects and variance components (Wu, 2010).

Our motivating datasets in this study are on HIV-1 viral load, (i) after unstructured treatment interruption, or UTI (Saitoh et al., 2008) and (ii) set point for acutely infected subjects from the AIEDRP program (Vaida and Liu, 2009). The former has about 7% observations below (left-censored) the detection-limits, whereas the later has about 22% lying above (right-censored) the limits of assay quantifications. As an alternatives to crude imputation methods, Hughes (1999) proposed a likelihood-based Monte Carlo EM algorithm (MCEM) for LME with censored responses (LMEC). Vaida et al. (2007) proposed a hybrid EM using a more efficient Hughes algorithm, extending it to NLME with censored data (NLMEC). Recently, Vaida and Liu (2009) proposed an exact EM-type algorithm for LMEC/NLMEC, which uses closed-form expressions at the E-step, as opposed to Monte Carlo simulations. Strictly speaking, these algorithms are Space Alternating Generalized EM (SAGE) algorithms (see Vaida et al., 2007).

In this chapter, in order to perform diagnostics analysis in LMEC/NLMEC models, we first propose a slight modification to the EM-type algorithm proposed by Vaida and Liu (2009), wherein all the parameters are updated (M-step) by considering the random effects and the censoring observations as missing data. Then, the diagnostic measures for assessing the local influence in LMEC/NLMEC are developed and presented. Finally, the methodology has been illustrated with the analysis of two examples involving HIV viral measure and an empirical study.

2.2 The multivariate normal and truncated normal-distribution

A random variable $\mathbf{Y}$ is said to follow a p-variate *Normal* distribution with mean vector $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$ (positive definite), denoted by $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if the probability density function (pdf) of $\mathbf{Y}$, is given by

$$\phi_{\mathbf{p}}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}}|\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\},$$

where $\Phi_p(\mathbf{u}|\mathbf{a}, \mathbf{A})$ and $\phi_p(\mathbf{u}|\mathbf{a}, \mathbf{A})$ are the cdf (left tail) and pdf, respectively, of $N_p(\mathbf{a}, \mathbf{A})$ computed at vector $\mathbf{u}$. In order to introduce some notation, for a Normal random vector, we establish the following proposition which is important for our subsequent research.

Proposition 1 *Let* $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $\mathbf{Y}$ *is partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, *with* $dim(\mathbf{Y}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *and* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *be the corresponding partitions of* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\mu}$. *Then*

i) $\mathbf{Y}_1 \sim N_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$,

ii) The conditional cdf of $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1$ *is given by*

$$P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = \Phi_{p_2}\left(\mathbf{y}_2|\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}\right), \tag{2.1}$$

i.e., $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1 \sim N_{p_2}\left(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}\right)$, *where* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, *and* $F_{\mathbf{Y}}(.|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *denotes the cdf of the p-variate Normal distribution with parameters* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$.

Now, let $TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A})$ represent a p-variate truncated Normal distribution for $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ lying within a right-truncated hyperplane

$$\mathbb{A} = \{\mathbf{x} = (x_1, \ldots, x_p)^\top | x_1 \leq a_1, \ldots, x_p \leq a_p\}. \tag{2.2}$$

Specifically, we say that the p-dimensional vector $\mathbf{X} \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A})$, if its density is given by:

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{\phi_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{\Phi_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}\mathbb{I}_{\mathbb{A}}(\mathbf{x}), \tag{2.3}$$

where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbb{I}_{\mathbb{A}}(\mathbf{x})$ is the indicator function whose value equals one if $\mathbf{x} \in \mathbb{A}$ and zero otherwise.

2.3 Linear mixed effects with censored response

Ignoring censoring for the moment, the classical normal LME models is specified as follows (Laird and H.Ware, 1982):

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{2.4}$$

where $\mathbf{b}_i \overset{iid}{\sim} N_q(\mathbf{0}, \mathbf{D})$ is independent of $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} N_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$, $\quad i = 1, \ldots, n$; the subscript i is the subject index; $\mathbf{I}_p$ denotes the $p \times p$ identity matrix; $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is a

$n_i \times 1$ vector of observed continuous responses for subject i; $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the $p \times 1$ vector of fixed effects $\boldsymbol{\beta}$; $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$; $\boldsymbol{\epsilon}_i$ of dimension $(n_i \times 1)$ is the vector of random errors, and the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$. In the present formulation, we consider the case where the response Y_{ij} is not fully observed for all $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Let the observed data for the i-th subject be $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i$ represents the vector of uncensored reading or censoring level, and $\mathbf{C}_i$ the vector of censoring indicators, such that

$$\begin{aligned} y_{ij} &\leq V_{ij} \quad \text{if} \quad C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{if} \quad C_{ij} = 0. \end{aligned} \tag{2.5}$$

For simplicity, we will assume that the data are left-censored and thus the LMEC is defined. The extensions of theses results to arbitrary censoring can be easily presented.

2.3.1 The likelihood function

Following Vaida and Liu (2009), classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$. For complete data, we have that marginally $\mathbf{y}_i \overset{\text{ind.}}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{ni} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. For responses with censoring pattern as in (2.5), we have that $\mathbf{y}_i \sim TN_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i; \mathbb{A}_i)$, where $TN_{n_i}(.; \mathbb{A}_i)$ denotes the truncated normal distribution on the interval $\mathbb{A}_i$, where $\mathbb{A}_i = A_{i1} \times \ldots, \times A_{ini}$, with A_{ij} as the interval $(-\infty, \infty)$, if $C_{ij} = 0$ and $(-\infty, V_{ij}]$, if $C_{ij} = 1$. For computing the likelihood function associated with model (2.4)–(2.5), the first step is to treat separately the observed and censored components of $\mathbf{y}_i$. Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i with $(n_i = n_i^o + n_i^c)$, such that, $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$, and 1 for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{V}_i$, $\mathbf{X}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as follows:

$$\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c),\ \mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c),\ \mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c) \text{ and } \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} & \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} & \boldsymbol{\Sigma}_i^{cc} \end{pmatrix},$$

where $vec(.)$ denote the function which stacks vectors or matrices of the same number of columns. Then we have $\mathbf{y}_i^o \sim N_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo})$, $\mathbf{y}_i^c|\mathbf{y}_i^o \sim N_{n_i^c}(\boldsymbol{\mu}_i, \mathbf{S}_i)$, where $\boldsymbol{\mu}_i =$

$\mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$ and $\mathbf{S}_i = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}$. From Vaida and Liu (2009) and Jacqmin-Gadda et al. (2000), the likelihood function for cluster i (using conditional probability arguments) is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i|\boldsymbol{\theta}) = f(\mathbf{y}_i^o|\boldsymbol{\theta}) f(\mathbf{y}_i^c|\mathbf{y}_i^o, \boldsymbol{\theta}) = f(\mathbf{y}_i^o|\boldsymbol{\theta}) P(\mathbf{y}_i^c \leq \mathbf{V}_i^c|\mathbf{y}_i^o, \boldsymbol{\theta}) \\ &= \phi_{n_i^o}(\mathbf{y}_i^o|\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}) \Phi_{n_i^c}(\mathbf{V}_i^c|\boldsymbol{\mu}_i, \mathbf{S}_i) = L_i, \end{aligned} \tag{2.6}$$

which can be evaluated without much computational burden through the routine *mvtnorm()* available in R Genz et al. (2008); R Development Core Team (2009). The log-likelihood function for the observed data is thus given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \{\log L_i\}$. Thus the estimates obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y})$ are thus the maximum likelihood estimates (MLEs).

2.3.2 The EM algorithm

As the observed log-likelihood function involves complex expressions, it is very difficult to work directly with $\ell(\boldsymbol{\theta}|\mathbf{y})$, either for the ML estimation or to carry out the influence analysis. For LMEC/NLMEC, an EM-type algorithm was developed by Vaida and Liu (2009) for the ML estimation, in which $\boldsymbol{\beta}$ and σ^2 are updated by integrating out $\mathbf{b}_i$ (marginal model), while $\mathbf{D}$ is updated with $\mathbf{y}_i$ and $\mathbf{b}_i$ as missing data. We proposed here an expectation conditional maximization (ECM) algorithm by considering $\mathbf{y}_i$ and $\mathbf{b}_i$ as missing data to update (M-step) all the parameters involved in the model.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$, and that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the ith subject. Treating $\mathbf{b}$ and $\mathbf{y}$ as hypothetical missing data, and augmented with the observed data $\mathbf{V}, \mathbf{C}$, we set $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$. Hence, the EM-type algorithm is applied to the complete-data log-likelihood function $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\mathbf{y}_c)$, where

$$\begin{aligned} \ell_i(\boldsymbol{\theta}|\mathbf{y}_c) &= -\frac{1}{2}\Big[\log\sigma^2 + \frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ &\quad + \log|\mathbf{D}| + \mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\Big] + C, \end{aligned} \tag{2.7}$$

and C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete

log-likelihood function given by (see appendix)

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}^{(k)}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \\ &= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta},\sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}), \end{aligned} \tag{2.8}$$

where

$$Q_{1i}(\boldsymbol{\beta},\sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left[\widehat{a}_i^{(k)} - 2\boldsymbol{\beta}^\top\mathbf{X}_i^\top(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}) + \boldsymbol{\beta}^\top\mathbf{X}_i^\top\mathbf{X}_i\boldsymbol{\beta}\right]$$

and

$$Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{b}_i^2}^{(k)}\mathbf{D}^{-1}\right),$$

with $\widehat{a}_i^{(k)} = \mathrm{tr}\left(\widehat{\mathbf{y}_i^2}^{(k)} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top + \widehat{\mathbf{b}_i^2}^{(k)}\mathbf{Z}_i^\top\mathbf{Z}_i\right)$, $\widehat{\mathbf{b}_i}^{(k)} = E\{\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{\mathbf{y}_i}^{(k)} - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})$, $\widehat{\mathbf{b}_i^2}^{(k)} = E\{\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\boldsymbol{\Lambda}}_i^{(k)} + \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{\mathbf{y}_i^2}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\mathbf{y}}_i^{(k)\top} + \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$, $\widehat{\mathbf{y}_i\mathbf{b}_i^\top} = E\{\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = (\widehat{\mathbf{y}_i^2}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$, with $\widehat{\boldsymbol{\Lambda}}_i^{(k)} = (\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top\mathbf{Z}_i/\widehat{\sigma^2}^{(k)})^{-1}$ and $\widehat{\boldsymbol{\varphi}}_i^{(k)} = \widehat{\boldsymbol{\Lambda}}_i^{(k)}\mathbf{Z}_i^\top/\widehat{\sigma^2}^{(k)}$.

It is clear that the E-step reduces to the computation of $\widehat{\mathbf{y}_i^2} = E\{\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\}$ and $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\}$, that is, the mean and second moment of a truncated multinormal distribution. These can be determined in closed form, as a function of multinormal probabilities, using a sequence of simple transformations. For more details on the computation of these moments one may refer to Vaida and Liu (2009).

The conditional maximization (CM) steps then conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ and obtains a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^{n}\mathbf{X}_i^\top\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i^\top\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right), \tag{2.9}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^{n}\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}) + \widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right], \tag{2.10}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{b}_i^2}^{(k)}, \tag{2.11}$$

where $N = \sum_{i=1}^{n} n_i$. This process is iterated until some distance between two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y})$ given in subsection 2.3.1, such as $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})|$ or $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, is small enough.

These expected values can be determined in closed form, using proposition 1, as follows.

1. If $\mathbf{y}_i = \mathbf{y}_i^c$, i.e, the individual i has only censored components. we have:

$$\begin{aligned}\widehat{\mathbf{y}_i^2} &= E\{\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\},\\ \widehat{\mathbf{y}}_i &= E\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\},\end{aligned}$$

   where $\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i \sim TN_{n_i}(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i; \mathbb{A}_i)$, $\widehat{\boldsymbol{\mu}}_i = \mathbf{X}_i\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\sigma^2}\mathbf{I}_{ni} + \mathbf{Z}_i\widehat{\mathbf{D}}\mathbf{Z}_i^\top$ and $\mathbb{A}_i = \{\mathbf{y}_i = (y_1, \ldots, y_{n_i})^\top | y_1 \leq V_{i1}, \ldots, y_{n_i} \leq V_{in_i}\}$.

2. If $\mathbf{y}_i = \mathbf{y}_i^o$, i.e, the individual i has non censored components. Then,

$$\widehat{\mathbf{y}_i^2} = \mathbf{y}_i\mathbf{y}_i^\top,\ \widehat{\mathbf{y}}_i = \mathbf{y}_i,$$

   and finally

3. If $\mathbf{y}_i = (\mathbf{y}_i^{c\top}, \mathbf{y}_i^{o\top})^\top$, i.e., for individual i, we observed censored and uncensored components. Then from Proposition 1 and by the fact that $\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i\}$,

$\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$ and $\{\mathbf{y}_i^c|\mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$ are equivalent processes, we have

$$\begin{aligned}\widehat{\mathbf{y}_i^2} &= E\{\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top} & \mathbf{y}_i^o\widehat{\mathbf{y}}_i^{c\top} \\ \widehat{\mathbf{y}}_i^c\mathbf{y}_i^{o\top} & \widehat{\mathbf{y}^2}_i^c \end{pmatrix}, \\ \widehat{\mathbf{y}}_i &= E\{\mathbf{y}_i|\mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = vec(y_i^o, \widehat{\mathbf{y}}_i^c),\end{aligned}$$

where $\widehat{\mathbf{y}}_i^c = E\{\mathbf{y}_i^c\}$ and $\widehat{\mathbf{y}^2}_i^c = E\{\mathbf{y}_i^c\mathbf{y}_i^{c\top}\}$, with $\mathbf{y}_i^c \sim TN_{n_i^c}(\boldsymbol{\mu}_i^{co}, \boldsymbol{\Sigma}_i^{cc.o}; \mathbb{A}_i^c)$, $\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}$ and $\boldsymbol{\mu}_i^{co} = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$.

The variance of the fixed effects in the LMEC is given by (Hughes, 1999)

$$Var(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i - \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i)\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}. \tag{2.12}$$

2.4 Diagnostic analysis

Influence diagnostics techniques consist of evaluating the sensitivity of the parameter estimates of a particular model when perturbation occurs either in the data set or in the underlying assumptions of the model. There are primarily two approaches for detecting influential observations. The first approach is the case-deletion technique (Cook, 1977), in which is a common approach for analyzing one or more fitted models after excluding some observations and then assessing by some metrics such as the likelihood distance and the Cook's distance. The second method is the local influence approach (Cook, 1986), which evaluates the changes in the results of the analysis by incorporating a minor perturbation to the model. By using the results of Zhu and Lee (2001), we will introduce here the case-deletion measures and the local influence measures to the censored data on the basis of the following Q-function $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$. We discuss first the case-deletion measures, then the local influence, and finally the perturbation schemes used.

2.4.1 Case-deletion measures

Case-deletion is a common approach to study the effect of dropping the ith case from the data set. In the following, a quantity with a subscript "$[i]$"denotes the original

quantity with the ith case deleted. The log-likelihood function of $\boldsymbol{\theta}$, based on the data with the ith case deleted, is denoted by $\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})$. Let $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}^{\top}_{[i]}, \widehat{\sigma^2}_{[i]}, \widehat{\boldsymbol{\alpha}}^{\top}_{[i]})^{\top}$ be the maximizer of the function $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}\}$, where $\widehat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$. To assess the influence of the ith case on the ML estimate $\widehat{\boldsymbol{\theta}}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. If the deletion of a case seriously influences the estimates, more attention need to be paid to that case. Hence, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith case is regarded as influential. As $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, the required computational effort can be quite heavy, especially when a sample is large. Hence, the following one-step pseudo approximation $\widehat{\boldsymbol{\theta}}^{1}_{[i]}$ is used to reduce the burden (see Cook and Weisberg, 1982; Zhu and Lee, 2001)

$$\widehat{\boldsymbol{\theta}}^{1}_{[i]} = \widehat{\boldsymbol{\theta}} + \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{2.13}$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ is the Hessian matrix and $\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, $i = 1, \ldots, n$, has its elements as follows

$$\dot{Q}_{[i]\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\beta} = \frac{1}{\sigma^2}E_{1[i]}, \tag{2.14}$$

$$\dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\sigma^2 = -\frac{1}{2\sigma^2}E_{2[i]}, \tag{2.15}$$

$$\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\alpha}, \tag{2.16}$$

where $E_{1[i]} = \sum_{j\neq i}\mathbf{X}_j^{\top}(\widehat{\mathbf{y}}_j - \mathbf{Z}_j\widehat{\mathbf{b}}_j - \mathbf{X}_j\widehat{\boldsymbol{\beta}})$ and $E_{2[i]} = \sum_{j\neq i}(n_j - \frac{A_j}{\sigma^2})$, with $A_j = a_j - 2\widehat{\boldsymbol{\beta}}^{\top}\mathbf{X}_j^{\top}(\widehat{\mathbf{y}}_j - \mathbf{Z}_j\widehat{\mathbf{b}}_j) + \widehat{\boldsymbol{\beta}}^{\top}\mathbf{X}_j^{\top}\mathbf{X}_j\widehat{\boldsymbol{\beta}}$. Finally, $\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ has its elements as

$$\dot{Q}_{[i]\alpha_r}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{1}{2}\sum_{j\neq i}\mathrm{tr}[\mathbf{D}^{-1}\dot{\mathbf{D}}(r) - \mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{\mathbf{b}_j\mathbf{b}_j^{\top}}].$$

The Hessian matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$

Following Zhu and Lee (2001), to obtain the diagnostic measures for case-deletion diagnostic and for local influence of a particular perturbation scheme, it is necessary to compute $\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \sigma^2, \boldsymbol{\alpha}^{\top})^{\top}$ is the parameter vector. Hence, the Hessian matrix $\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}$ has its elements as follows:

$$\begin{aligned}
\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2}\mathbf{X}_i^\top\mathbf{X}_i, \quad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2} = -\frac{1}{\sigma^4}\mathbf{X}_i^\top(\mathbf{Y}_i - \mathbf{Z}_i\mathbf{b}_i - \mathbf{X}_i\boldsymbol{\beta}),\\
\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\alpha_r} &= \mathbf{0}, \quad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\sigma^2} = \frac{1}{2\sigma^4}[n_i - \frac{2}{\sigma^2}A_i],\\
\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\alpha_r} &= 0, \quad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_s\partial\alpha_r} = \frac{1}{2}\mathrm{tr}(\mathbf{A}(sr)) - \frac{1}{2}\mathrm{tr}(\mathbf{B}(sr)\widehat{\mathbf{b}_i\mathbf{b}_i^\top}),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{A}(sr) &= \mathbf{D}^{-1}[\dot{\mathbf{D}}(s)\mathbf{D}^{-1}\dot{\mathbf{D}}(r) - \ddot{\mathbf{D}}(s,r)] \quad \textit{and}\\
\boldsymbol{B}(sr) &= \mathbf{D}^{-1}[\dot{\mathbf{D}}(s)\mathbf{D}^{-1}\dot{\mathbf{D}}(r) + \dot{\mathbf{D}}(r)\mathbf{D}^{-1}\dot{\mathbf{D}}(s) - \ddot{\mathbf{D}}(s,r)]\mathbf{D}^{-1},
\end{aligned}$$

with $\dot{\mathbf{D}}(r) = \partial\mathbf{D}/\partial\alpha_r$, $\ddot{\mathbf{D}}(s,r) = \partial^2\mathbf{D}/\partial\alpha_s\partial\alpha_r$, $r,s = 1,\ldots,p^*$, $p^* = dim(\boldsymbol{\alpha})$ and $i = 1,\ldots,n$. After some rearrangement and evaluating these derivatives at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ we obtain the Hessian matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$, which is a block-diagonal matrix of the form $\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{diag}(\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \ddot{Q}_\alpha(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}))$, where

$$\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{1}{\widehat{\sigma}^2}\mathbf{X}^\top\mathbf{X}, \quad \ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{b}{2\widehat{\sigma}^4} \quad \text{and} \quad \ddot{Q}_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\left(\frac{\partial^2 Q_i(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_s\partial\alpha_r}\right),$$

where $\mathbf{X} = (\mathbf{X}_1^\top, ..., \mathbf{X}_n^\top)^\top$, $b = -\sum_{i=1}^n(n_i - 2A_i/\widehat{\sigma}^2)$.

Next, we will obtain the one-step approximation of $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\sigma^2}_{[i]}, \widehat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$, $i = 1,\ldots,n$, based on (2.13). Namely, the relationship between parameter estimates for full data and the data with the ith case deleted.

Theorem 1 *For the LMEC, the relationships between the parameter estimates for full data and the data with the ith case deleted are as follows:*

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{[i]}^1 &= \widehat{\boldsymbol{\beta}} + (\mathbf{X}^\top\mathbf{X})^{-1}\boldsymbol{E}_{1[i]},\\
\widehat{\sigma^2}_{[i]}^1 &= \widehat{\sigma^2} - \frac{1}{b}E_{2[i]},\\
\widehat{\boldsymbol{\alpha}}_{[i]}^1 &= \widehat{\boldsymbol{\alpha}} + \{-\ddot{Q}_\alpha(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}),
\end{aligned}$$

where $\boldsymbol{E}_{1[i]}$, $\boldsymbol{E}_{2[i]}$ and $\dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$, $i = 1,\ldots,n$, are as in (2.14), (2.15) and (2.16), respectively.

From Theorem 1, case-deletion measures can be developed for assessing influential observations, such as the generalized Cook distance and the likelihood distance (Zhu and Lee, 2001). To assess the influence of the ith case on the ML estimate $\widehat{\boldsymbol{\theta}}$, we need to compare $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, and if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith case is regarded as influential. Based on the metric for measuring the distance between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$ proposed by Zhu and Lee (2001) based on the EM algorithm, we consider here the following generalized Cook distance:

$$GD_i = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}), i = 1, \ldots, n. \tag{2.17}$$

Upon substituting (2.13) into (2.17), we obtain the approximation:

$$GD_i^1 = \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}),\ i = 1, \ldots, n.$$

Since $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is a diagonal matrix, from Xie et al. (2007), GD_i^1 can be decomposed into three parts that corresponds to the generalized Cook distance for parameter subset $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\alpha}$, which are denoted, respectively, by $GD_i^1(\boldsymbol{\beta})$, $GD_i^1(\sigma^2)$ and $GD_i^1(\boldsymbol{\alpha})$, as follows:

$$GD_i^1 = GD_i^1(\boldsymbol{\beta}) + GD_i^1(\sigma^2) + GD_i^1(\boldsymbol{\alpha}), \tag{2.18}$$

where

$$\begin{aligned}
GD_i^1(\boldsymbol{\beta}) &= \dot{Q}_{[i]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{1}{\widehat{\sigma}^2} \mathbf{E}_{1[i]}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{E}_{1[i]}, \\
GD_i^1(\sigma^2) &= \dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{1}{2b} \mathbf{E}_{2[i]}^2, \\
GD_i^1(\boldsymbol{\alpha}) &= \dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_\alpha(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}).
\end{aligned}$$

Another measure for the influence of the ith case is the following Q-distance function, similar to the likelihood distance LD_i (Cook and Weisberg, 1982) is defined as

$$QD_i = 2\{Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}})\}. \tag{2.19}$$

We can calculate an approximation of the likelihood displacement QD_i by substituting

(2.13) into (2.19), resulting in the following approximation QD_i^1 of QD_i:

$$QD_i^1 = 2\{Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{[i]}^1|\widehat{\boldsymbol{\theta}})\}. \tag{2.20}$$

2.4.2 Local influence

In this subsection we derive the normal curvature of local influence (Cook, 1986) for some common perturbation schemes either in the model or in the data. We will consider the case-weight, scale matrix perturbation schemes and response perturbation schemes, for this purpose.

Consider a perturbation vector $\boldsymbol{\omega} = (\omega_1, ..., \omega_g)^\top$ varying in an open region $\boldsymbol{\Omega} \subset \mathbb{R}^g$. Let $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$ be the complete-data log-likelihood to the perturbed model. We assume that there is a $\boldsymbol{\omega}_0$ in $\boldsymbol{\Omega}$ such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ for all $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ denote the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}]$. The influence graph is then defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q-displacement function defined as follows:

$$f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right].$$

Following the approach of Cook (1986) and Zhu and Lee (2001), the normal curvature $C_{f_Q,\mathbf{d}}$ of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\mathbf{d}$ can be used to summarize the local behavior of the Q-displacement function. It can be shown that

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_o}\mathbf{d} \text{ and } -\ddot{Q}_{\boldsymbol{\omega}_0} = \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}^\top \left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0},$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

Following the same procedure as in Cook (1986), the quantity $-\ddot{Q}_{\boldsymbol{\omega}_0}$ is quite useful for detecting influential observations. From the spectral decomposition of a symmetric matrix $-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^g \zeta_k \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^\top$, where $\{(\zeta_k, \boldsymbol{\varepsilon}_k), k = 1, \ldots, g\}$ are eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ with $\zeta_1 \geq \ldots \geq \zeta_r > \zeta_{r+1} = \ldots = 0$ and orthonormal eigenvectors $\{\boldsymbol{\varepsilon}_k, k = 1, \ldots, g\}$, Zhu and Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for capturing more information. Based on the work of Zhu and Lee (2001), we consider the following aggregated contribution vector of all eigenvectors that corresponding to nonzero eigenvalues. Let $\tilde{\zeta}_k = \zeta_k/(\zeta_1 + \ldots + \zeta_r)$, $\boldsymbol{\varepsilon}_k^2 = (\boldsymbol{\varepsilon}_{k1}^2, \ldots, \boldsymbol{\varepsilon}_{kg}^2)^\top$

and $M(0) = \sum_{k=1}^{r} \tilde{\zeta}_k \boldsymbol{\varepsilon}_k^2$. The lth component of $M(0)$, $M(0)_l$, is equal to $\sum_{k=1}^{r} \tilde{\zeta}_k \boldsymbol{\varepsilon}_{kl}^2$. The assessment of influential cases is based on the visual inspection of the $\{M(0)_l, l = 1, \ldots, g\}$ plotted against the index l. The lth case may be regarded as influential if $M(0)_l$ is larger than the benchmark.

The inconvenience on the use of the normal curvature is in deciding about the influence of the observations, since $C_{f_Q,\mathbf{d}}(\boldsymbol{\theta})$ may assume any value and it is not invariant under a uniform change of scale. Based on the work of Poon and Poon (1999) in using a conformal normal curvature, Zhu and Lee (2001) considered the following conformal normal curvature $B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) = C_{f_Q,\mathbf{d}}(\boldsymbol{\theta})/tr[-2\ddot{Q}\boldsymbol{\omega}_0]$, whose computation is quite simple and also has the property that $0 \leq B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) \leq 1$. Let $\mathbf{d}_l$ be a basic perturbation vector with lth entry as 1 and all other entries as zero. Zhu and Lee (2001) showed that for all l, $M(0)_l = B_{f_Q,\mathbf{d}_l}$. We can therefore obtain $M(0)_l$ via $B_{f_Q,\mathbf{u}_l}$.

So far, there is not a general rule to judge the largeness of the influence of a specific case in the data. Let $\overline{M}(0)$ and $SM(0)$ denote, respectively, the mean and the standard error of $\{M(0)_l : l = 1, \ldots, g\}$, where $\overline{M(0)} = 1/g$. Poon and Poon (1999) proposed to use $2\overline{M}(0)$ as benchmarks for $M(0)$. However, we may use different functions of $M(0)$. For instance, Zhu and Lee (2001) proposed to use $\overline{M}(0) + 2SM(0)$ as a benchmark to take into account the variance of $\{M(0)_l : l = 1, \ldots, g\}$. According to Lee and Xu (2004), the exact choice of the function of $\overline{M}(0)$ as the benchmark is subjective. More recently, Lee and Xu (2004) proposed to use $\overline{M}(0) + c^* SM(0)$, where c^* is a selected constant, and depending on the application, c^* may be taken to be any value. In this work, we will consider $c^* = 3, 5$.

2.4.3 Perturbation schemes

Now, in this subsection, we will evaluate the $\boldsymbol{\Delta}$ matrix under the following perturbation schemes for LMEC models. *Case-weight* made for detecting observations with outstanding contribution on the log-likelihood function and that may exercise high influence on the maximum likelihood estimates. *Scale perturbation* made on the scale matrix $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{ni} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top$. It also can be made on either σ^2 or $\mathbf{D}$ which may reveal individuals that are most influential, in the sense, of the likelihood displacement on the scale structure. Finally, *perturbation of response variables* made on the response values, which may indicate observations with large influence on the MLE. In our case

the response variables are $\mathbf{V}'s$.

For each perturbation scheme, one has the partitioned form

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_o} = (\boldsymbol{\Delta}_{\beta}^{\top}, \boldsymbol{\Delta}_{\sigma^2}^{\top}, \boldsymbol{\Delta}_{\alpha}^{\top})^{\top},$$

where $\boldsymbol{\Delta}_{\beta} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^{\top}}|_{\boldsymbol{\omega}_o} \in \mathbb{R}^{p \times g}$, $\boldsymbol{\Delta}_{\sigma^2} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \boldsymbol{\omega}^{\top}}|_{\boldsymbol{\omega}_o} \in \mathbb{R}^{1 \times g}$ and $\boldsymbol{\Delta}_{\alpha} = (\boldsymbol{\Delta}_{\alpha 1}^{\top}, \ldots, \boldsymbol{\Delta}_{\alpha p^*}^{\top})^{\top}$, with $\boldsymbol{\Delta}_{\alpha r} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_r \partial \boldsymbol{\omega}^{\top}}|_{\boldsymbol{\omega}_o} \in \mathbb{R}^{1 \times g}$, $r = 1, \ldots, p^*$ and g being the dimensions of the perturbation vector $\boldsymbol{\omega}$.

Case weight perturbation

First, we consider an arbitrary attribution of weights for the expected value of the complete-data log-likelihood function (perturbed Q–function), which may capture departures in general directions, represented by writing

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}] = \sum_{i=1}^{n} \omega_i \mathrm{E}[\ell_i(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}] = \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}).$$

Here $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\top}$ is an $n \times 1$ vector and $\boldsymbol{\omega}_o = (1, \ldots, 1)^{\top}$. In addition, it is possible to show that the local influence for this perturbation scheme is equivalent to the deletion method discussed in previous subsection. For this perturbation scheme, we find

$$\begin{aligned}
\boldsymbol{\Delta}_{\beta} &= \frac{1}{\sigma^2} \boldsymbol{X}^{\top} D(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n), \\
\boldsymbol{\Delta}_{\sigma^2} &= -\frac{1}{2\sigma^2} \mathbf{n}^{\top} + \frac{1}{2\sigma^4} \mathbf{m}^{\top}, \\
\boldsymbol{\Delta}_{\alpha_r} &= [\frac{\partial Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_r}, \ldots, \frac{\partial Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_r}], \ r = 1, \ldots, p^*,
\end{aligned}$$

where $\mathbf{n} = (n_1, \ldots, n_n)^{\top}$, $\mathbf{m} = (A_1, \ldots, A_n)^{\top}$, $D(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$ is a block-diagonal matrix, with $\boldsymbol{\epsilon}_i = \widehat{\mathbf{y}}_i - \mathbf{Z}_i \widehat{\mathbf{b}}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_i$ and $\dfrac{\partial Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_r} = -\dfrac{1}{2} \mathrm{tr}[\mathbf{D}^{-1} \dot{\mathbf{D}}(r) - \mathbf{D}^{-1} \dot{\mathbf{D}}(r) \mathbf{D}^{-1} \widehat{\mathbf{b}_i^2}]$.

Scale matrix perturbation

To study the effects of departures from the assumption regarding the scale matrix, we consider the perturbations $\mathbf{D}(\omega_i) = \omega_i^{-1}\mathbf{D}$ or $\sigma^2(\omega_i) = \omega_i^{-1}\sigma^2$, for $i = 1, \ldots, n$. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_o = (1, \ldots, 1)^\top$. Moreover, the perturbed Q-function is as in (2.8), switching $\mathbf{D}(\omega_i)$ and $\sigma^2(\omega_i)$ with $\mathbf{D}$ and σ^2, respectively. The matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has its elements as follows:

- Perturbation on $\mathbf{D}$: $\boldsymbol{\Delta}_\beta = \mathbf{0}$, $\boldsymbol{\Delta}_{\sigma^2} = \mathbf{0}$ and $\boldsymbol{\Delta}_{\alpha_r} = \frac{1}{2}[g_1, \ldots, g_n]$, where $g_i = \mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{\mathbf{b}_i^2}$, $r = 1, \ldots, p^*$.

- Perturbation on σ^2: $\boldsymbol{\Delta}_\beta = \frac{1}{\sigma^2}\boldsymbol{X}^\top D(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$, $\boldsymbol{\Delta}_{\sigma^2} = \dfrac{1}{2\sigma^4}\mathbf{m}^\top$ and $\boldsymbol{\Delta}_{\boldsymbol{\alpha}} = \mathbf{0}$.

Response perturbation

A perturbation of the response variables V_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, can be introduced by replacing V_{ij} by $V_{ij}(\omega) = V_{ij} + \omega_i s_{ij}$, where s_{ij} is a scale factor. Now substituting $V_{ij}(\omega)$ into (2.5), we can write perturbed model as

$$\begin{aligned} y_{ij}(\omega) &\leq V_{ij} \quad \text{if } C_{ij} = 1, \\ y_{ij}(\omega) &= V_{ij} \quad \text{if } C_{ij} = 0, \end{aligned}$$

where $\mathbf{y}_{ij}(\omega) = \mathbf{y}_{ij} - \omega_i s_{ij}$. Hence, the perturbed Q-function $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})$ is as in subsection 2.3.2, with $\widehat{\mathbf{y}_i}$, $\widehat{\mathbf{y}_i^2}$ and $\widehat{\mathbf{y}_i\mathbf{b}_i^\top}$ replaced by with $\widehat{\mathbf{y}_{i\omega}} = \widehat{\mathbf{y}}_i - \omega_i\mathbf{s}_i$, $\widehat{\mathbf{y}_{i\omega}^2} = \widehat{\mathbf{y}_i^2} - \omega_i(\widehat{\mathbf{y}}_i\mathbf{s}_i^\top + \mathbf{s}_i\widehat{\mathbf{y}_i^\top}) + \omega_i^2\mathbf{s}_i\mathbf{s}_i^\top$ and $\widehat{\mathbf{y}_{i\omega}\mathbf{b}_{i\omega}^\top} = \widehat{\mathbf{y}_i\mathbf{b}_i^\top} - \omega_i\mathbf{s}_i\widehat{\mathbf{b}_i^\top}$, respectively, with $\mathbf{s}_i = (s_{i1}, \ldots, s_{in_i})^\top$. Under this perturbation scheme the vector $\boldsymbol{\omega}_0$, representing no perturbation, is given by $\boldsymbol{\omega}_0 = \mathbf{0}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has the following elements:

$$\begin{aligned} \boldsymbol{\Delta}_\beta &= -\frac{1}{\sigma^2}\boldsymbol{X}^\top D(\mathbf{s}_1, \ldots, \mathbf{s}_n), \\ \boldsymbol{\Delta}_{\sigma^2} &= -\frac{1}{\sigma^4}(\mathbf{Y} - \mathbf{Z}\mathbf{b} - \mathbf{X}\boldsymbol{\beta})^\top D(\mathbf{s}_1, \ldots, \mathbf{s}_n), \\ \boldsymbol{\Delta}_\alpha &= \mathbf{0}, \end{aligned}$$

where $\mathbf{Y} = (\widehat{\mathbf{y}}_1^\top, \ldots, \widehat{\mathbf{y}}_n^\top)^\top$, $\mathbf{b} = (\widehat{\mathbf{b}}_1, \ldots, \widehat{\mathbf{b}}_n)^\top$ and $D(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ is a block-diagonal matrix.

2.5 The nonlinear case

The NLME (Pinheiro and Bates, 2000) is defined as:

$$\mathbf{y}_i = \eta(\boldsymbol{\phi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \;\; \boldsymbol{\phi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \;\; i = 1, \ldots, n, \tag{2.21}$$

where $\mathbf{b}_i \sim N_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \sim N_{n_i}(0, \sigma^2\mathbf{I})$ are independent; $\mathbf{y}_i$ is a $(n_i \times 1)$ vector of observed continuous responses for subject i; η is a nonlinear function of the individual random parameter $\boldsymbol{\phi}_i$; $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$ respectively, possibly depending on some covariable values; $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of fixed effects and $\mathbf{b}_i$ is the $(q \times 1)$ vector of random effects.

As mentioned by Vaida and Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ involves of taking the first-order Taylor expansion of η_i around the current parameter estimate $\widetilde{\boldsymbol{\beta}}$ and the random effect estimates $\widetilde{\mathbf{b}}_i$ (empirical predictors), which is equivalent to iteratively solving the following LME model (L-step)

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \;\; i = 1, \ldots, n, \tag{2.22}$$

where $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \widetilde{\eta}(\boldsymbol{\phi}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i)$, $\mathbf{b}_i \overset{ind}{\sim} N_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} N_{n_i}(\mathbf{0}, \sigma_e^2\mathbf{I}_{n_i})$, $\widetilde{\mathbf{H}}_i = \frac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^\top}|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$ and $\widetilde{\mathbf{W}}_i = \frac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\boldsymbol{\beta}_i^\top}|_{\boldsymbol{\beta}_i=\widetilde{\boldsymbol{\beta}}_i}$, and $\widetilde{\eta}(\boldsymbol{\phi}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i) = \eta(\boldsymbol{\phi}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i) - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}} - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i$. Thus, for censored response the linearized model (2.22) is an LME with censored data, with same structure as (2.4), which is then solved as detailed in the previous section. The model matrices in (2.22) depends on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates to convergence between L-, E-, and CM-steps. Moreover, the influence diagnostic procedures discussed earlier in Section 2.4 can be incorporated along with the approximation in (2.22) to obtain approximate influence diagnostics measures for NLMEC.

2.6 Application

We illustrate the performance of the proposed methods with the analysis of two HIV datasets, previously analyzed by Vaida and Liu (2009), and the analysis of a simulated example.

2.6.1 UTI data

The first application is a study of 72 perinatally HIV-infected children (Saitoh et al., 2008). The data set is available in the R package *lmec*. Primarily due to treatment fatigue, unstructured treatment interruptions (UTI) is common in this population. Suboptimal adherence can lead to antiretroviral (ARV) resistance and diminished treatment options in the future. The subjects in the study had taken ARV therapy for at least 6 months before UTI, and the medication was discontinued for more than 3 months. Out of 362 observations, 26 (7%) observations were below the detection limits (50 or 400 copies/mL) and considered left-censored at these values. The individual profiles of viral load at different follow-up times after UTI is presented in Figure 3.1 (right panel). We consider a profile LME model with random intercepts b_i as $y_{ij} = b_i + \beta_j + \epsilon_{ij}$, where y_{ij} is the $\log_{10}$ HIV RNA for subject i at time t_j, $t_1 = 0, t_2 = 1, t_3 = 3, t_4 = 6, t_5 = 9, t_6 = 12, t_7 = 18, t_8 = 24$. The $\log_{10}$ transformation of HIV viral load is used to stabilized the variance of viral load and make the viral load more normally distributed. A summary of these parameter estimates and their respective p-values are presented in Table 2.1. These results are coherent with those indicated in Vaida and Liu (2009). From Table 2.1, we note that all the regression parameters are significant at 5% level.

Tabela 2.1: Parameter estimates of the LMEC model and p-values for the UTI data. SE indicates the standard error.

Parameter	Estimate	SE	p-value
$\widehat{\beta_1}$	3.6038	0.1253	< 0.01
$\widehat{\beta_2}$	4.1664	0.1285	< 0.01
$\widehat{\beta_3}$	4.2413	0.1304	< 0.01
$\widehat{\beta_4}$	4.3604	0.1307	< 0.01
$\widehat{\beta_5}$	4.5662	0.1398	< 0.01
$\widehat{\beta_6}$	4.5692	0.1485	< 0.01
$\widehat{\beta_7}$	4.6773	0.1646	< 0.01
$\widehat{\beta_8}$	4.7935	0.2018	< 0.01
$\widehat{\sigma^2}$	0.3414		
$\widehat{\alpha}$	0.76535		

Global influence

In order to identify outlying observations under the fitted model, the index plot of the Mahalanobis distance $d_i = (\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_i^{-1}(\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})$, $i = 1, \ldots, 72$, is displayed in Figure 2.1(a). We can see from this figure that observations #42 appear as possible outliers. To evaluate the effect on the ML estimates when some observation is eliminated, we analyze the QD_i^1 and GD_i^1 index plots, which are shown in Figures 2.1(b) and 2.2(a), respectively. We note from these figures that two cases $(\#20, \#42)$ are potentially influential on the parameter estimates. Figures 2.2(b)-(d) present the index plots of $GD_i^1(\gamma)$, for $\boldsymbol{\gamma} = \boldsymbol{\beta}, \sigma^2, \alpha$, respectively. From these figures we see that observation #42 is influential with regard to the parameters $\boldsymbol{\beta}$ and σ^2, while observation #20 is influential with regard to the parameter α.

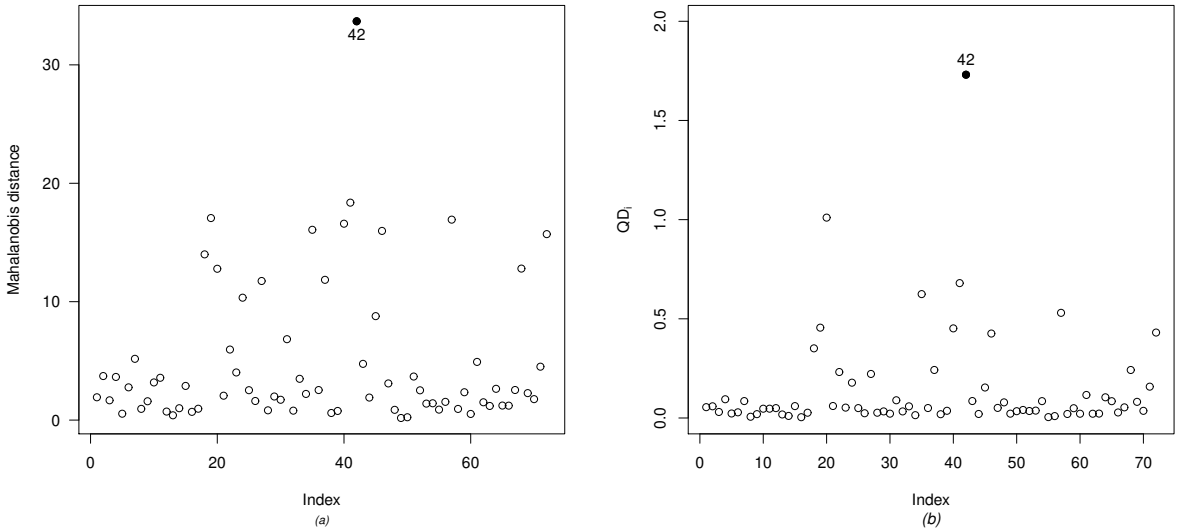


Figura 2.1: UTI data. (a) Mahalanobis distance and (b) Approximate likelihood displacement QD_i^1. The influential observations are numbered.

Local influence

Next, we conduct a local influence study on the UTI data, based on $M(0)$ with interest focussing on $\boldsymbol{\theta}$. Here we use the criterion $M(0)_i > \overline{M}(0) + 3SM(0)$, $i = 1, \ldots, 72$, to discriminate whether an observation is influential or not. Figure 2.3 presents the index plots of $M(0)$ under the four perturbation schemes discussed in 2.4.3.

From this figure it is noted that observation #42 appears as influential under case weight and scale matrix σ^2 perturbation, while observation #20 is more influential under perturbation on the scale matrix $\mathbf{D}$. However, no one observation appears to be influential under response variable perturbation.

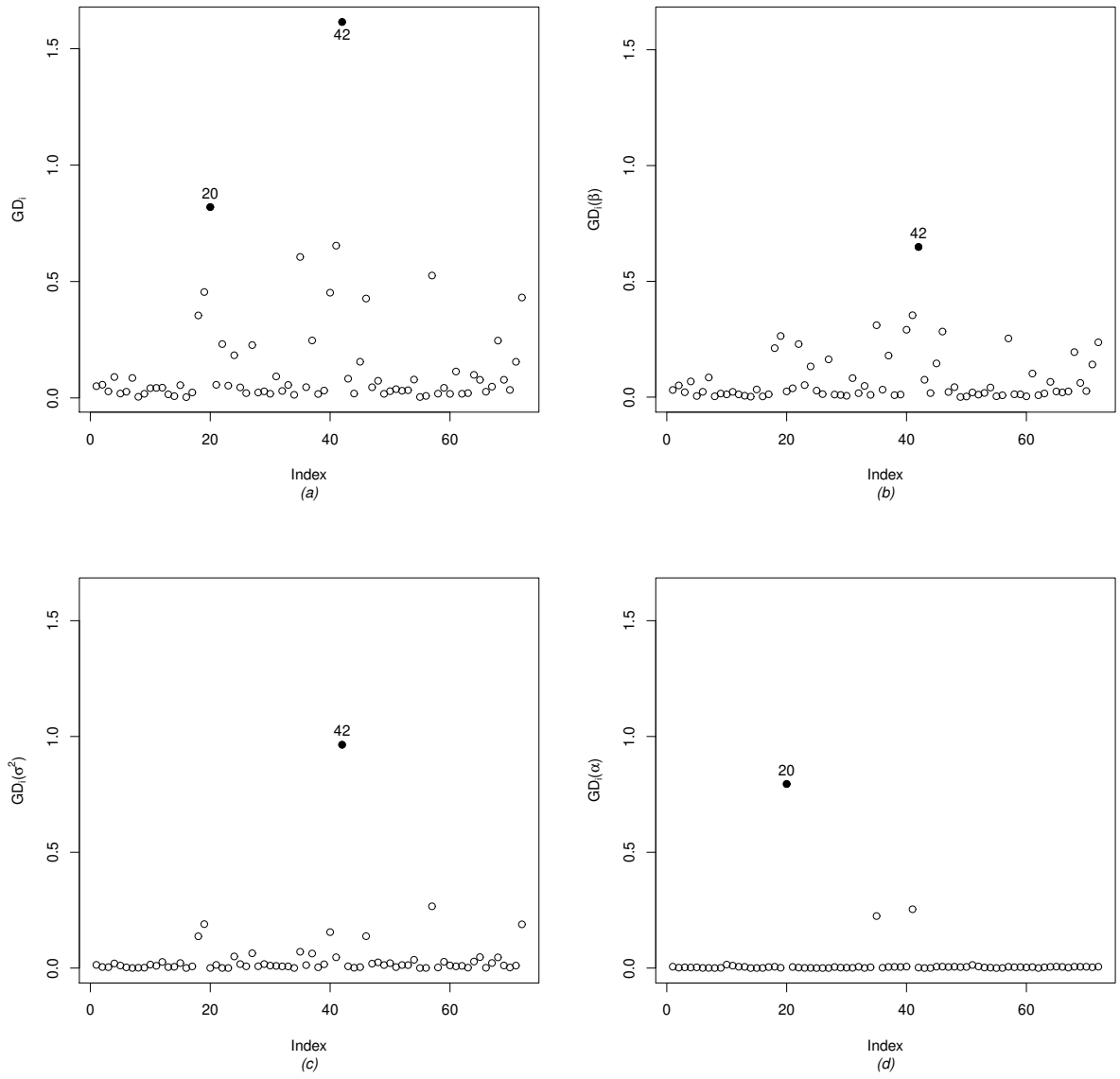


Figura 2.2: UTI data. (a) Approximate generalized Cook distance GD_i^1, (b) GD_i^1 for subset $\boldsymbol{\beta}$, (c) GD_i^1 for subset σ^2 and (d) GD_i^1 for subset $\boldsymbol{\alpha}$. The influential observations are numbered.

In order to assess the impact of the two observations that have been highlighted as potentially influential on the ML estimates, we refitted the proposed LMEC model by

Tabela 2.2: RC (in %) for the UTI data.

Dropped	$RC_{\widehat{\beta_1}}$	$RC_{\widehat{\beta_2}}$	$RC_{\widehat{\beta_3}}$	$RC_{\widehat{\beta_4}}$	$RC_{\widehat{\beta_5}}$	$RC_{\widehat{\beta_6}}$	$RC_{\widehat{\beta_7}}$	$RC_{\widehat{\beta_8}}$	$RC_{\widehat{\sigma^2}}$	$RC_{\widehat{\alpha}}$
$\{\#20\}$	1.28	1.13	1.14	1.14	1.07	0.74	0.71	0.75	0.41	19.07
$\{\#42\}$	0.49	0.44	0.72	1.10	0.29	0.26	0.59	1.04	10.40	0.93
$\{\#20, \#42\}$	0.93	1.69	1.97	2.36	0.89	0.62	0.24	0.16	10.05	18.48

dropping each one of these cases. Let $I_1 = \{20\}$, $I_2 = \{42\}$ and $I_3 = \{20, 42\}$ denotes the sets of observations identified as influential. Table 2.2 presents the relative changes (RC) in percentage of these estimates defined by

$$RC_{\widehat{\gamma}} = \left| \frac{\widehat{\gamma} - \widehat{\gamma}_{[i]}}{\widehat{\gamma}} \right|,$$

where $\gamma = \beta_1, \ldots, \beta_8, \sigma^2, \alpha$ and $\widehat{\gamma}_{[i]}$ denotes the ML estimate of $\widehat{\gamma}$ after the set I_i, $(i = 1, 2, 3)$ has been removed. Even though some RC are large, significants changes in β are not observed. It is of interesting to notice from Table 2.2 the coherence with the diagnostic graphics shown in Figure 2.2 (as we would expect). For instance, the elimination of the observation #20 leads to a large change in the RC of α and elimination of the observation #42 leads to a large change in the RC of σ^2. Moreover, the elimination of the set of observations $\#20, \#42$ leads to a large change in the RC of α and σ^2.

2.6.2 AIEDRP study

The second AIDS case study is from the AIEDRP program, a large multicenter observational study of subjects with acute and early HIV infection. We consider 320 untreated individuals with acute HIV infection; for more details on this dataset see Vaida and Liu (2009). Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification, hence they were considered as right-censored. Following Vaida and Liu (2009), we choose a five-parameter NLME model (inverted S-shaped curve) as follows:

$$y_{ij} = \alpha_{1i} + \frac{\alpha_2}{(1 + \exp((t_{ij} - \alpha_3)/\alpha_4))} + \alpha_{5i}(t_{ij} - 50) + \epsilon_{ij}, \tag{2.23}$$

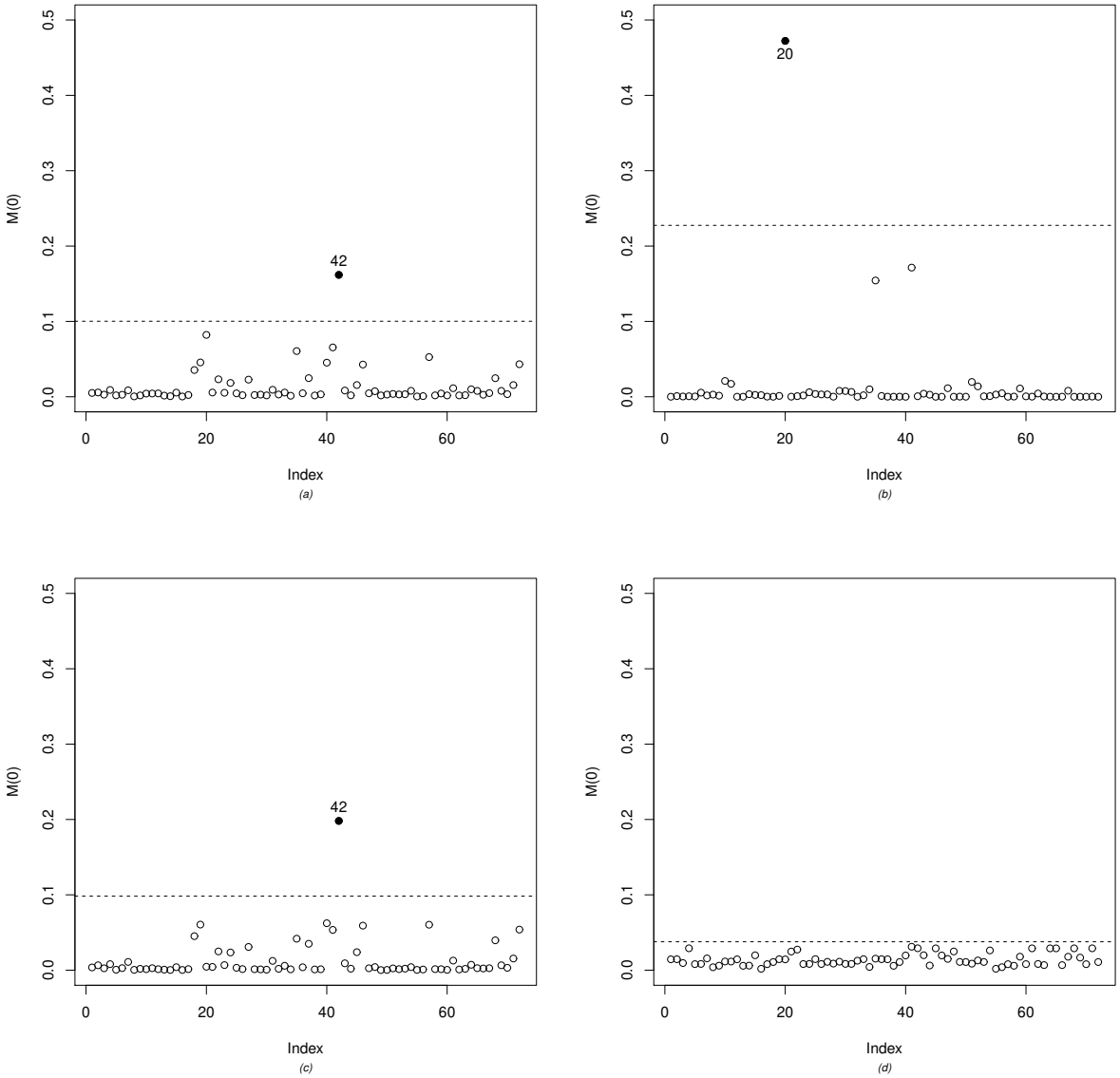


Figura 2.3: UTI data. Index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under (a) Case weight perturbation, (b) Perturbation on $\mathbf{D}$, (c) Perturbation on σ^2 and (d) Perturbation on the response variable for the UTI data. The influential observations are numbered.

where y_{ij} is the $\log_{10}$ HIV RNA for subject i at time t_{ij}. The parameters α_{1i} and α_2 are the setpoint value and the decrease from the maximum HIV RNA. In the absence of treatment (following acute infection), the HIV RNA varies around a set-point which may differ among individuals, hence the set point is chosen to be subject-specific. The location parameter α_3 indicates the time point at which half of the change in HIV RNA

is attained, α_4 is a scale parameter modeling the rate of decline and α_{5i} allows for increasing HIV RNA trajectory after day 50. To force the parameters to be positive, we re-parameterize as follows: $\beta_{1i} = \log(\alpha_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\alpha_k), k = 2, 3, 4$ and $\alpha_{5i} = \beta_5 + b_{2i}$.Table 2.3 lists the ML estimates for the parameters, together with their corresponding standard errors, calculated via Equation (2.12). From Table 2.3, we note that all the regression parameters are significant at 5% level, except the parameter β_2.

Tabela 2.3: Parameter estimates of the NLMEC model and p-values for the AIEDRP data. SE indicates the standard error.

Parameter	Estimate	SE	p-value
$\widehat{\beta_1}$	1.6096	0.0137	<0.01
$\widehat{\beta_2}$	0.1422	0.0949	0.1340
$\widehat{\beta_3}$	3.5262	0.0237	<0.01
$\widehat{\beta_4}$	1.0559	0.2677	0.01
$\widehat{\beta_5}$	-0.0035	0.0014	0.01
$\widehat{\sigma^2}$	0.2652		
$\widehat{\alpha_{11}}$	0.0177		
$\widehat{\alpha_{12}}$	0.0002		
$\widehat{\alpha_{22}}$	0.00004		

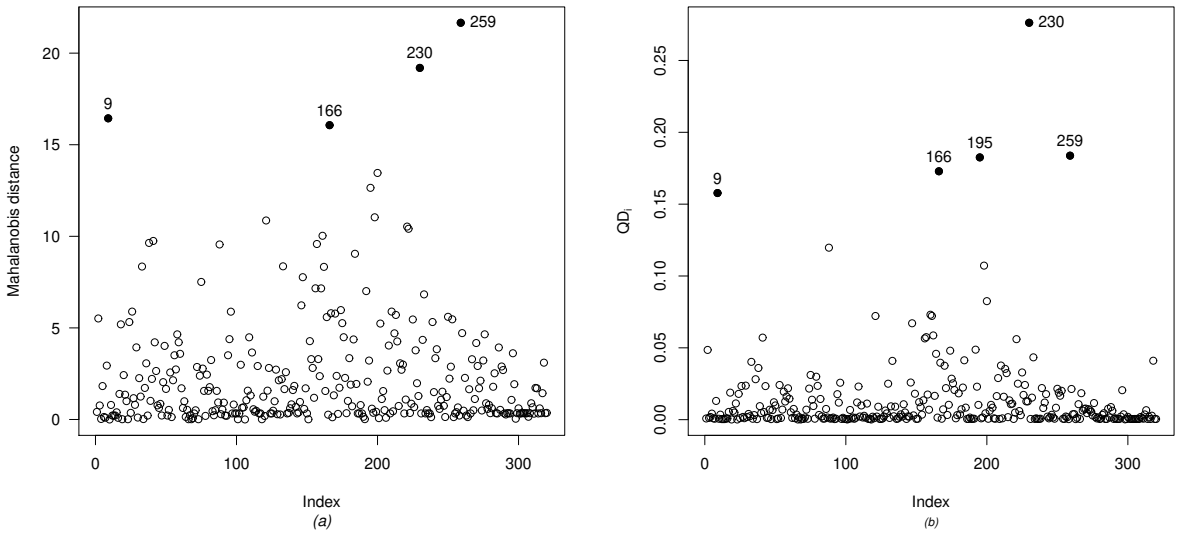


Figura 2.4: AIEDRP data. (a) Mahalanobis distance and (b) Approximate likelihood displacement QD_i^1. The influential observations are numbered.

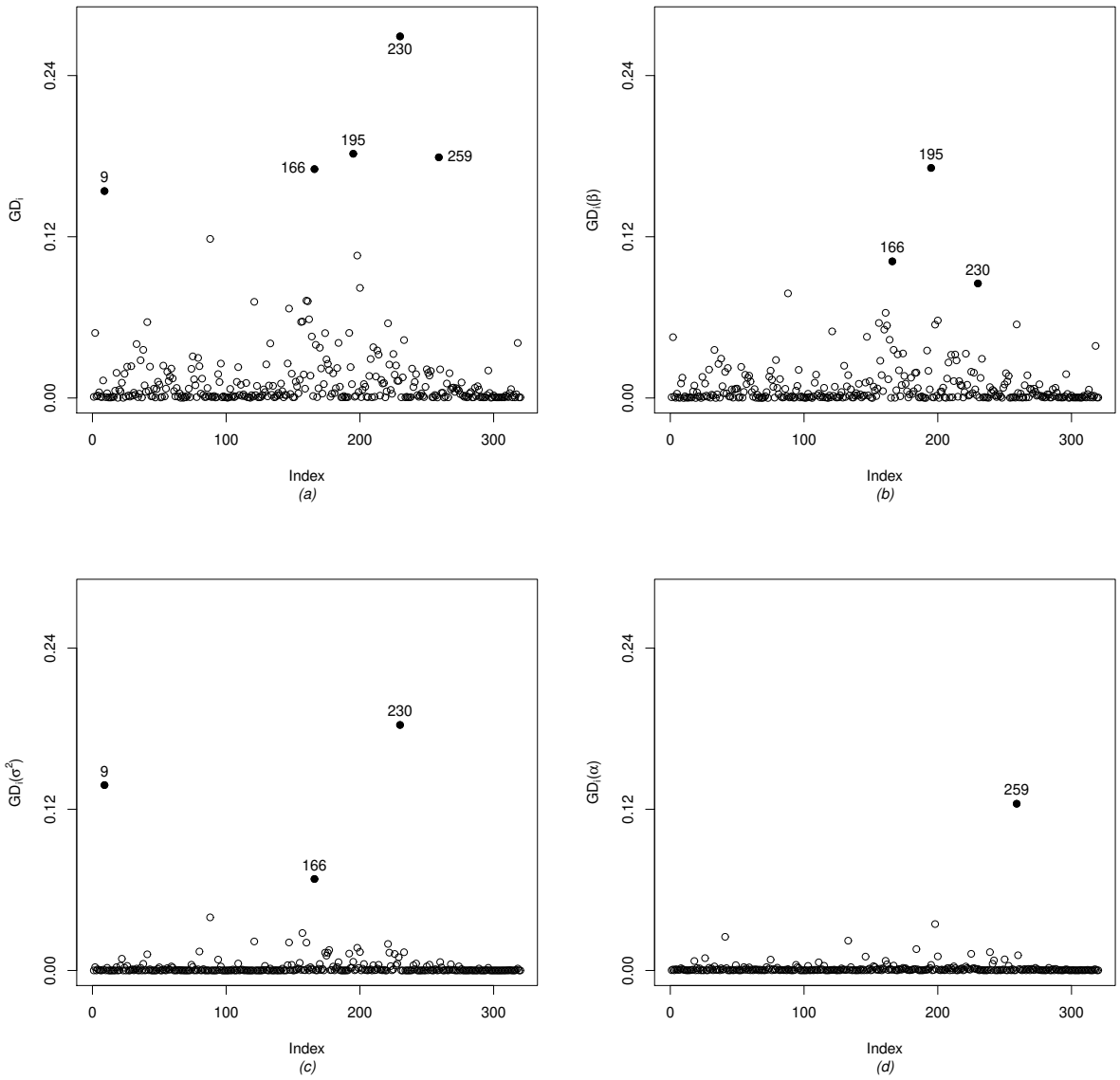


Figura 2.5: AIEDRP data. (a) Approximate generalized Cook distance GD_i^1, (b) GD_i^1 for subset $\boldsymbol{\beta}$, (c) GD_i^1 for subset σ^2 and (d) GD_i^1 for subset $\boldsymbol{\alpha}$. The influential observations are numbered.

Global influence

In order to identify outlying observations under the fitted model, the index plot of the Mahalanobis distance is displayed in Figure 2.4(a). We can see from this figure that observations #9, #166, #230 and #259 appear as possible outliers. As in the previous application, to evaluate the effect on the ML estimates when some

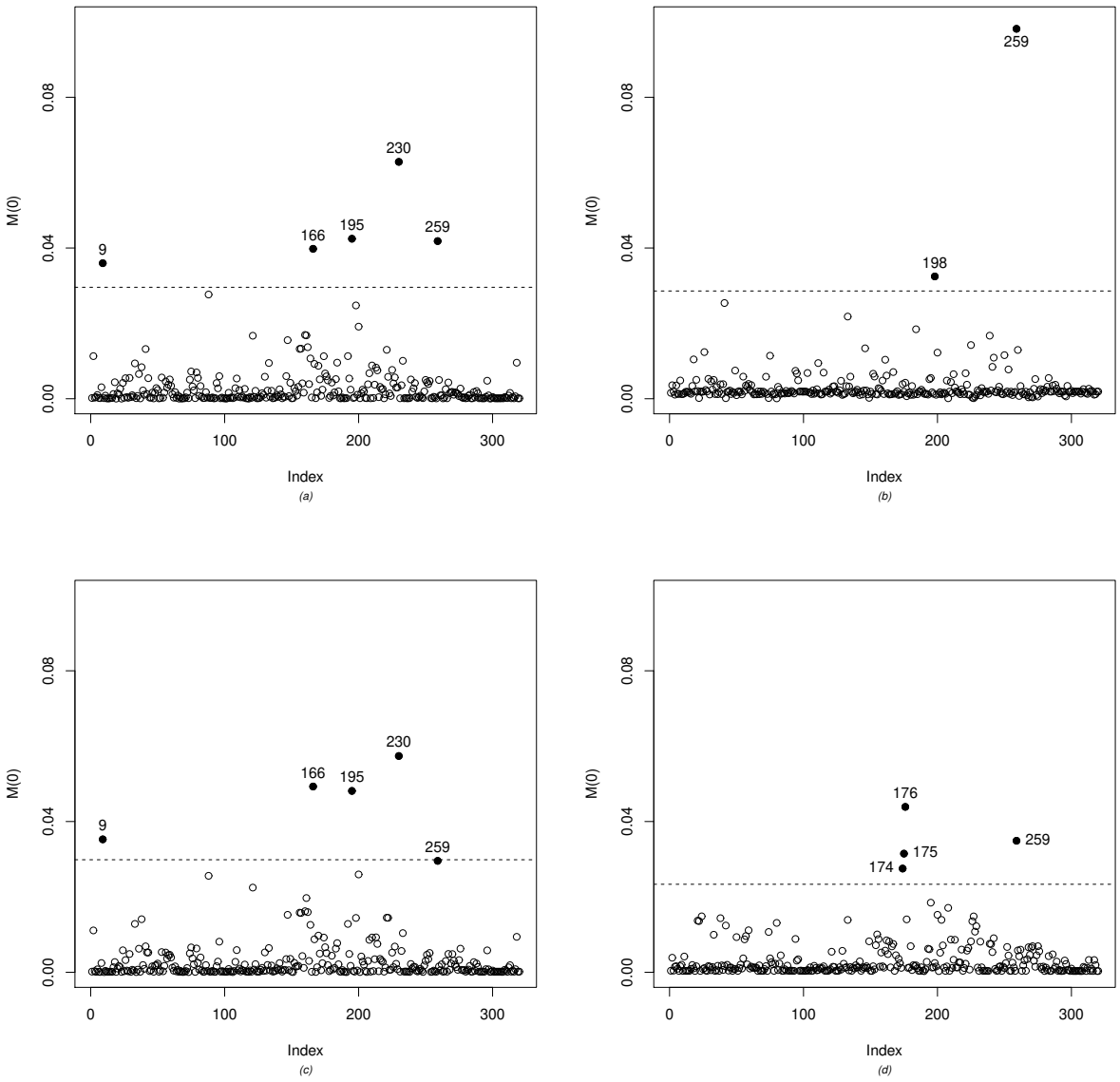


Figura 2.6: AIEDRP data. Index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under (a) Case weight perturbation, (b) Perturbation on $\mathbf{D}$, (c) Perturbation on σ^2 and (d) Perturbation on the response variable. The influential observations are numbered.

observation is eliminated, we analyze the case deletion measures QD_i^1 and GD_i^1, which are shown in Figures 2.4(b) and 2.5(a), respectively. We note from these figures that cases #9, #166, #195, #230 and #259 are all potentially influential with regard to the full parameter estimate $\boldsymbol{\theta}$. On the other hand, from figures 2.5(b)-(d), where we present the index plots of $GD_i^1(\gamma)$, for $\boldsymbol{\gamma} = \boldsymbol{\beta}, \sigma^2, \alpha$, respectively, we can see that observations

#166, #195 and #230 are influential with regard to the regression parameters $\boldsymbol{\beta}$, while only observation #259 is influential with regard to the parameter α.

Local influence

Next, we conduct a local influence study for the AIEDRP data, based on $M(0)$ with interest focussing on $\boldsymbol{\theta}$. Figure 2.6 presents the index plots of $M(0)$ under the four perturbation schemes discussed in 2.4.3. From this figure it is noted that observations #9, #166, #195, #230 and #259 all appear as influential under case weight and scale- σ^2 perturbation, while only observations #198 and #259 are more influential under perturbation on the scale matrix $\mathbf{D}$. It is noted also that different observations #174, #175, #176 and #259 appear out as influential under response variable perturbation. It is important to emphasize that, as in the uncensored case, the influence measure GD_i^1 considered here is closely related to the local influence measure based on the case weight perturbation.

2.6.3 Simulation Study

Results from analysis of a simulated example are presented here to illustrate the performance of the proposed diagnostic measures. We consider a logistic model similar to the one in 2.23, with random set points α_{1i} and random decline rates α_{4i}, as follows

$$y_{ij} = \alpha_{1i} + \frac{\alpha_2}{(1 + \exp((t_{ij} - \alpha_3)/\alpha_{4i}))} + \epsilon_{ij},$$

where $i = 1, \ldots, 100$, $j = 1, \ldots, 10$, $\alpha_{1i} = \exp(\beta_1 + b_{1i})$, $\beta_k = \log(\alpha_k), k = 2, 3$ and $\alpha_{4i} = \exp(\beta_4 + b_{2i})$, $(b_{1i}, b_{2i}) \overset{ind.}{\sim} N_2(\mathbf{0}, \mathbf{D})$ and $\epsilon_{ij} \overset{ind.}{\sim} N_{n_i}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_i})$.
We set $\boldsymbol{\beta} = (1.6094, 0.6931, 3.8067, 2.3026)^\top$, $\sigma^2 = 0.55$, and $\mathbf{D}$ with elements $D_{11} = 0.0025$, $D_{12} = -0.001$ and $D_{22} = 0.0100$. In addition, and twenty percent (20%) of all observations were censored.

After generating y_{ij}, $i = 1, \ldots, 100$, $j = 1, \ldots, 10$, we perturbed the response variable of the individual #85 as follows: $y_{85j} \leftarrow y_{85j} + 1.5$. By using the approach describe in previous sections, we compute for global influence the case deletion measure QD_i^1 and local influence based on $M(0)$ for the response and case weight perturbations.

As expected, we observe from Figure 2.7 the influence of the observation #85, This reveals that the influence measures have detected what they are supposed to detect, but at the same time suggest and give no false influential cases.

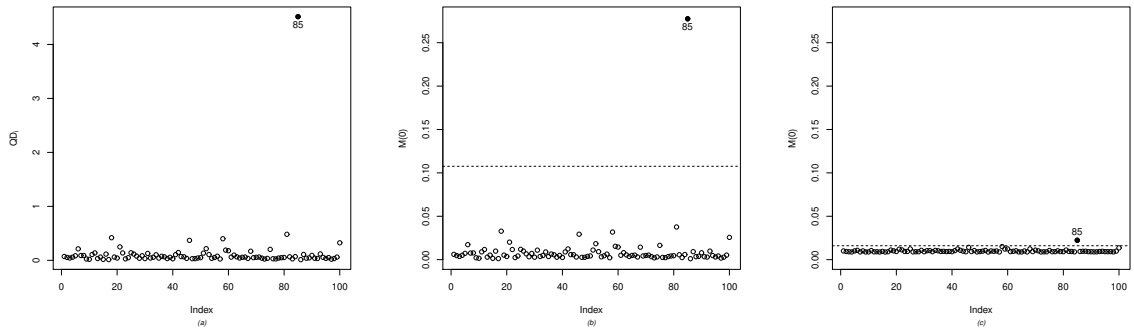


Figura 2.7: Simulated data sets. (a)Approximate likelihood displacement, (b) Case weight perturbation, (c) Perturbation on the response variable. The influential observations are numbered.

2.7 Conclusions

This chapter provides a new insight into classical diagnostics methods for censored linear and nonlinear mixed effects models, typically used for analyzing censored HIV viral load outcomes, and also presents an useful expectation conditional maximization (EMC) algorithm, which enable the development of diagnostics influence measures. Explicit expressions are obtained for the Hessian matrix $\ddot{\mathbf{Q}}$ and for matrix the $\boldsymbol{\Delta}$ under different perturbation schemes. For NLMEC, the analysis is mathematically (and computationally) feasible through a linearization procedure. The proposed methodology has been applied to two recent (left and right-censoring) AIDS studies, which is freely downloadable from `R`. Our findings about the influential observations for these two datasets agree with those presented in Lachos et al. (2011) from a Bayesian perspective.

The proposed methods can be extended to interval-censored longitudinal data, following the work of Sinha et al. (1999). On the other hand, the models developed here do not consider skewness in the responses because typically in HIV-AIDS studies, the responses (censored viral load) is log-transformed to achieve a 'close to normality' shape. However, features of non-normality, such as skewness and thick-tails, need to be incorporated into the proposed methodology to come up with a more general framework

for censored mixed models. So in the next chapter we will propose a robust parametric modeling of LMEC/NLMEC based on the multivariate-t distribution.

Capítulo 3

The Student-t linear and nonlinear mixed-effect models with censored data

3.1 Introduction

In this chapter we propose a robust parametric modeling of LMEC/NLMEC based on the multivatiate-t distribution, so that the t-LMEC/t-NLMEC is defined and a fully likelihood based approach is considered, including the implementation of an exact ECM algorithm for maximum likelihood (ML) estimation. As in Vaida and Liu (2009), we show that the E-step reduces to computing the first two moments of certain truncated multivariate-t distributions. The general formulas for these moments were derived by Lin et al. (2011) (eq. 12 and 13). They require the multivariate-t cumulative density function (cdf), for which we use the *mvtnorm*() package (Genz et al., 2008) in R (R Development Core Team, 2009). The likelihood function is easily computed as a by-product of the E-step and is used for monitoring convergence and for model selection, such as, the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the likelihood ratio test (LR). The methodology has been illustrated with the analysis of two examples involving HIV viral measure and an empirical study.

3.2 The multivariate t and truncated t-distribution

A random variable $\mathbf{Y}$ is said to follow a p-variate t distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ (positive definite) and degrees of freedom ν ($\nu > 0$), denoted by $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, if it can be represented by

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \;\; \mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \; U \sim Gamma(\nu/2, \nu/2), \tag{3.1}$$

where $\mathbf{Z}$ and U are independent and $Gamma(\alpha, \beta)$ stands for a gamma distribution with mean α/β, and density denoted by $G(.|\alpha, \beta)$. We then obtain the probability density function (pdf) of $\mathbf{Y}$, given by

$$t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}}\nu^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\left(1 + \frac{\delta}{\nu}\right)^{-(p+\nu)/2},$$

where $\Gamma(.)$ is the standard gamma function and $\delta = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is the Mahalanobis distance. The cdf will be denoted by $T_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. If $\nu > 1$, $\boldsymbol{\mu}$ is the mean of $\mathbf{Y}$, and if $\nu > 2$, $\nu(\nu-2)^{-1}\boldsymbol{\Sigma}$ is its covariance matrix. As ν tends to infinity, U converges to one with probability one, and so $\mathbf{Y}$ becomes marginally multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The family of t-distributions thus provides a heavy-tailed alternative to the normal family with mean $\boldsymbol{\mu}$ and covariance matrix that is equal to a scalar multiple of $\boldsymbol{\Sigma}$ (if $\nu > 2$). In order to introduce some notation, for a Student-t random vector, we establish the following proposition which is important for our subsequent research.

Proposition 2 *Let* $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *and* $\mathbf{Y}$ *is partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, *with* $dim(\mathbf{Y}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *and* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *be the corresponding partitions of* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\mu}$. *Then*

i) $\mathbf{Y}_1 \sim t_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)$,

ii) The conditional cdf of $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1$ *is given by*

$$P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = T_{p_2}\left(\mathbf{y}_2|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1\right), \tag{3.2}$$

i.e., $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1 \sim t_{p_2}\left(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1\right)$, *where* $\widetilde{\boldsymbol{\Sigma}}_{22.1} = \left(\frac{\nu + \delta_1}{\nu + p_1}\right)\boldsymbol{\Sigma}_{22.1}$, $\delta_1 =$

$(\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, *and* $T_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *denotes the cdf of the p-variate Student-t distribution with parameters* $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ *and* ν.

Proof 1 *The proof of i) is straightforward from (3.1). The proof of (ii), follows from Proposition 4 given in Arellano-Valle and Genton (2010) by setting* $\lambda = \tau = 0$.

Now, let $Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ represent a p-variate truncated t distribution for $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ lying within a right-truncated hyperplane

$$\mathbb{A} = \{\mathbf{x} = (x_1, \ldots, x_p)^\top | x_1 \leq a_1, \ldots, x_p \leq a_p\}. \tag{3.3}$$

Specifically, we say that the p-dimensional vector $\mathbf{X} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, if its density is given by:

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} \mathbb{I}_{\mathbb{A}}(\mathbf{x}), \tag{3.4}$$

where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbb{I}_{\mathbb{A}}(\mathbf{x})$ is the indicator function whose value equals one if $\mathbf{x} \in \mathbb{A}$ and zero elsewhere. The following propositions are crucial for evaluating some conditional expectations of the proposed ECM algorithm for censored mixed effects models.

Proposition 3 *If* $\mathbf{X} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, *with* $\mathbb{A}$ *as defined in (3.3), then*

$$E\left\{\left(\frac{\nu + p}{\nu + \delta}\right)^r \mathbf{X}^{(k)}\right\} = c_p(\nu, r)\frac{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r)}{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} E_{\mathbf{W}}\{\mathbf{W}^{(k)}\}, \; k = 0, 1, 2,$$

where $c_p(\nu, r) = \left(\frac{\nu + p}{\nu}\right)^r \left(\frac{\Gamma((p + \nu)/2)\Gamma((\nu + 2r)/2)}{\Gamma(\nu/2)\Gamma((p + \nu + 2r)/2)}\right)$, $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r; \mathbb{A})$, $\delta = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$, $\mathbf{a} = (a_1, \ldots, a_p)^\top$, $\boldsymbol{\Sigma}^* = \frac{\nu}{\nu + 2r}\boldsymbol{\Sigma}$, $\mathbf{W}^{(0)} = 1$, $\mathbf{W}^{(1)} = \mathbf{W}$, $\mathbf{W}^{(2)} = \mathbf{W}\mathbf{W}^\top$ *and* $\nu + 2r > 0$.

Proof 2 *First note that if* $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, *then we can write*

$$\left(\frac{\nu + p}{\nu + \delta}\right)^r t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = c_p(\nu, r) t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r). \tag{3.5}$$

It follows that

$$E\left\{\left(\frac{\nu + p}{\nu + \delta}\right)^r \mathbf{X}^{(k)}\right\} = c_p(\nu, r)\frac{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r)}{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} E\left\{\mathbf{X}^{(k)}|\mathbf{X} \leq \mathbf{a}\right\},$$

which concludes the proof.

Proposition 4 *Let* $\mathbf{X} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, *with* $\mathbb{A}$ *as defined in (3.3). Consider the partition* $\mathbf{X}^\top = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)$ *with* $dim(\mathbf{X}_1) = p_1$, $dim(\mathbf{X}_2) = p_2$, $p_1 + p_2 = p$, *and the corresponding partition of the parameters* $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{a} = (\mathbf{a}^{x_1}, \mathbf{a}^{x_2})$ *and* $\mathbb{A} = (\mathbb{A}^{x_1}, \mathbb{A}^{x_2})$. *Then under the notation given in Proposition 2 we have*

$$E\left\{\left(\frac{\nu+p}{\nu+\delta}\right)^r \mathbf{X}_2^{(k)}|\mathbf{X}_1\right\} = \frac{d_p(p_1,\nu,r)}{(\nu+\delta_1)^r}\frac{T_{p_2}(\mathbf{a}^{x_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r)}{T_{p_2}(\mathbf{a}^{x_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu+p_1)} E_{\mathbf{W}}\{\mathbf{W}^{(k)}\},$$

where $d_p(p_1,\nu,r) = (\nu+p)^r \left(\frac{\Gamma((p+\nu)/2)\Gamma((p_1+\nu+2r)/2)}{\Gamma((p_1+\nu)/2)\Gamma((p+\nu+2r)/2)}\right)$, $\mathbf{W} \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r; \mathbb{A}^{x_2})$, $\delta = (\mathbf{X}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$, $\delta_1 = (\mathbf{X}_1-\boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1-\boldsymbol{\mu}_1)$, $\mathbf{a}^{x_2} = (a_1,\ldots,a_{p2})^\top$, $\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\frac{\nu+\delta_1}{\nu+2r+p_1}\right)\boldsymbol{\Sigma}_{22.1}$, $\mathbf{W}^{(0)} = 1$, $\mathbf{W}^{(1)} = \mathbf{W}$ $\mathbf{W}^{(2)} = \mathbf{W}\mathbf{W}^\top$, $\nu + p_1 + 2r > 0$ *and* $k = 0, 1, 2$.

Proof 3 *First note that if* $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, *then using the result given in Proposition 2-(ii), we have*

$$\left(\frac{\nu+p}{\nu+\delta}\right)^r t_{p_2}\left(\mathbf{x}_2|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu+p_1\right) = \frac{d_p(p_1,\nu,r)}{(\nu+\delta_1)^r} t_{p_2}(\mathbf{x_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r) \tag{3.6}$$

and the proof concludes by noting that

$$E\left\{\left(\frac{\nu+p}{\nu+\delta}\right)^r \mathbf{X}_2^{(k)}|\mathbf{X}_1\right\} = \frac{d_p(p_1,\nu,r)}{(\nu+\delta_1)^r}\frac{T_{p_2}(\mathbf{a}^{x_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r)}{T_{p_2}(\mathbf{a}^{x_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu+p_1)} E\left\{\mathbf{X}_2^{(k)}|\mathbf{X}_2 \le \mathbf{a}^{x_2}\right\},$$

where $\mathbf{X}_2^{(k)} \sim t_{p_2}\left(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r\right)$.

Formulas for $E\{\mathbf{W}\}$ and $E\{\mathbf{W}\mathbf{W}^\top\}$, where $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, have been recently developed in closed form by Lin et al. (2011) (eq. 12 and 13), which depend on the multivariate-t cdf. The computation uses existing functions for the multivariate-t cumulative distribution, for which the *pmvt*() of the *mvtnorm* library (Genz et al., 2008) from *R* can be used.

3.3 Linear mixed effects with censored response

For robust estimation of the parameters, we proceed as in Pinheiro et al. (2001) by considering a generalization of the classical N–LME as follows:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{3.7}$$

with the assumption that

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{ind.}{\sim} t_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I}_{n_i} \end{pmatrix}, \nu\right), i = 1, \ldots, n, \tag{3.8}$$

where the subscript i is the subject index; $\mathbf{I}_p$ denotes the $p \times p$ identity matrix; $\mathbf{y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ is a $n_i \times 1$ vector of observed continuous responses for sample unit i, $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector of population-averaged regression coefficients called fixed effects, $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$, $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors, and the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$.

From (3.8), it is clear that marginally

$$\mathbf{b}_i \overset{iid}{\sim} t_q(\mathbf{0}, \mathbf{D}, \nu) \quad \text{and} \quad \boldsymbol{\epsilon}_i \overset{iid}{\sim} t_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \nu), \quad i = 1, \ldots, n. \tag{3.9}$$

Note that $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated, once $Cov(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = E\{\mathbf{b}_i\boldsymbol{\epsilon}_i^\top\} = E\{E\{\mathbf{b}_i\boldsymbol{\epsilon}_i^\top|U_i\}\} = \mathbf{0}$, where U_i is defined in 3.1. Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \nu)^\top$ is based on the marginal distribution for $\mathbf{y}_i$, which are marginally distributed as

$$\mathbf{y}_i \overset{\text{ind.}}{\sim} t_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu), \tag{3.10}$$

for $i = 1, \ldots, n$, where $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{ni} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. The estimates from the multivariate t-LME are more robust against outliers than those based on the standard LME. In a simulation study, Pinheiro et al. (2001) showed that the t-LME substantially outperforms the normal or standard LME when outliers are present in the data. The gains in efficiency in estimating the parameter is particularly high for the variance - covariance parameters.

This problem has been also discussed by Wu (2010) in the context of censored mixed effects models.

Following Vaida and Liu (2009), we consider the case in which the response Y_{ij} is not fully observed for all i, j. Thus, let the observed data for the i-th subject be $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i$ represents the vector of uncensored reading or censoring level, and $\mathbf{C}_i$ the vector of censoring indicators:

$$\begin{aligned} y_{ij} &\leq V_{ij} \quad \text{if} \quad C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{if} \quad C_{ij} = 0, \end{aligned} \tag{3.11}$$

so that, the t-LMEC is defined. For simplicity we will assume that the data are left-censored. The extensions to arbitrary censoring are immediate. It follows that for responses with censoring pattern as in (3.11), we have that marginally

$$\mathbf{y}_i \sim Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu; \mathbb{A}_i),$$

where $\mathbb{A}_i = A_{i1} \times \ldots, \times A_{ini}$, with A_{ij} as the interval $(-\infty, \infty)$ if $C_{ij} = 0$ and the interval $(-\infty, V_{ij}]$ if $C_{ij} = 1$.

3.3.1 The likelihood function

The likelihood function, can be easily computed by using a sequence of simple steps. The first step is to treat separately the observed and censored components of $\mathbf{y}_i$. Partition $\mathbf{y}_i$ into the observed and censored parts: $\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c)$, that is, $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$, and 1 for all elements in $\mathbf{y}_i^c$; write accordingly $\mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $vec(.)$ denote the function which stacks vectors or matrices of the same number of columns. Then, from Proposition 2, we have that $\mathbf{y}_i^o \sim t_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}, \nu)$, $\mathbf{y}_i^c|\mathbf{y}_i^o, \sim t_{n_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta}), \tag{3.12}$$

$$\mathbf{S}_i^{co} = \left(\frac{\nu + \delta(\mathbf{y}_i^o)}{\nu + n_i^o}\right)\boldsymbol{\Sigma}_i^{cc.o}, \tag{3.13}$$

with $\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}\boldsymbol{\Sigma}_i^{oc}$ and $\delta(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$. Thus, the likelihood for cluster i is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i|\boldsymbol{\theta}) = f(\mathbf{y}_i^o|\boldsymbol{\theta}) f(\mathbf{y}_i^c|\mathbf{y}_i^o, \boldsymbol{\theta}) = f(\mathbf{y}_i^o|\boldsymbol{\theta}) P(\mathbf{y}_i^c \leq \mathbf{V}_i^c|\mathbf{y}_i^o, \boldsymbol{\theta}) \\ &= t_{n_i^o}(\mathbf{V}_i^o|\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}, \nu) T_{n_i^c}(\mathbf{V}_i^c|\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o) = L_i. \end{aligned} \tag{3.14}$$

Therefore, the log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \{\log L_i\}$. This can be computed at each step of the EM-type algorithm without additional computational burden, because L_i's are computed at the E-step (see subsection 3.3.2). In addition, the log-likelihood can be used to monitor the convergence of the algorithm and for model selection (AIC, BIC, LR).

Lucas (1997) developed an interesting study on the robust aspects of the Student-t M-estimator in the univariate case using influence functions. He showed that the protection against outliers is preserved only if the degrees of freedom parameter is fixed. Otherwise, if the degrees of freedom is also estimated by maximum likelihood, the influence functions for ν and the change of variance function of the location parameter are not bounded. In this work we will maintain fixed the degrees of freedom and the shape parameters for Student-t, and we will use a model selection procedure based on the AIC or BIC to choose the most appropriate values of ν (see Lange et al., 1989; Meza and Osorio, 2011). Thus, hereafter we consider that the parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$.

3.3.2 The EM algorithm

The EM algorithm originally proposed by Dempster, Laird and Rubin (1977) has several appealing features such as stability of monotone convergence with each iteration increasing the likelihood and simplicity of implementation. However, ML estimation in model (3.7), (3.8) and (3.11) is complicated such that the EM algorithm is less advisable due to a computational difficulty in the M-step. To cope with this problem, we apply an extension of EM algorithm, called the ECM (Meng and Rubin, 1993) algorithm, which shares the appealing features of the EM and has a typically faster convergence rate than the EM in the sense of a small amount of iterations or actual computer time.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$, such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i-th subject. Treating $\mathbf{b}$, $\mathbf{u}$ and $\mathbf{y}$ as hypothetical missing data, and augmented with the observed data $\mathbf{V}, \mathbf{C}$, we set $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$. Hence, the ECM algorithm is applied to the complete-data

log-likelihood function $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\mathbf{y}_c)$, given by

$$\begin{aligned}\ell_i(\boldsymbol{\theta}|\mathbf{y}_c) &= -\frac{1}{2}\Big[n_i \log\sigma^2 + \frac{u_i}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ &\quad + \log|\mathbf{D}| + u_i\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\Big] + h(u_i|\nu) + C, \end{aligned} \tag{3.15}$$

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$ and $h(u_i|\nu)$ is a density of a $Gamma(\nu/2, \nu/2)$. Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete log-likelihood function given by (see appendix)

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \\ &= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}), \end{aligned} \tag{3.16}$$

where

$$\begin{aligned} Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\Big[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}) \\ &\quad + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\Big] \end{aligned}$$

and

$$Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathrm{tr}\left(\widehat{u\mathbf{b}_i^2}^{(k)}\mathbf{D}^{-1}\right),$$

with $\widehat{a}_i^{(k)} = \mathrm{tr}\left(\widehat{u\mathbf{y}_i^2}^{(k)} - 2\widehat{u\mathbf{y}\mathbf{b}_i}^{(k)}\mathbf{Z}_i^\top + \widehat{u\mathbf{b}_i^2}^{(k)}\mathbf{Z}_i^\top\mathbf{Z}_i\right)$; $\widehat{u\mathbf{b}_i^2}^{(k)} = E\{u_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\sigma^2}^{(k)}\widehat{\boldsymbol{\Lambda}}_i^{(k)} + \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{u\mathbf{y}}_i^{(k)\top} + \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$; $\widehat{u\mathbf{b}_i}^{(k)} = E\{u_i\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}}_i^{(k)} - \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})$; $\widehat{u\mathbf{y}\mathbf{b}_i}^{(k)} = E\{u_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = (\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$, where $\widehat{\boldsymbol{\Lambda}}_i^{(k)} = (\widehat{\sigma^2}^{(k)}\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top\mathbf{Z}_i)^{-1}$ and $\widehat{\boldsymbol{\varphi}}_i^{(k)} = \widehat{\boldsymbol{\Lambda}}_i^{(k)}\mathbf{Z}_i^\top$.

Note that in this case we do not consider the computation of $E[h(u_i|\nu)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$, because ν is fixed.

The conditional maximization (CM) steps then conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ and obtains a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described below:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^{n} \widehat{u}_i^{(k)} \mathbf{X}_i^\top \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \left(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{u\mathbf{b}}_i^{(k)}\right), \tag{3.17}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^{n}\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}) + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right], \tag{3.18}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\widehat{u\mathbf{b}_i^2}^{(k)}, \tag{3.19}$$

where $N = \sum_{i=1}^{n} n_i$. This process is iterated until some distance involving two successive evaluations of the log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y})$ described in subsection 3.3.1, like $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})|$ or $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, is small enough. That is, convergence is declared when the improvement in log-likelihood falls below a certain preset limit. In practice, *pmvt*() shows small random variability, which leads to non-increasing log-likelihood beyond a certain level. The variability due to *pmvt*() can be controlled using the *algorithm* = *GenzBretz*(value) argument.

From (3.17)-(3.19) it is clear that the E-step reduces only to the computation of $\widehat{u\mathbf{y}_i^2}$, $\widehat{u\mathbf{y}}_i$ and $\widehat{u}_i$. These expected values can be determined in closed form, using propositions 2-4, as follows.

1. If $\mathbf{y}_i = \mathbf{y}_i^c$, i.e, the individual i has only censored components. Then from Proposition 3, we have:

$$\begin{aligned}
\widehat{u\mathbf{y}_i^2} &= E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = \frac{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i^*, \nu+2)}{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu)} E\{\mathbf{W}_i\mathbf{W}_i^\top\},\\
\widehat{u\mathbf{y}}_i &= E\{u_i\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = \frac{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i^*, \nu+2)}{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu)} E\{\mathbf{W}_i\},\\
\widehat{u}_i &= E\{u_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = \frac{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i^*, \nu+2)}{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu)},
\end{aligned}$$

where $\mathbf{W}_i \sim Tt_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i^*, \nu+2; \mathbb{A}_i)$, $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i^* = \frac{\nu}{\nu+2}\boldsymbol{\Sigma}_i$, $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{ni} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$ and $\mathbb{A}_i = \{\mathbf{W}_i = (w_1, \ldots, w_{n_i})^\top | w_1 \leq V_{i1}, \ldots, w_{n_i} \leq V_{in_i}\}$.

2. If $\mathbf{y}_i = \mathbf{y}_i^o$, i.e, the individual i has non censored components. Then,

$$\widehat{u\mathbf{y}_i^2} = \frac{\nu+n_i}{\nu+\delta(\mathbf{y}_i)}\mathbf{y}_i\mathbf{y}_i^\top,\ \widehat{u\mathbf{y}}_i = \frac{\nu+n_i}{\nu+\delta(\mathbf{y}_i)}\mathbf{y}_i,\ \widehat{u}_i = \frac{\nu+n_i}{\nu+\delta(\mathbf{y}_i)},$$

where $\delta(\mathbf{y}_i) = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, and finally

3. If $\mathbf{y}_i = (\mathbf{y}_i^{c\top}, \mathbf{y}_i^{o\top})^\top$, i.e., for individual i, we observed censored and uncensored

components. Then from Proposition 4 and by the fact that $\{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i\}$, $\{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\mathbf{y}_i^o\}$ and $\{\mathbf{y}_i^c|\mathbf{V}_i,\mathbf{C}_i,\mathbf{y}_i^o\}$ are equivalent processes, we have

$$\begin{aligned}
\widehat{u\mathbf{y}_i^2} &= E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{y}_i^o,\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\} = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top}\widehat{u}_i & \widehat{u}_i\mathbf{y}_i^o\widehat{\mathbf{w}}_i^{c\top} \\ \widehat{u}_i\widehat{\mathbf{w}}_i^c\mathbf{y}_i^{o\top} & \widehat{u}_i\widehat{\mathbf{w}^{\mathbf{2}}}_i^c \end{pmatrix}, \\
\widehat{u\mathbf{y}}_i &= E\{u_i\mathbf{y}_i|\mathbf{y}_i^o,\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\} = vec(y_i^o\widehat{u}_i,\widehat{\mathbf{w}}_i^c), \\
\widehat{u}_i &= E\{u_i|\mathbf{y}_i^o,\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\} = \left(\frac{n_i^o+\nu}{\nu+\delta(\mathbf{y}_i^o)}\right)\frac{T_p(\mathbf{V}_i|\boldsymbol{\mu}_i^{co},\widetilde{\mathbf{S}}^{co},\nu+n_i^o+2)}{T_p(\mathbf{V}_i|\boldsymbol{\mu}_i^{co},\mathbf{S}^{co},\nu+n_i^o)},
\end{aligned}$$

where $\widetilde{\mathbf{S}}^{co} = \left(\frac{\nu+\delta(\mathbf{y}_i^o)}{\nu+2+n_i^o}\right)\boldsymbol{\Sigma}_i^{cc.o}$, $\widehat{\mathbf{w}}_i^c = E\{\mathbf{W}_i\}$ and $\widehat{\mathbf{w}^{\mathbf{2}}}_i^c = E\{\mathbf{W}_i\mathbf{W}_i^\top\}$, with $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co},\widetilde{\mathbf{S}}^{co},\nu+n_i^o+2;\mathbb{A}_i^c)$ and $\boldsymbol{\Sigma}_i^{cc.o}$, $\boldsymbol{\mu}_i^{co}$ and $\mathbf{S}^{co}$ are as in (3.12)-(3.13).

3.3.3 Estimation of random effects and the expected information matrix

In this subsection we consider an empirical Bayes inference for the random effects, that is, the minimum mean squared error (MSE) predictor of $\mathbf{b}_i$, that is useful for evaluating subject-specific quantities such as individual intercepts and slopes. Thus, if values of parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top,\sigma^2,\boldsymbol{\alpha}^\top)^\top$ and ν were known, the conditional mean of $\mathbf{b}_i$ given $\mathbf{C}_i$, $\mathbf{V}_i$ is

$$\begin{aligned}
\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) &= E\{\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i\} = E\{E\{E\{\mathbf{b}_i|u_i\}|\mathbf{y}_i,u_i\}|\mathbf{V}_i,\mathbf{C}_i\} \\
&= E\{\boldsymbol{\Lambda}_i\mathbf{Z}_i^\top(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})|\mathbf{V}_i,\mathbf{C}_i\} = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top(\widehat{\mathbf{y}}_i-\mathbf{X}_i\boldsymbol{\beta}),
\end{aligned} \tag{3.20}$$

where $\boldsymbol{\Lambda}_i$ is defined in subsection 3.3.2 and $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i\}$ is the first moment of the truncated multivariate-t distribution $(Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta},\boldsymbol{\Sigma}_i,\nu;\mathbb{A}_i))$. In practice, the empirical Bayes estimators of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimate $\widehat{\boldsymbol{\theta}}$ into (3.20), which leads to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\theta}})$. The conditional covariance matrix of $\mathbf{b}_i$ given $\mathbf{C}_i$, $\mathbf{V}_i$ is

$$\begin{aligned}
Var\{\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i\} &= E\{\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i\} - \widehat{\mathbf{b}}_i(\boldsymbol{\theta})\widehat{\mathbf{b}}_i(\boldsymbol{\theta})^\top \\
&= \frac{\nu+n_i}{\nu+n_i-2}E\left\{(\frac{\nu+n_i}{\nu+\delta(\mathbf{y}_i)})^{-1}|\mathbf{V}_i,\mathbf{C}_i\right\}\boldsymbol{\Lambda}_i\sigma^2 + \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top(\widehat{\mathbf{y}_i^2}-\widehat{\mathbf{y}}_i\widehat{\mathbf{y}}_i^\top)\mathbf{Z}_i\boldsymbol{\Lambda}_i,
\end{aligned}$$

where $\widehat{\mathbf{y}_i^2} = E\{\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i\}$ is the second moment of the truncated multivariate-t distribution $(Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta},\boldsymbol{\Sigma}_i,\nu;\mathbb{A}_i))$ These expected values can be easily accomplished from steps [1]-[3] given above as a by-product of our proposed ECM algorithm (E-step).

Louis (1982) derives a result that can be used to adjust the variances of the estimated fixed effects for the information lost due to censoring. Using this method, from the results given in Appendix B in Lange et al. (1989), an asymptotic approximation for the variances of the fixed effects is given by (see Appendix A.2):

$$\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = Var(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \frac{\nu + n_i}{\nu + n_i + 2} \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - \sum_{i=1}^{n} \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{B}_i \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1}, \tag{3.21}$$

where $\mathbf{B}_i = Var\left\{ \frac{\nu + n_i}{\nu + \delta(\mathbf{y}_i)} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) | \mathbf{V}_i, \mathbf{C}_i \right\}$, with $\mathbf{y}_i \sim Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu; \mathbb{A}_i)$. Asymptotic confidence intervals and hypothesis tests for the fixed effects are obtained assuming that the MLE $\boldsymbol{\beta}$ has approximately a $N_p(\boldsymbol{\beta}, \mathbf{J}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}})$ distribution. In practice, $\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is usually unknown and has to be replaced by its MLE $\widehat{\mathbf{J}}_{\boldsymbol{\beta}\boldsymbol{\beta}}$.

3.4 The nonlinear case

Extending the notation of the previous section and ignoring censoring, we first propose the following general mixed-effects model in which the random terms are assumed to follow a multivariate-t distribution (t-NLME).

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ denote the (continuous) response vector for subject i and $\eta = (\eta(\mathbf{X}_{i1}, \boldsymbol{\phi}_i), \ldots, \eta(\mathbf{X}_{in_i}, \boldsymbol{\phi}_i))^\top$ be a nonlinear vector valued differentiable function of the individuals random parameter $\boldsymbol{\phi}_i$ and a vector of covariates $\mathbf{X}_i$. The t-NLME can then be expressed as:

$$\mathbf{y}_i = \eta(\boldsymbol{\phi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\phi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \tag{3.22}$$

where the joint distribution of $(\mathbf{b}_i, \boldsymbol{\epsilon}_i)$ is as in (3.8), $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$ respectively, possibly depending on some covariable values, $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of fixed effects, $\mathbf{b}_i$ is the $(q \times 1)$ vector of random effects. Thus, from the properties of the multivariate-t distribution, we have that marginally, $\boldsymbol{\phi}_i \overset{ind}{\sim} t_r(\mathbf{A}_i\boldsymbol{\beta}, \mathbf{B}_i\mathbf{D}\mathbf{B}_i^\top, \nu)$ and $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} t_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \nu)$, and as in the linear case, they are uncorrelated because $\text{Cov}(\boldsymbol{\phi}_i, \boldsymbol{\epsilon}_i) = \mathbf{0}$. For NI-NLME with non censoring responses, the marginal distribution is given by

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \int_0^\infty \int_{\mathbb{R}^q} \phi_{n_i}(\mathbf{y}_i; \eta(\boldsymbol{\phi}_i, \mathbf{X}_i), u_i^{-1}\sigma^2\mathbf{I}_{n_i}) \phi_q(\boldsymbol{\phi}_i; \mathbf{A}_i\boldsymbol{\beta}, u_i^{-1}\mathbf{B}_i\mathbf{D}\mathbf{B}_i^\top) \times G(u_i|\nu/2, \nu/2) d\boldsymbol{\phi}_i du_i, \tag{3.23}$$

which generally does not have a closed form expression because the model function is not

linear in the random effect. In the normal case, various approximations (viz. first-order Taylor series expansion of the model function around the conditional mode of $\mathbf{b}_i$, says $\widetilde{\mathbf{b}}_i$) have been proposed to achieve tractable numerical optimizations (Wu, 2010). Most algorithms for computing the approximate MLE $\widehat{\boldsymbol{\theta}}$ and empirical Bayes estimators (predictors) for the random effects $\widehat{\mathbf{b}}_i$ considers iterative maximization of the approximate log-likelihood functions $\ell(\boldsymbol{\theta}, \widetilde{\mathbf{b}}) = \sum_{i=1}^{n} \log f(\mathbf{y}_i|\boldsymbol{\theta}, \widetilde{\mathbf{b}}_i)$. Following Taylor series expansions, we have the following theorems. The first uses a point in a neighborhood of the conditional mode $\widetilde{\mathbf{b}}_i$ as the expansion point and it has been proven useful for implementation of model selection, in a Bayesian context (Lachos et al., 2011). The second, useful for the implementation of the EM algorithm, uses simultaneously neighborhood of $\mathbf{b}_i$ and $\boldsymbol{\beta}$ as expansions points, with the advantage that the likelihood is completely linearized (in $\mathbf{b}_i$ and $\boldsymbol{\beta}$). We call these LME approximations and can be considered as extensions of the result given in Lindstrom and Bates (1990) and Pinheiro and Bates (2000) for the Student-t case.

Theorem 2 *Let $\widetilde{\mathbf{b}}_i$ be an expansion point in a neighborhood of $\mathbf{b}_i$, then under the t–NLME model as in (3.22), the marginal distribution of $\mathbf{y}_i$, can be approximated as $\mathbf{y}_i \dot{\sim} t_{ni}(\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i, \widetilde{\mathbf{V}}_i, \nu)$, where $\widetilde{\mathbf{V}}_i = \widetilde{\mathbf{H}}_i\mathbf{D}\widetilde{\mathbf{H}}_i^{\top} + \sigma^2\mathbf{I}_{n_i}$,*
$\widetilde{\mathbf{H}}_i = \dfrac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^{\top}}|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$ *and $\dot{\sim}$ denotes approximated in distribution.*

Proof 4 *See Lachos et al. (2011).*

The next theorem allows the implementation of the EM algorithm.

Theorem 3 *Let $\widetilde{\mathbf{b}}_i$ and $\widetilde{\boldsymbol{\beta}}$ be expansion points in a neighborhood of $\mathbf{b}_i$ and $\boldsymbol{\beta}$, respectively, then under the t-NLME model as (3.22)–(3.8), we have the following linearized model*

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{3.24}$$

where $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \widetilde{\eta}(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i)$, $\mathbf{b}_i \overset{ind}{\sim} t_q(0, \mathbf{D}, \nu)$ and $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} t_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \nu)$, $\widetilde{\mathbf{H}}_i = \dfrac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^{\top}}|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$ and $\widetilde{\mathbf{W}}_i = \dfrac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\boldsymbol{\beta}_i^{\top}}|_{\boldsymbol{\beta}_i=\widetilde{\boldsymbol{\beta}}_i}$ and $\widetilde{\eta}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i) = \eta(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}$,

Proof 5 *Based on first-order Taylor expansion of the function η around $\widetilde{\mathbf{b}}_i$ and $\widetilde{\boldsymbol{\beta}}$, we have that*

$$\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i) \approx [\eta(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) + \widetilde{\mathbf{H}}_i\mathbf{b}_i - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i + \widetilde{\mathbf{W}}_i\boldsymbol{\beta} - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}]$$

with $\widetilde{\mathbf{H}}_i = \dfrac{\partial \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial \mathbf{b}_i^\top}\Big|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$ *and* $\widetilde{\mathbf{W}}_i = \dfrac{\partial \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial \boldsymbol{\beta}_i^\top}\Big|_{\boldsymbol{\beta}_i=\widetilde{\boldsymbol{\beta}}_i}$. *It follows that*

$$\begin{aligned}\boldsymbol{\epsilon}_i &= \mathbf{y}_i - \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i) \approx \mathbf{y}_i - [\eta(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) + \widetilde{\mathbf{H}}_i\mathbf{b}_i - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i + \widetilde{\mathbf{W}}_i\boldsymbol{\beta} - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}] \\ &= \mathbf{y}_i - [\widetilde{\eta}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i) + \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i] = \widetilde{\mathbf{y}}_i - [\widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i],\end{aligned}$$

which concludes the proof.

The empirical Bayes estimates of the random effects $\widetilde{\mathbf{b}}$, given in (3.20), can be used iteratively in the linearization procedure from Theorem 2. Note that the distribution of $\mathbf{b}_i|\mathbf{y}_i$ is approximately symmetric (Student-t), and thus $\widetilde{\mathbf{b}}_i$ is the mode of the distribution at each step. As commented by Vaida and Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ consists to iteratively solving the LME model (L-step) in (3.24). For censored response the linearized model (3.24) is an LME with censored data, with same structure as (3.7)-(3.8), which is then solved as indicated in the previous section. The model matrices in (3.24) depends on the current parameter value, and needs to be recalculated at each iteration. The algorithm iterates to convergence between L-, E-, and CM-steps.

3.5 Model choice

A variety of information criteria exist to properly determine the best choice among a set of competing models. To identify the best selected model support by the data, we adopt the AIC and the BIC, which are the two most commonly used model selection tools. Both criteria can be applied to non-nested and to nested models, but not always lead to the same choice. Basically, there is no clear consensus regarding which criterion is better to use. A combined use of AIC and BIC would be of help to screening reasonable candidate models.

A formal test concerning the appropriateness of using the normal model $H_0 : \nu^{-1} = 0$ versus t model $H_1 : \nu^{-1} > 0$ is nontrivial since the null hypothesis is on the boundary of the parameter space. For testing parameters under non-standard settings, Self and Liang (1987) have shown the limiting distribution of the likelihood ratio test (LR) statistic will follow a mixture of chi-square distributions. Referring to Case 5 of Self and Liang (1987), the LR statistic under $H_0 : \nu^{-1} = 0$ is an equally weighted mixture of χ^2_0 and χ^2_1 distributions, where χ^2_0 denotes a degenerate distribution with all of its mass or probability at zero. In this case, the critical values are 1.65, 2.71 and 5.41 at the 10%, 5% and 1% significance levels, respectively.

3.6 Application

We illustrate the performance of the proposed methods with the analysis of two HIV datasets, previously analyzed in chapter 2, and the analysis of a simulated example.

3.6.1 UTI data

The first application is the same study used in 2.6.1, a study of 72 perinatally HIV-infected children (Saitoh et al., 2008).

Vaida and Liu (2009) and we analyzed the same dataset by fitting a similar N-LMEC via EM algorithm, but from Figure 1 given in Lachos et al. (2011) it is clear that inference based on normality assumptions are questionable (presence of thick tails). We revisit the UTI data with the aim of providing robust inferences, from a frequentist perspective, by using the Student-t distribution. The ML estimates were obtained using the ECM algorithm described in subsection 3.3.2. Starting values were obtained by using the library *lmec*.

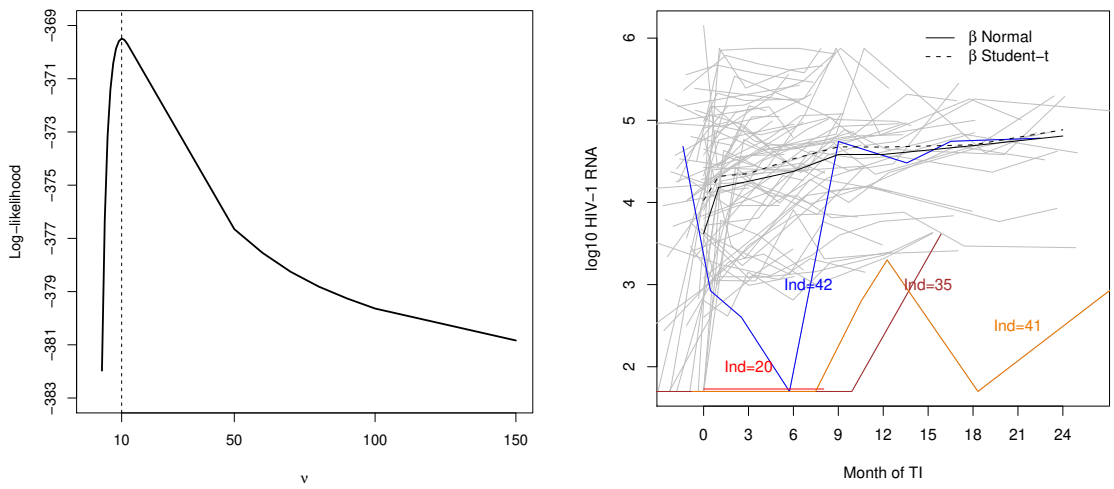


Figura 3.1: UTI data. (Left panel) Plot of the profile log-likelihood of the degrees of freedom ν. (Right panel) Individual profiles and overall mean (in $\log_{10}$ scale) using the Normal and t distributions for HIV viral load at different follow-up times. The trajectories for the influential individuals are numbered.

For the Student-t model, we assumed that the degree of freedom ν is know and by using the AIC criterion we found $\nu = 10$ (see left panel in Figure 1). It is a first indication that the normal model is inadequate. Table 3.1 presents the ML estimate of $\boldsymbol{\theta}$ and the corresponding

Tabela 3.1: ML estimates under normal and Student-t models fitted to the UTI data. SE are the corresponding standard errors.

	N-LMEC		T-LMEC	
Parameter	estimate	SE	estimate	SE
β_1	3.6038	0.1253	3.6182	0.1238
β_2	4.1664	0.1285	4.2532	0.1311
β_3	4.2413	0.1304	4.3137	0.1332
β_4	4.3604	0.1307	4.4580	0.1338
β_5	4.5662	0.1398	4.6229	0.1435
β_6	4.5692	0.1485	4.6112	0.1532
β_7	4.6773	0.1646	4.6978	0.1709
β_8	4.7935	0.2018	4.7874	0.2111
σ^2	0.3414		0.3503	
α	0.7653		0.6662	
ν	-	-	10	-
AIC	844.1172		759.0148	
BIC	883.0337		797.9312	

standard errors of the fixed effects. Comparing these values we notice a similarity between the estimates under normal and Student-t models. Additionally, the inferences for the variance components are similar for the two models, but are not comparable since they are on different scales. According to the AIC or BIC values, given at the bottom of Table 3.1, we notice also that the t-LMEC model perform better than the N-LMEC model. For the LR statistics described in Subsection 3.5, we have that the maximum log-likelihood for the N-LMEC model is -412.059 and for the t-LMEC model is -369.507, corresponding to a likelihood ratio statistics of $LR = 42.552$. Here the LR statistic follows a equally weighted mixture of χ^2_0 and χ^2_1 distributions. Therefore, the resulting p-value 3.441×10^{-11} guarantees the appropriateness of the use of the multivariate-t distribution.

With missing-at-random assumption as in Vaida and Liu (2009), our dropout (censored) model does not bias the inference regarding the mean of β_j. For both models the mean viral load $E(y_{ij}) = \beta_j$ increases gradually throughout 24 months for the two models. For the best model (t-LMEC), the mean viral load increases from 3.62 at the time of UTI to 4.79 at 24 months. The estimates of the between-subject (α) and within-subject (σ^2) scale parameters (in log10 scale) are 0.6662 and 0.3503, respectively.

To determine possible influential observations, we use the Mahalanobis distance $d_i^2(\boldsymbol{\theta}) = (\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_i^{-1}(\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})$, $i = 1, \ldots, 72$. As in Pinheiro et al. (2001), replacing $\boldsymbol{\theta}$ and $\mathbf{b}_i$ with

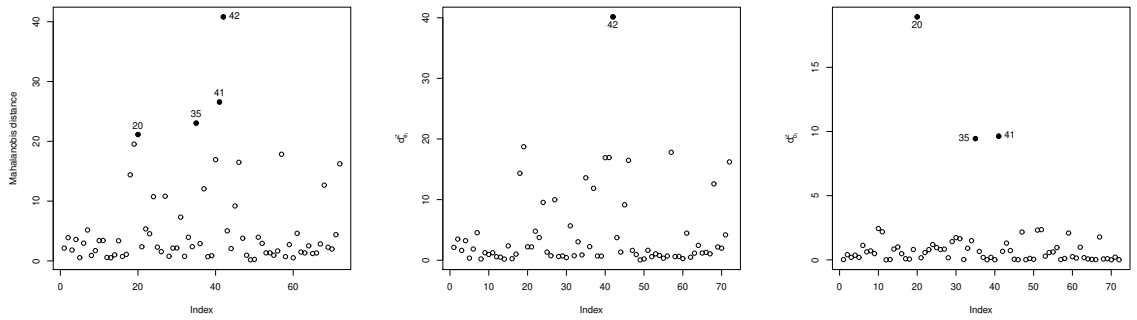


Figura 3.2: UTI data. (a) Mahalanobis distance, (b) Estimated $d^2_{\mathbf{e}_i}$ (error) and (c) Estimated $d^2_{\mathbf{b}_i}$ (R.E.), for the N-LMEC model.

their current estimates, we obtain the following decomposition for the Mahalanobis distance:

$$\begin{aligned} d_i^2(\widehat{\boldsymbol{\theta}}) &= (\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top(\widehat{\sigma^2}\mathbf{I}_{ni} + \mathbf{Z}_i^\top\widehat{\mathbf{D}}\mathbf{Z}_i)^{-1}(\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}) \\ &= -\frac{1}{\widehat{\sigma^2}}\widehat{\mathbf{e}}_i{}^\top\widehat{\mathbf{e}}_i + \widehat{\mathbf{b}}_i{}^\top\widehat{\mathbf{D}}\widehat{\mathbf{b}}_i, = \widehat{d^2_{\mathbf{e}_i}} + \widehat{d^2_{\mathbf{b}_i}} \end{aligned}$$

where $\widehat{\mathbf{e}}_i = \widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i\widehat{\mathbf{b}}_i$ where $\widehat{\mathbf{b}}_i$ is as in (3.20). The estimated distances $d^2_{\mathbf{e}_i}$ (Error) and $d^2_{\mathbf{b}_i}$ (Random Effect-R.E.) provide a useful diagnostic statistics for identifying subjects with outlying observations (see, for example, Meza and Osorio, 2011). Figure 3.2 presents these diagnostic statistics for N-LMEC model. Subjects #42 present large values of d_i^2 and $d^2_{\mathbf{e}_i}$, suggesting an outlying observation at the within-subject level (**e**-outlier). Moreover, observations #20, #35 and #41 presents large value of $d^2_{\mathbf{b}_i}$, suggesting outlying observations at the between-subject level (**b**-outlier). Under a Bayesian paradigm, these observations were also detected as influential in the work by Lachos et al. (2011).

It is well known that outlying observations may affect the estimation of the parameters under assumptions of normality. However, when we use the Student-t distribution, the EM algorithm allows to accommodate these discrepant observations attributing to them small weights in the estimation procedure. The estimated weights ($\widehat{u}_i$, $i = 1, \ldots, 72$) for the t-LMEC model are presented in Figure 3.3. We observe from Figure that observations #20, #35, #41 and #42, indicated as outliers under the normal model, take smaller weights, confirming the robust aspects of the MLE against outlying observations under the t-LMEC model. The robustness of the t-LMEC is also observed in Figure 3.1 (right panel), where the presence of these outliers might have underestimated the predicted mean curve for the N-LMEC model as compared to the t-LMEC model. In summary, we can see from this

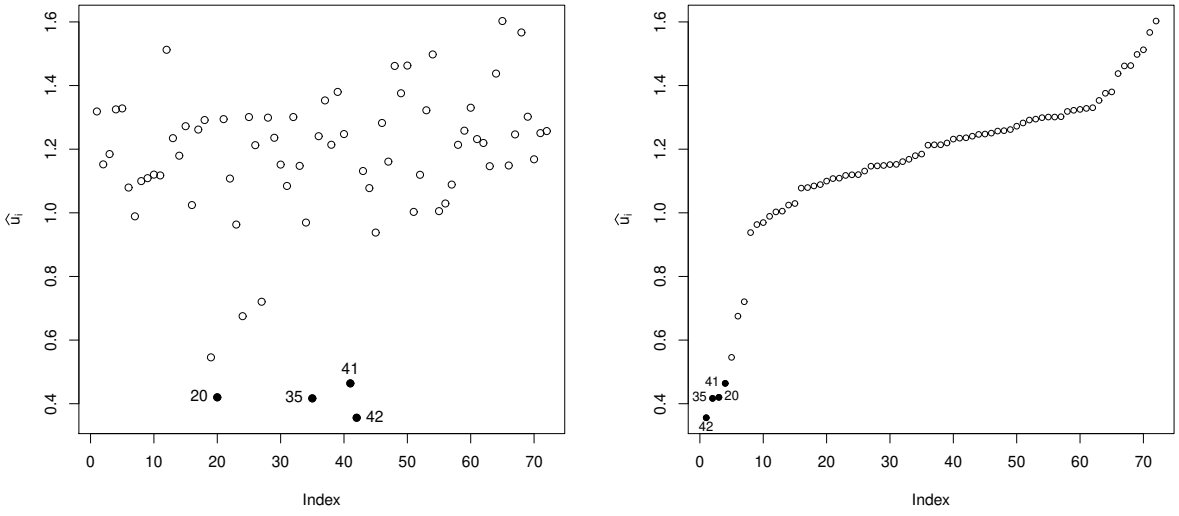


Figura 3.3: UTI data. Estimated weight $\widehat{u}_i$ for the t-LMEC fit. The influential observations for the N-LMEC are numbered.

example that the robust aspects of the t-LME models (Pinheiro et al., 2001) against outlying observations are also extended to the case in which censoring components are present.

3.6.2 AIEDRP study

The second AIDS case study is from the AIEDRP program, and is the same study in 2.6.2.

Within a classical framework, we use the Student-t (t-NLMEC) with the ECM algorithm as described in subsection 3.3.2. As in the previous application, the estimation of the parameters ν was chosen following the strategy proposed by Lange et al. (1989), which selects a small value for $\nu = 10$ (see left panel in Figure 3.4). This parameter act as tuning constant in robust estimation methods and in our case we see that this choice provide adequate protection against outliers. For the sake of model comparison, we also fit the N-NLMEC counterparts, which can be treated as the reduced t-NLMEC as ν tends to infinity.

Table 3.2 lists the ML estimates parameters for the N-LMEC model and the t-LMEC model, together with the corresponding standard errors of the fixed effects and the associated AIC and BIC values. From this table, we observe that the standard errors of the t-NLMEC are smaller, indicating that the Student-t model to produce more precise estimates. According to the AIC or BIC values, the t-NLMEC provided much improved model fits over the N-NLMEC.

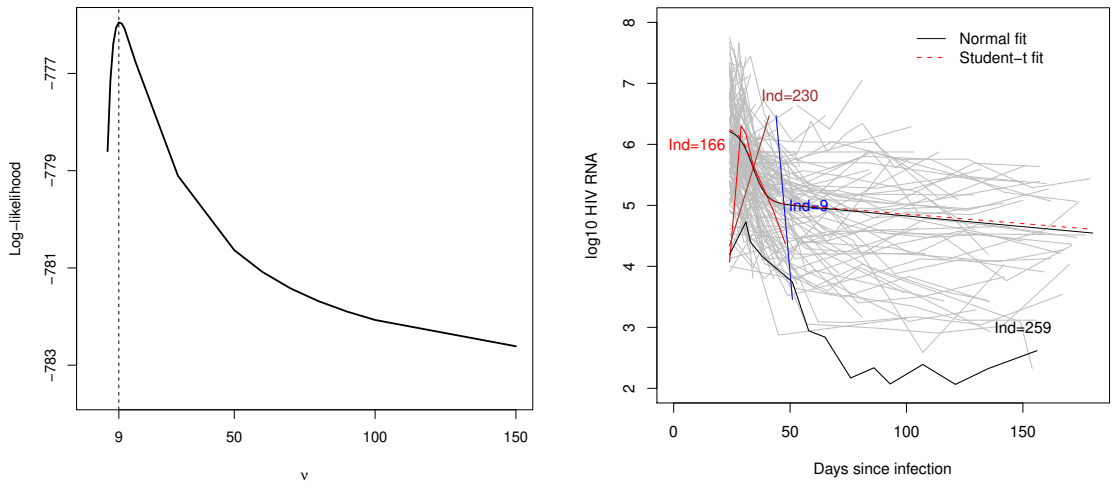


Figura 3.4: AIEDRP data. (Left panel) Plot of the profile log-likelihood of the degrees of freedom ν. (Right panel) Individual profiles and overall mean (in $\log_{10}$ scale) using the Normal and t distributions for HIV viral load at different follow-up times. The trajectories for the influential individuals are numbered.

Tabela 3.2: ML estimates under normal and Student-t models fitted to the AIEDRP data. SE are the corresponding standard errors.

	N-LMEC		T-LMEC	
Parameter	estimate	SE	estimate	SE
β_1	1.60964	0.0147	1.61148	0.0133
β_2	0.14217	0.0949	0.16122	0.0849
β_3	3.52617	0.0237	3.52370	0.0208
β_4	1.05585	0.2677	0.98713	0.2458
β_5	-0.0035	0.0014	-0.0031	0.0013
σ^2	0.26521		0.20726	
α_{11}	0.01769		0.01611	
α_{12}	0.00016		0.00013	
α_{22}	0.00004		0.00004	
ν	-	-	10	-
AIC	1610.814		1581.416	
BIC	1700.521		1623.908	

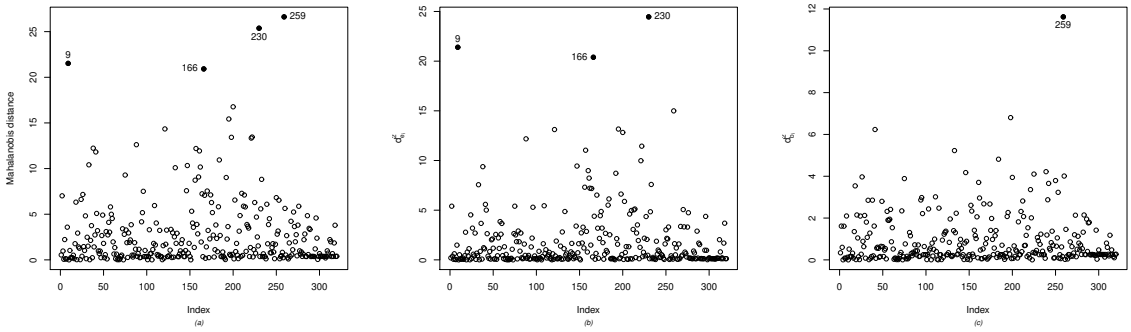


Figura 3.5: AIEDRP data. (a) Mahalanobis distance, (b) Estimated $d^2_{\mathbf{e}_i}$ (error) and (c) Estimated $d^2_{\mathbf{b}_i}$ (R.E.). The influential observations are numbered.

In fact, the maximum log-likelihood for the N-LMEC is -781.708 and for the t-LMEC model is -775.951, corresponding to the likelihood ratio statistics of 11.508 (p-value = 0.00035), this also reinforce the conclusion that the t-LMEC model fits the data significantly better than N-LMEC model.

To identify outlying observations, we compute the Mahalanobis distance $d_i^2(\widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, 320$, the estimated distances $d^2_{\mathbf{e}_i}$ (Error) and $d^2_{\mathbf{b}_i}$ (Random Effect), were also computed for the normal case. Figure 3.5 presents these diagnostic statistics for the N-LMEC model. We can see from this figures that observations #9, #166, #230 and #259 appear as possible outliers. The observations #9, #166 and #230 presents large value of $d^2_{\mathbf{e}_i}$, suggesting an **e**-outlier. Moreover, observation #259 presents large value of $d^2_{\mathbf{b}_i}$, suggesting an **b**-outlier. From figure 3.4 (right panel), the fitted viral load curve appears to be underestimated as compared to the t-NLMEC due to the presence of these outliers. This suggests that t-NLMEC, which down-weights the influence of outliers, provides an appropriate way for achieving robust inference.

The robustness of the t-LMEC model can be assessed by considering the influence of a single outlying observation on the ML estimate of θ. In particular, we can assess how much the ML estimates of θ influences by a change of δ units in a single observation y_{ik}. We replace a single observation y_{ik} by $y_{ik}(\delta) = y_{ik} + \delta$, and record the relative change in the estimates $((\widehat{\theta}(\delta) - \widehat{\theta})/\widehat{\theta})$, where $\widehat{\theta}$ denotes the original estimate and $\widehat{\theta}(\delta)$ the estimate for the contaminated data. In this application we contaminated the first observation on subject 198 and varied δ between -10 and 10. In Figure 3.6 we present the results of the relatives changes of the estimates $\boldsymbol{\beta}$ and σ^2 for different values of δ, under the N-NLMEC and t-NLMEC models. As expected, the estimates from the t-NLMEC is less affected by variations of δ than the

N-NLMEC.

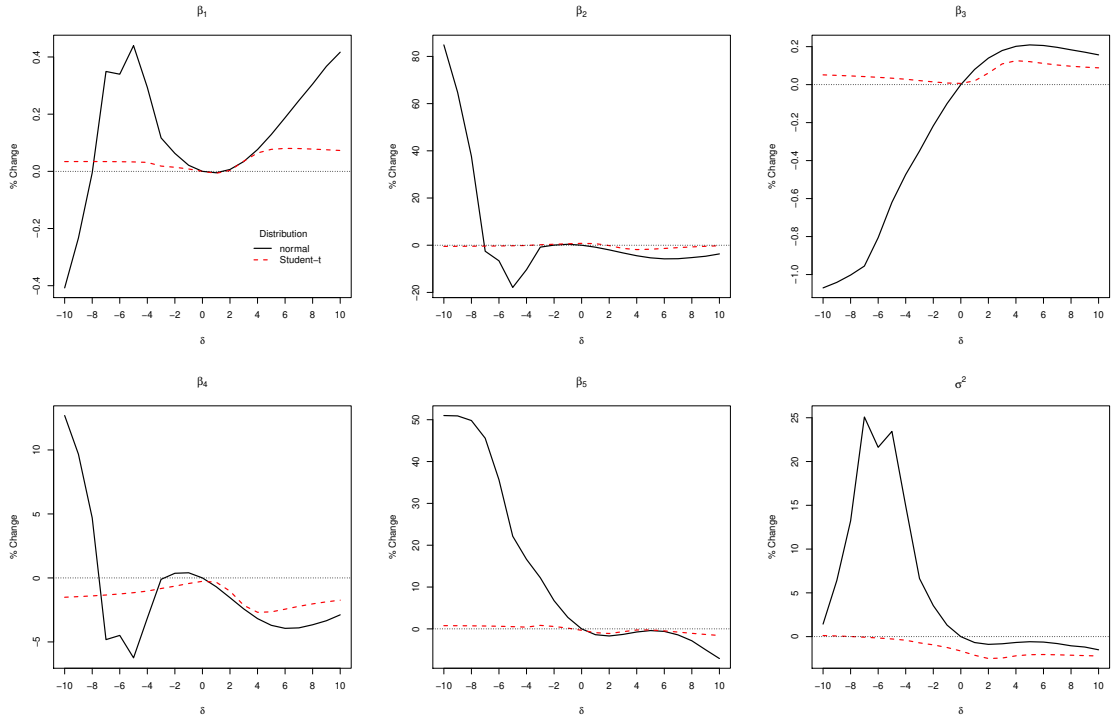


Figura 3.6: AIEDRP data. Relative changes on the ML estimates of $\boldsymbol{\theta}$ from the N-NLMEC (solid line) and the t-NLMEC (dashed line) for different contaminations δ.

3.6.3 Simulation studies

To study the performance of our proposed methodology we conduct a simulation study to illustrate the linear and nonlinear cases. The goal of this simulation study is to investigate the consequences on parameter inference when the normality assumption is inappropriate as well as to investigate whether the model comparison measures, AIC and BIC determines the best-fitting model to the simulated data.

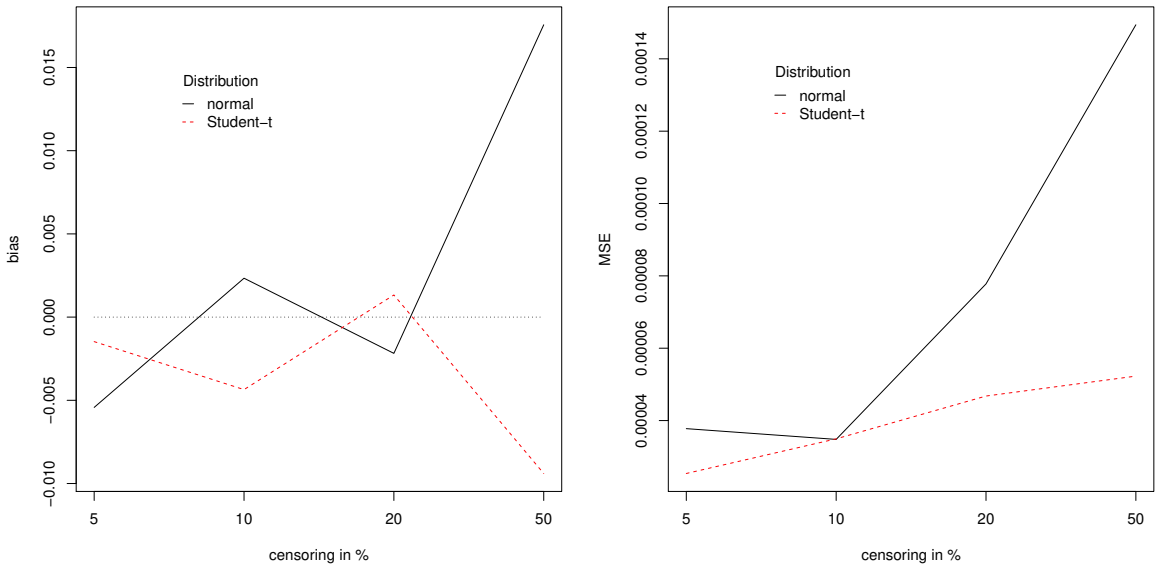
The linear case

To study the linear regression, we consider the following linear mixed model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, 100, \quad j = 1, \ldots, 6, \tag{3.25}$$

where $(b_{0i}, b_{1i}) \overset{\text{iid.}}{\sim} t_2(\mathbf{0}, \mathbf{D}, \nu)$, $\epsilon_{ij} \sim t(0, \sigma^2, \nu)$. We set $t_{ij} = (2, 4, 6, 8, 10, 24)$, $\boldsymbol{\beta}^\top = (\beta_0, \beta_1) = (-2.83, -0.18)$, $\mathbf{D} = \begin{bmatrix} 0.049 & 0.001 \\ 0.001 & 0.002 \end{bmatrix}$, $\sigma^2 = 0.15$ and $\nu = 4$.

We choose various settings of censoring proportions, 5%, 10%, 20% and 50%, to study the effect of the level of censoring in the estimation. This way, we have 4 different simulation settings with 100 simulated datasets for each setting. Once the simulated data is generated, we fit the LMEC model assuming normal and Student-t distributions. For each of the simulations, we fit the model given in (3.25) assuming normal and Student-t distributions. For each simulation, the parameters estimation as well as AIC and BIC were recorded. Table 3.3 presents the summary statistics for β (the fixed-effects parameters) assuming normal and Student-t distributions for the 4 censoring patterns. In the Table, MC Mean denotes the arithmetic average of the 100 estimates given by $\sum_{j=1}^{100} \hat{\gamma}_j/100$ and MC Sd is the arithmetic average of the 100 posterior standard deviations given by $\sum_{j=1}^{100} sd(\hat{\gamma}_j)/100$, where $\gamma = \beta_1, \beta_2$ or σ^2. In addition, we also estimate the MC coverage of β_1 and β_2, i.e. the proportion of times the 95% confidence interval includes the true value of the fixed effects.



(a) Bias of β_1 (b) MSE of β_1

Figura 3.7: Simulation studies. (a) Represents the bias of β_1 in comparison with the true value for the normal and Student-t models for the 4 censoring patterns (5%, 10%, 20%, 50%)in the LMEC setup. (b) Presents the Mean Square Error (MSE) for β_1 for the normal and Student-t models.

From Table 3.3, we observe that the Student-t distribution over perform the normal distribution at all levels of censoring. Figure 3.7 shows that for the normal distribution there

Tabela 3.3: Monte Carlo results based on 100 simulated Student-t samples. MC mean and MC Sd (in parenthesis) and MC Coverage are the respective mean estimates, standard deviations and coverage proportion average from fitting LMEC with Student-t and normal assumptions with different settings of censoring proportions. IM Sd are the average values of the approximate standard errors obtained through the information-based method. MC AIC and MC BIC are the arithmetic average of the respective model comparison measures.

			Simulated Student-t data				
Censoring	Fit		β_1	β_2	σ^2	MC AIC	MC BIC
5%	Normal	MC Mean	-2.839	-0.179	0.285	604.261	626.484
		IM Sd	0.068	0.010			
		MC Sd	0.065	(0.006)	(0.072)		
		MC Coverage	98%	99%			
	Student-t	MC Mean	-2.831	-0.180	0.154	554.302	576.525
		IM Sd	0.055	0.008			
		MC Sd	(0.052)	(0.005)	(0.023)		
		MC Coverage	95%	100%			
10%	Normal	MC Mean	-2.822	-0.180	0.281	569.744	591.966
		IM Sd	0.070	0.010			
		MC Sd	(0.061)	(0.006)	(0.078)		
		MC Coverage	99%	99%			
	Student-t	MC Mean	-2.830	-0.179	0.150	526.334	548.557
		IM Sd	0.057	0.008			
		MC Sd	(0.059)	(0.006)	(0.024)		
		MC Coverage	97%	100%			
20%	Normal	MC Mean	-2.824	-0.180	0.270	505.704	527.927
		IM Sd	0.079	0.013			
		MC Sd	(0.076)	(0.009)	(0.073)		
		MC Coverage	97%	99%			
	Student-t	MC Mean	-2.832	-0.180	0.151	474.053	496.276
		IM Sd	0.068	0.011			
		MC Sd	(0.063)	(0.007)	(0.031)		
		MC Coverage	100%	99%			
50%	Normal	MC Mean	-2.810	-0.183	0.285	407.693	429.916
		IM Sd	0.090	0.016			
		MC Sd	(0.088)	(0.012)	(0.072)		
		MC Coverage	98%	99%			
	Student-t	MC Mean	-2.840	-0.178	0.154	387.582	409.805
		IM Sd	0.081	0.015			
		MC Sd	(0.066)	(0.007)	(0.023)		
		MC Coverage	98%	100%			

is a strong increase of the bias (the deviations of the parameter estimates from the true value) as well as the mean square error (MSE). Clearly, the Student-t model shows much less bias and thus more precise estimations. Therefore, models with heavier tails than normal produce more accurate estimates in the context of censored data; the degree and direction of the bias in fixed effects depends both on the relative proportions of censoring as well as

model assumption. Observe that from Table 3.3 $\hat{\sigma}^2$ for the normal distribution is almost twice the true σ^2. This is due to the fact that in the normal scenario σ^2 represents the variance and therefore should be compared with $\frac{\nu}{\nu-2}\sigma^2$, which is 0.30. Notice also that, the Student-t model has a smaller confidence interval due to the smaller standard deviation but its coverage is slightly better than the normal method. This fact provides (once again) that the estimation of the Student-t method is more robust when dealing with censored data. Table 3.3 also provides the average values of the approximate standard deviations of the ML estimates obtained through the information-based method described in Subsection 3.3.3 (IM Sd) and the Monte Carlo standard deviation (Mc Sd) for the parameters. As we can see, the estimation method of the standard deviation provides relatively close results for the normal and Student-t methods, showing that the proposed asymptotic approximation for the variances of the fixed effects is reliable.

We also present the arithmetic average (MC AIC and MC BIC) of the model comparison criterions mentioned earlier. All the measures strongly favored the Student-t model, demonstrating the ability of these measures to detect an obvious departure from normality. The percentage of samples when these criteria chooses the t-LMEC also remains high.

The nonlinear case

As in the linear case we fix the censoring proportion as presented in Section 3.6.3 and also generated 100 simulated data sets. Following Vaida and Liu (2009), to study the nonlinear regression, we consider the following nonlinear mixed model:

$$y_{ij} = \alpha_{1i} + \frac{\alpha_2}{(1+\exp((t_{ij}-\alpha_3)/\alpha_{4i}))} + \epsilon_{ij}, \quad i = 1,\ldots,100, \quad j = 1,\ldots,10, \tag{3.26}$$

where $(b_{1i}, b_{2i}) \overset{\text{iid.}}{\sim} t_2(\mathbf{0}, \mathbf{D}, \nu)$ and $\epsilon_{ij} \sim t(0, \sigma^2, \nu)$. We reparametrize $\beta_{1i} = \log(\alpha_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\alpha_k), k = 2, 3$, $\alpha_{4i} = \beta_4 + b_{2i}$ and in addition, we set $\nu = 4$, $\sigma^2 = 0.55$, $\mathbf{D} = \begin{bmatrix} 0.0025 & -0.0010 \\ -0.0010 & 0.0100 \end{bmatrix}$, $\boldsymbol{\beta}^\top = (\beta_1, \beta_2, \beta_3, \beta_4) = (1.6094, 0.6931, 3.8067, 2.3026)$ and $t_{ij} = (1, 10, 20, 30, 40, 50, 60, 70, 80, 90)$.

We fit the NLMEC model (3.26) assuming normal and Student-t distributions. For each of the simulations, we fit the re-parameterized model given in (3.26) assuming normal and Student-t distributions. The model selection criterion AIC and BIC as well as the parameters estimation were recorded for each simulation. For the 4 censoring patterns, the summary statistics for β (the fixed-effects parameters) are presented in Table 3.4 assuming normal and Student-t distributions.

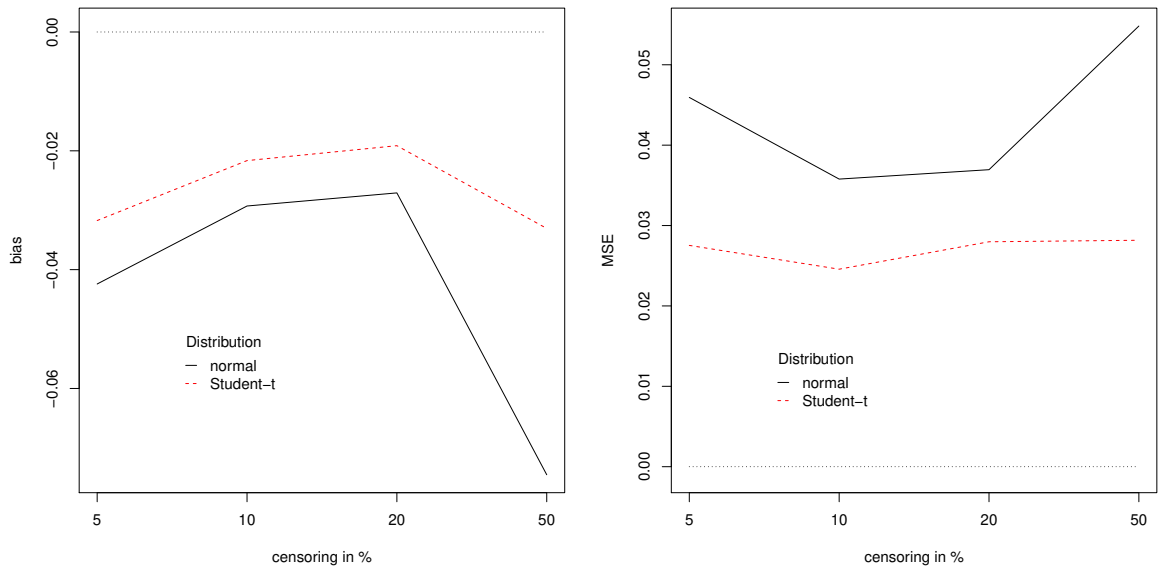
Tabela 3.4: Monte Carlo results based on 100 simulated Student-t samples. MC mean, MC Sd (in parenthesis) and MC Coverage are the respective mean estimates, standard deviations and coverage proportion average from fitting NLMEC with Student-t and normal assumptions with different settings of censoring proportions. IM Sd are the average values of the approximate standard error obtained through the information-based method. MC AIC and MC BIC are the arithmetic average of the respective model comparison measures.

			Simulated Student-t data						
Censoring	Fit		β_1	β_2	β_3	β_4	σ^2	MC AIC	MC BIC
5%	Normal	MC Mean	1.627	0.642	3.796	2.205	0.967	2865.279	2904.541
		IM Sd	0.017	0.068	0.041	0.191			
		MC Sd	(0.016)	(0.073)	(0.043)	(0.192)	(0.146)		
		MC coverage	81%	87%	96%	95%			
	Student-t	MC Mean	1.615	0.667	3.805	2.230	0.642	2654.928	2694.190
		IM Sd	0.015	0.058	0.035	0.161			
		MC Sd	(0.012)	(0.056)	(0.031)	(0.150)	(0.060)		
		MC coverage	96%	93%	99%	95%			
10%	Normal	MC Mean	1.623	0.657	3.801	2.235	0.970	2815.475	2854.737
		IM Sd	0.018	0.070	0.042	0.191			
		MC Sd	(0.017)	(0.069)	(0.046)	(0.178)	(0.141)		
		MC coverage	86%	88%	92%	95%			
	Student-t	MC Mean	1.613	0.676	3.803	2.253	0.629	2608.471	2647.733
		IM Sd	0.015	0.059	0.035	0.160			
		MC Sd	(0.014)	(0.057)	(0.036)	(0.150)	(0.057)		
		MC coverage	94%	94%	95%	97%			
20%	Normal	MC Mean	1.616	0.683	3.806	2.240	0.975	2705.762	2494.963
		IM Sd	0.019	0.070	0.042	0.190			
		MC Sd	(0.016)	(0.069)	(0.042)	(0.183)	(0.145)		
		MC coverage	95%	95%	98%	96%			
	Student-t	MC Mean	1.616	0.678	3.797	2.259	0.579	2494.963	2534.225
		IM Sd	0.015	0.059	0.035	0.157			
		MC Sd	(0.015)	(0.060)	(0.032)	(0.162)	(0.044)		
		MC coverage	89%	92%	99%	95%			
50%	Normal	MC Mean	1.614	0.684	3.781	2.131	0.978	1982.382	2021.644
		IM Sd	0.022	0.073	0.043	0.208			
		MC Sd	(0.023)	(0.069)	(0.045)	(0.160)	(0.186)		
		MC coverage	94%	95%	90%	93%			
	Student-t	MC Mean	1.624	0.650	3.789	2.226	0.546	1879.266	1918.528
		IM Sd	0.022	0.075	0.041	0.187			
		MC Sd	(0.016)	(0.066)	(0.040)	(0.151)	(0.038)		
		MC coverage	90%	93%	95%	95%			

From Table 3.4, we observe that for all levels of censoring the Student-t distribution performs better than the normal distribution and have a small standard deviation in the estimates providing more accurate estimation. The arithmetic average (MC AIC and MC BIC) of the model comparison criteria are also presented and strongly favors the Student-t model in comparison to the normal model. This, reinforce that these measures are capable

of detecting departures from normality. Like in the linear case, we have that the estimates $\hat{\sigma}^2$ of σ^2 for the normal distribution must be compared with $\frac{\nu}{\nu-2}\sigma^2$, which now is 1.10. As in the linear setup we can see that the Student-t model continues to have smaller confidence interval with a usually bigger coverage of the parameters. This is a strong evidence of the robustness in estimation of the Student-t method. Again, as observed in the linear case the IM Sd and MC Sd for the nonlinear regression provides close results for both models (normal and Student-t). This emphasize that the estimation of the standard error provided by the proposed asymptotic approximation of the fixed effects (Equation 3.21) is reliable.

In Figure 3.8 we represent the bias and MSE for the parameter estimates of β_4 for the normal and Student-t distributions. It is clear that the normal model has a much bigger bias and MSE than the Student-t model. Therefore, for censored data the Student-t model is more robust, providing more accurate estimations when the data has departures from the normality assumption. Although Figure 3.8 only presents the results for the estimates of β_4 a similar pattern was observed for all the other parameters.



(a) Bias of β_1 (b) MSE of β_1

Figura 3.8: Simulation studies. (a) Represents the bias of β_1 in comparison with the true value for the normal and Student-t models for the 4 censoring patterns (5%, 10%, 20%, 50%) in the NLMEC setup. (b) Presents the Mean Square Error (MSE) for β_1 for the normal and Student-t models.

3.7 Conclusions

We have proposed in this chapter a robust approach to linear and nonlinear mixed effects models with censored observation based on the multivariate-t distribution, called the t-LMEC/t-NLMEC. It offers a great deal of flexibility in dealing with longitudinal data in the presence of outliers. A novel ECM algorithm to obtain approximated MLEs is developed by exploring the statistical properties of the multivariate truncated Student-t distribution. Our proposed algorithm has a closed-form expression for the E-step, based on formulas for the mean and variance of the truncated Student-t distribution. Thus, the proposed methodologies allow the practitioner to fit longitudinal data in a broad variety of considerations. For NLMEC, the analysis is computationally feasible through approximating the t-NLMEC for a multivariate t distribution with specified parameters. We apply our methodology to two recent AIDS studies as well as simulated data to illustrate how the procedures can be used to evaluate model assumptions, identify outliers, and obtain robust parameter estimates. From these results it is encouraging that the use of t-LMEC/t-NLMEC models offer better fitting, protection against outliers and more precise inferences than the usual normal counterpart.

We conjecture that the methodology presented in this chapter should yield satisfactory results in other areas where multivariate data appears frequently, for instance, survival models, dynamics linear models, spatially censored data, etc., at expense of moderate complexity of implementation. Finally, the proposed EM algorithm has been coded and implemented in the R package *t-lmec*(R Development Core Team, 2009).

Capítulo 4

Considerações finais

Neste trabalho discutimos vários aspectos envolvendo modelos lineares e não lineares com efeito misto para resposta censuradas. Desenvolvemos métodos de diagnósticos clássicos para modelos lineares e não lineares com efeito misto para resposta censuradas, usando a distribuição normal multivariada.

Propusemos uma abordagem robusta para modelos lineares e não lineares com efeito misto de resposta censuradas com base na distribuição t-multivariada , denominado t-LMEC/t-NLMEC, oferecendo uma grande flexibilidade em lidar com dados longitudinais na presença de outliers.

Os resultados foram aplicados a dois conjuntos de dados de HIV considerados por Vaida and Liu (2009). Os estudos de simulação foram realizados utilizando programas estatísticos tais como R.

4.1 Trabalhos futuros

Vários trabalhos de pesquisa poderão ser obtidos a partir dos resultados desta dissertação, entre eles podemos sugerir os seguintes:

- Realizar um estudo de diagnóstico em modelos lineares e não lineares de efeito misto para resposta censuradas, com base na distribuição t-multivariada.
- Realizar um estudo de inferência e de diagnóstico em modelos lineares e não lineares de efeito misto para respostas censuradas, com base na família de distribuições normal e t assimétrica (Lachos et al., 2010).

- Realizar um estudo de inferência e diagnósticos em modelos com erros nas variáveis para respostas censuradas.

Referências

Arellano-Valle, R. and M. Genton (2010). Multivariate extended skew-t distributions and related families. *Metron LXVIII*, 201–234.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B, 48*, 133–169.

Cook, R. D. and S. Weisberg (1982). *Residuals and Influence in Regression*. Boca Raton, FL: Chapman & Hall/CRC.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

Genz, A., F. Bretz, T. Hothorn, T. Miwa, X. Mi, F. Leisch, and F. Scheipl (2008). mvtnorm: Multivariate Normal and t Distribution. *R package version 0.9-2, URL http://CRAN. R-project. org/package= mvtnorm*.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics 55*(2), 625–629.

Jacqmin-Gadda, H., R. Thiebaut, G. Chene, and D. Commenges (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics 1*(4), 355–368.

Lachos, V., D. Bandyopadhyay, and D. Dey (2011). Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions. *Biometrics*.

Lachos, V. H., P. Ghosh, and R. B. Arellano-Valle (2010). Likelihood based inference for skew–normal independent linear mixed models. *Statistica Sinica 20*, 303–322.

Laird, N. M. and J. H.Ware (1982). Random effects models for longitudinal data. *Biometrics 38*, 963–974.

Lange, K. L., R. Little, and J. Taylor (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association 84*, 881–896.

Lee, S. Y. and L. Xu (2004). Rinfluence analysis of nonlinear mixed-effects models. *Computational Statistics and Data Analysis 45*, 321–341.

Lesaffre, E. and G. Verbeke (1998). Local influence in linear mixed models. *Biometrics 54*, 570–582.

Lin, T., H. Ho, H. Chen, and W. Wang (2011). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference*.

Lin, T. and J. Lee (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference 137*(2), 484–495.

Lindstrom, M. and D. Bates (1990). Nonlinear mixed-effects models for repeated-measures data. *Biometrics 46*, 673687.

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association 91*, 1219–1227.

Louis, T. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 226–233.

Lucas, A. (1997). Robustness of the student t based m estimator. *Communications in Statistics: Theory and Methods 26*, 1165–118.

McLachlan, G. and T. Krishnan (1996). *The EM Algorithm and Extensions*. Wiley.

Meng, X. and D. B. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 81*, 633–648.

Meza, M. and F. Osorio (2011). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*.

Osorio, F., G. A. Paula, and M. Galea (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis 51*, 4354–4368.

Pinheiro, J. C. and M. Bates, Douglas (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer.

Pinheiro, J. C., C. H. Liu, and Y. N. Wu (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics 10*, 249–276.

Poon, W. and Y. Poon (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*(1), 51–61.

R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rosa, G. J. M., C. R. Padovani, and D. Gianola (2003). Robust linear mixed models with normal/independent distributions and bayesian mcmc implementation. *Biometrical Journal 45*, 573–590.

Russo, C., G. Paula, and R. Aoki (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics & Data Analysis 53*(12), 4143–4156.

Saitoh, A., M. Foca, R. Viani, S. Heffernan-Vacca, F. Vaida, J. Lujan-Zilbermann, P. Emmanuel, J. Deville, and S. Spector (2008). Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. *Pediatrics 121*(3), e513–e521.

Self, S. and K. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio test under nonstandard conditions. *Journal of the American Statistical Association*, 605–610.

Sinha, D., M. Chen, and S. Ghosh (1999). Bayesian Analysis and Model Selection for Interval-Censored Survival Data. *Biometrics 55*(2), 585–590.

Song, P., P. Zhang, and A. Qu (2007). Maximum likelihood inference in robust linear mixed-effects models using multivariate t distributions. *Statistica Sinica 17*, 929–943.

Vaida, F., A. Fitzgerald, and V. DeGruttola (2007). Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. *Computational statistics & data analysis 51*(12), 5718–5730.

Vaida, F. and L. Liu (2009). Fast Implementation for Normal Mixed Effects Models With Censored Response. *Journal of Computational and Graphical Statistics 18*(4), 797–817.

Wang, W. and T. Fan (2011). Estimation in multivariate t linear mixed models for multivariate longitudinal data. *Statistica Sinica 21*, 1857–1880.

Wu, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research 14*(2), 171–192.

Wu, L. (2010). *Mixed Effects Models for Complex Data*. Boca Raton, FL: Chapman & Hall/CRC.

Xie, F., B. Wei, and J. Lin (2007). Case-deletion influence measures for the data from multivariate t distributions. *Journal of Applied Statistics 34*(8), 907–921.

Zhu, H. and S. Lee (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B 63*, 111–126.

Apêndice A

Additional results of Chapter 2 and Chapter 3

A.1 The EM algorithm

A.1.1 Normal distribution

We include here the derivation of the EM equations (2.9) - (2.11).

Recall that the vector of parameters to be estimated is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha})$ and that $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$, such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the ith subject. In their estimation procedure, $\mathbf{b}$, $\mathbf{V}$ and $\mathbf{C}$ are treated as hypothetical missing data, and augmented with the observed data set $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top$,

$$L(\mathbf{y}_c|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{y}_i, \mathbf{b}_i) = \prod_{i=1}^{n} f(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \mathbf{b}_i) f(\mathbf{b}_i).$$

The complete log-likelihood is given by

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \log(L(\mathbf{y}_c|\boldsymbol{\theta})) = C - \frac{1}{2}\sum_{i=1}^{n}\Big[n_i \log\sigma^2 + \log|\mathbf{D}| + \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i \\ &\quad + \frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\Big], \end{aligned}$$

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. The EM function is given by

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \boldsymbol{\theta}^*].$$

So we have that,

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) &= C^* - \frac{1}{2}\sum_{i=1}^{n}\left\{ n_i \log\sigma^2 + \log|\mathbf{D}| + \operatorname{tr}\left(E[\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}^*]\mathbf{D}^{-1}\right)\right. \\ &\quad \left. + E\left[\frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)|\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}^*\right]\right\}, \end{aligned}$$

where C^* is a constant that is independent of the parameter vector $\boldsymbol{\theta}$.

In order to introduce some important results, we establish the following lemma,

Lemma 1 *Let* $\mathbf{Y} \overset{\text{ind.}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $\mathbf{X} \overset{\text{ind.}}{\sim} N_q(\boldsymbol{\eta}, \boldsymbol{\Omega})$. *So,*

$$\begin{aligned} \phi_p(\mathbf{y}|\boldsymbol{\mu} + \mathbf{A}x, \boldsymbol{\Sigma})\phi_q(x, \boldsymbol{\Omega}) &= \phi_p(\mathbf{y}|\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\eta}, \boldsymbol{\Sigma} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top) \\ &\quad \times\phi_q(x|\boldsymbol{\eta} + \boldsymbol{\Lambda}\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{\eta}), \boldsymbol{\Sigma}), \end{aligned}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\Omega}^{-1} + \mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}$.

Then to compute the expectation term above, note first that,

$$\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i \overset{\text{ind.}}{\sim} TN_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i),$$

and using the Lemma 1,

$$\mathbf{b}_i|\mathbf{y}_i \overset{\text{ind.}}{\sim} N_q\left(\left(\mathbf{D}^{-1} + \frac{1}{\sigma^2}\mathbf{Z}_i^\top\mathbf{Z}_i\right)^{-1}\frac{1}{\sigma^2}\mathbf{Z}_i^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \left(\mathbf{D}^{-1} + \frac{1}{\sigma^2}\mathbf{Z}_i^\top\mathbf{Z}_i\right)^{-1}\right),$$

$$\mathbf{b}_i|\mathbf{y}_i \overset{\text{ind.}}{\sim} N_q\left(\boldsymbol{\varphi}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \boldsymbol{\Lambda}_i\right),$$

with $\boldsymbol{\Lambda}_i = \left(\mathbf{D}^{-1} + \frac{1}{\sigma^2}\mathbf{Z}_i^\top\mathbf{Z}_i\right)^{-1}$ and $\boldsymbol{\varphi}_i = \frac{1}{\sigma^2}\boldsymbol{\Lambda}_i\mathbf{Z}_i^\top$.

Now using the proposition (1) we compute this expectation term:

$$\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}[E_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{y}_i)] = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[\mathbf{y}_i\right],$$

$$\widehat{\mathbf{y}_i^2} = E\{\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}[E_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{y}_i\mathbf{y}_i^\top)] = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[\mathbf{y}_i\mathbf{y}_i^\top\right],$$

$$\begin{aligned}\widehat{\mathbf{b}_i} &= E\{\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}[E_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_i)] = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[\boldsymbol{\varphi}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\right]\\ &= \boldsymbol{\varphi}_i\left[E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}(\mathbf{y}_i)-\mathbf{X}_i\boldsymbol{\beta}\right] = \boldsymbol{\varphi}_i\left(\widehat{\mathbf{y}}_i-\mathbf{X}_i\boldsymbol{\beta}\right),\end{aligned}$$

$$\begin{aligned}\widehat{\mathbf{b}_i^2} &= E\{\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}[E_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_i\mathbf{b}_i^\top)]\\ &= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[\left(\boldsymbol{\Lambda}_i+\boldsymbol{\varphi}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\varphi}_i^\top\right)\right]\\ &= \boldsymbol{\Lambda}_i+\boldsymbol{\varphi}_i E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\right]\boldsymbol{\varphi}_i^\top\\ &= \boldsymbol{\Lambda}_i+\boldsymbol{\varphi}_i\left\{E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\mathbf{y}_i\mathbf{y}_i^\top\right)-E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}(\mathbf{y}_i)\,\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right.\\ &\quad\left.-\mathbf{X}_i\boldsymbol{\beta}\left[E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}(\mathbf{y}_i)\right]^\top+\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right\}\boldsymbol{\varphi}_i^\top\\ &= \boldsymbol{\Lambda}_i+\boldsymbol{\varphi}_i\left[\widehat{\mathbf{y}_i^2}-\widehat{\mathbf{y}}_i\boldsymbol{\beta}^\top\mathbf{X}_i^\top-\mathbf{X}_i\boldsymbol{\beta}\widehat{\mathbf{y}}_i^\top+\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right]\boldsymbol{\varphi}_i^\top,\end{aligned}$$

$$\begin{aligned}\widehat{\mathbf{y}\mathbf{b}_i} &= E\{\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}[E_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{y}_i\mathbf{b}_i^\top)] = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}[\mathbf{y}_i E_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_i^\top)]\\ &= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[\mathbf{y}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\varphi}_i^\top\right] = \left[E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}(\mathbf{y}_i\mathbf{y}_i^\top)-E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}(\mathbf{y}_i)\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right]\boldsymbol{\varphi}_i^\top\\ &= \left(\widehat{\mathbf{y}_i^2}-\widehat{\mathbf{y}}_i\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right)\boldsymbol{\varphi}_i^\top.\end{aligned}$$

Replacing the expectation in $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = C^* - \frac{1}{2}\sum_{i=1}^{n}\left[n_i\log\sigma^2+\log|\mathbf{D}|+\operatorname{tr}\left(\widehat{\mathbf{b}_i^2}\mathbf{D}^{-1}\right)+\frac{A_i}{\sigma^2}\right],$$

where

$$\begin{aligned}A_i &= \operatorname{tr}(\widehat{\mathbf{y}_i^2})-\widehat{\mathbf{y}}_i^\top\mathbf{X}_i\boldsymbol{\beta}-\operatorname{tr}(\widehat{\mathbf{y}\mathbf{b}_i}^\top\mathbf{Z}_i)-\boldsymbol{\beta}^\top\mathbf{X}_i^\top\widehat{\mathbf{y}}_i+\boldsymbol{\beta}^\top\mathbf{X}_i^\top\mathbf{X}_i\boldsymbol{\beta}\\ &\quad+\boldsymbol{\beta}^\top\mathbf{X}_i^\top\mathbf{Z}_i\widehat{\mathbf{b}}_i-\operatorname{tr}(\widehat{\mathbf{y}\mathbf{b}_i}\mathbf{Z}_i^\top)+\widehat{\mathbf{b}}_i^{\,\top}\mathbf{Z}_i^\top\mathbf{X}_i\boldsymbol{\beta}+\operatorname{tr}(\widehat{\mathbf{b}_i^2}\mathbf{Z}_i^\top\mathbf{Z}_i).\end{aligned}$$

The differential with respect to $\boldsymbol{\beta}$, σ^2 and $\mathbf{D}$ are

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial\boldsymbol{\beta}} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}-\mathbf{X}_i^\top(\widehat{\mathbf{y}}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\widehat{\mathbf{b}}_i),$$

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \sigma^2} = -\frac{1}{2}\sum_{i=1}^{n}\left[\frac{n_i}{\sigma^2} - \frac{A_i}{(\sigma^2)^2}\right],$$

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \mathbf{D}^{-1}} = -\frac{n}{\sigma^2}(-2\mathbf{D} + diag(\mathbf{D})) - \frac{1}{2}\sum_{i=1}^{n}\left(\widehat{\mathbf{y}}_i + \widehat{\mathbf{y}}_i{}^{\top} - diag(\widehat{\mathbf{b}_i^2})\right).$$

The solution of $\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \boldsymbol{\beta}} = 0$ is

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\mathbf{X}_i^{\top}\mathbf{X}_i\right)^{-1}\left[\sum_{i=1}^{n}\mathbf{X}_i(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i)\right].$$

The solution of $\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \sigma^2} = 0$ is

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} n_i}.$$

For unstructured $\mathbf{D}$, the solution of $\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \mathbf{D}^{-1}} = 0$, for all $\mathbf{D}$, is

$$\widehat{\mathbf{D}} = \frac{\sum_{i=1}^{n}\widehat{\mathbf{b}_i^2}}{n}.$$

A.1.2 Student-t distribution

We include here the derivation of the EM equations (3.17) - (3.19).

Recall that the vector of parameters to be estimated is $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \sigma^2, \boldsymbol{\alpha})$ and that $\mathbf{y} = (\mathbf{y}_1^{\top}, \ldots, \mathbf{y}_n^{\top})^{\top}$, $\mathbf{b} = (\mathbf{b}_1^{\top}, \ldots, \mathbf{b}_n^{\top})^{\top}$, $\mathbf{u} = (u_1, \ldots, u_n)^{\top}$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$, such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the ith subject. In their estimation procedure, $\mathbf{b}$, $\mathbf{V}$ and $\mathbf{C}$ are treated as hypothetical missing data, and augmented with the observed data set $\mathbf{y}_c = (\mathbf{C}^{\top}, \mathbf{V}^{\top}, \mathbf{y}^{\top}, \mathbf{b}^{\top}, \mathbf{u}^{\top})^{\top}$,

$$L(\mathbf{y}_c|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{y}_i, \mathbf{b}_i, u_i) = \prod_{i=1}^{n} f(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \mathbf{b}_i, u_i) f(\mathbf{b}_i|u_i) f(u_i).$$

The complete log-likelihood is given by

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \log(L(\mathbf{y}_c|\boldsymbol{\theta})) = C + \sum_{i=1}^{n}\Big\{h(u_i|\nu) - \frac{1}{2}\Big[n_i\log\sigma^2 + \log|\mathbf{D}| + u_i\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i \\ &\quad + \frac{u_i}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\Big]\Big\},\end{aligned}$$

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$ and $h(u_i|\nu)$ is a density of a $Gamma(\nu/2, \nu/2)$. The EM function is given by

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \boldsymbol{\theta}^*].$$

So we have that,

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) &= C^* - \frac{1}{2}\sum_{i=1}^{n}\Big\{n_i\log\sigma^2 + \log|\mathbf{D}| + \mathrm{tr}\left(E[u_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}^*]\mathbf{D}^{-1}\right) \\ &\quad + E\left[\frac{u_i}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)|\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}^*\right]\Big\},\end{aligned}$$

where C^* is a constant that is independent of the parameter vector $\boldsymbol{\theta}$.

Then to compute the expectation term above, note first that,

$$\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i \overset{\text{ind.}}{\sim} Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu),$$

$$E(u_i|\mathbf{y}_i) = \frac{\nu + n_i}{\nu + \delta},$$

where $\delta = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, and using the Lemma 1

$$\mathbf{b}_i|\mathbf{y}_i, u_i \overset{\text{ind.}}{\sim} N_q\left(\frac{u_i}{\sigma^2}\left(u_i\mathbf{D}^{-1} + \frac{u_i}{\sigma^2}\mathbf{Z}_i^\top\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \left(u_i\mathbf{D}^{-1} + \frac{u_i}{\sigma^2}\mathbf{Z}_i^\top\mathbf{Z}_i\right)^{-1}\right),$$

$$\mathbf{b}_i|\mathbf{y}_i, u_i \overset{\text{ind.}}{\sim} N_q\left(\boldsymbol{\varphi}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \frac{\sigma^2}{u_i}\boldsymbol{\Lambda}_i\right),$$

with $\boldsymbol{\Lambda}_i = (\sigma^2\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathbf{Z}_i)^{-1}$ and $\boldsymbol{\varphi}_i = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top$. Using the propositions (2)-(4) we compute

this expectation term:

$$\begin{aligned}
\widehat{u\mathbf{y}_i} &= E\{u_i\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[E_{\mathbf{b}_i|\mathbf{y}_i,u_i}(u\mathbf{y}_i)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[E_{u_i|\mathbf{y}_i}(u_i\mathbf{y}_i)\right] = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\right) \\
&= \frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\},
\end{aligned}$$

$$\begin{aligned}
\widehat{u\mathbf{y}_i^2} &= E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[E_{\mathbf{b}_i|\mathbf{y}_i,u_i}(u\mathbf{y}_i\mathbf{y}_i^\top)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[E_{u_i|\mathbf{y}_i}(u_i\mathbf{y}_i\mathbf{y}_i^\top)\right] = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\mathbf{y}_i^\top\right) \\
&= \frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\mathbf{W}_i^\top\},
\end{aligned}$$

$$\begin{aligned}
\widehat{u_i} &= E\{u_i|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[E_{\mathbf{b}_i|\mathbf{y}_i,u_i}(u_i)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[E_{u_i|\mathbf{y}_i}(u_i)\right] = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\right) \\
&= \frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i^0\} = \frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)},
\end{aligned}$$

$$\begin{aligned}
\widehat{u\mathbf{b}_i} &= E\{u_i\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[E_{\mathbf{b}_i|\mathbf{y}_i,u_i}(u_i\mathbf{b}_i)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[u_iE_{\mathbf{b}_i|\mathbf{y}_i,u_i}(\mathbf{b}_i)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}\left[u_i\boldsymbol{\varphi}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\right]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[\boldsymbol{\varphi}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})E_{u_i|\mathbf{y}_i}(u_i)\right] \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left\{\boldsymbol{\varphi}_i\left[\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\right)-\mathbf{X}_i\boldsymbol{\beta}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\right)\right]\right\} \\
&= \boldsymbol{\varphi}_i\left[E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\right)-\mathbf{X}_i\boldsymbol{\beta}E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\right)\right] \\
&= \boldsymbol{\varphi}_i\left[\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\}-\mathbf{X}_i\boldsymbol{\beta}\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}\right] \\
&= \boldsymbol{\varphi}_i\left(\widehat{u\mathbf{y}_i}-\mathbf{X}_i\boldsymbol{\beta}\widehat{u_i}\right),
\end{aligned}$$

$$
\begin{aligned}
\widehat{u\mathbf{b}_i^2} &= E\{u_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[E_{\mathbf{b}_i|\mathbf{y}_i,u_i}(u_i\mathbf{b}_i\mathbf{b}_i^\top)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[u_iE_{\mathbf{b}_i|\mathbf{y}_i,u_i}(\mathbf{b}_i\mathbf{b}_i^\top)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left\{E_{u_i|\mathbf{y}_i}\left[u_i\left(\boldsymbol{\Lambda}_i(u_i^{-1}\sigma^2+\mathbf{Z}_i^\top(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top)\boldsymbol{\varphi}_i^\top\right)\right]\right\} \\
&= \boldsymbol{\Lambda}_i\sigma^2+E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[E_{u_i|\mathbf{y}_i}(u_i)\left(\boldsymbol{\varphi}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\varphi}_i^\top\right)\right] \\
&= \boldsymbol{\Lambda}_i\sigma^2+\boldsymbol{\varphi}_i\left\{E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\mathbf{y}_i^\top\right)-E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\right)\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right. \\
&\quad \left.-\mathbf{X}_i\boldsymbol{\beta}\left[E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\right)\right]^\top+E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\right)\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right\}\boldsymbol{\varphi}_i^\top \\
&= \boldsymbol{\Lambda}_i\sigma^2+\boldsymbol{\varphi}_i\left\{\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\mathbf{W}_i^\top\}-\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right. \\
&\quad \left.-\mathbf{X}_i\boldsymbol{\beta}\left[\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\}\right]^\top+\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right\}\boldsymbol{\varphi}_i^\top \\
&= \boldsymbol{\Lambda}_i\sigma^2+\boldsymbol{\varphi}_i\left(\widehat{u\mathbf{y}_i^2}-\widehat{u\mathbf{y}_i}\boldsymbol{\beta}^\top\mathbf{X}_i^\top-\mathbf{X}_i\boldsymbol{\beta}\widehat{u\mathbf{y}_i}^\top+\widehat{u_i}\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right)\boldsymbol{\varphi}_i^\top,
\end{aligned}
$$

$$
\begin{aligned}
\widehat{u\mathbf{y}\mathbf{b}_i} &= E\{u_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}^*\} = E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{E_{u_i|\mathbf{y}_i}[E_{\mathbf{b}_i|\mathbf{y}_i,u_i}(u_i\mathbf{y}_i\mathbf{b}_i^\top)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\{\mathbf{y}_iE_{u_i|\mathbf{y}_i}[u_iE_{\mathbf{b}_i|\mathbf{y}_i,u_i}(\mathbf{b}_i^\top)]\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left\{\mathbf{y}_iE_{u_i|\mathbf{y}_i}\left[u_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\varphi}_i^\top\right]\right\} \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left[\mathbf{y}_iE_{u_i|\mathbf{y}_i}(u_i)(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\varphi}_i^\top\right] \\
&= E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left\{\left[\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\mathbf{y}_i^\top\right)-\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\right)\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right]\boldsymbol{\varphi}_i^\top\right\} \\
&= \left[E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\mathbf{y}_i^\top\right)-E_{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i}\left(\frac{(\nu+n_i)}{(\nu+\delta)}\mathbf{y}_i\right)\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right]\boldsymbol{\varphi}_i^\top \\
&= \left[\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\}-\frac{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2)}{T_{n_i}(a|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}E\{\mathbf{W}_i\mathbf{W}_i^\top\}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right]\boldsymbol{\varphi}_i^\top \\
&= \left(\widehat{u\mathbf{y}_i^2}-\widehat{u\mathbf{y}_i}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\right)\boldsymbol{\varphi}_i^\top.
\end{aligned}
$$

Replacing the expectation in $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$

$$
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = C^*-\frac{1}{2}\sum_{i=1}^{n}\left[n_i\log\sigma^2+\log|\mathbf{D}|+\operatorname{tr}\left(\widehat{u\mathbf{b}_i^2}\mathbf{D}^{-1}\right)+\frac{A_i}{\sigma^2}\right],
$$

where

$$
\begin{aligned}
A_i &= \operatorname{tr}(\widehat{u\mathbf{y}_i^2})-\widehat{u\mathbf{y}}_i^\top\mathbf{X}_i\boldsymbol{\beta}-\operatorname{tr}(\widehat{u\mathbf{y}\mathbf{b}_i}^\top\mathbf{Z}_i)-\boldsymbol{\beta}^\top\mathbf{X}_i^\top\widehat{u\mathbf{y}_i}+\boldsymbol{\beta}^\top\mathbf{X}_i^\top\widehat{u}_i\mathbf{X}_i\boldsymbol{\beta} \\
&\quad +\boldsymbol{\beta}^\top\mathbf{X}_i^\top\mathbf{Z}_i\widehat{u\mathbf{b}_i}-\operatorname{tr}(\widehat{u\mathbf{y}\mathbf{b}_i}\mathbf{Z}_i^\top)+\widehat{u\mathbf{b}_i}^\top\mathbf{Z}_i^\top\mathbf{X}_i\boldsymbol{\beta}+\operatorname{tr}(\widehat{u\mathbf{b}_i^2}\mathbf{Z}_i^\top\mathbf{Z}_i).
\end{aligned}
$$

The differential with respect to $\boldsymbol{\beta}$, σ^2 and $\mathbf{D}$ are

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2}\sum_{i=1}^{n} -\mathbf{X}_i^\top(\widehat{u\mathbf{y}_i} - \widehat{u}_i\mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\widehat{u\mathbf{b}_i}),$$

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \sigma^2} = -\frac{1}{2}\sum_{i=1}^{n}\left[\frac{n_i}{\sigma^2} - \frac{A_i}{(\sigma^2)^2}\right],$$

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \mathbf{D}^{-1}} = -\frac{n}{\sigma^2}(-2\mathbf{D} + diag(\mathbf{D})) - \frac{1}{2}\sum_{i=1}^{n}\left(\widehat{u\mathbf{y}_i} + \widehat{u\mathbf{y}_i}^\top - diag(\widehat{u\mathbf{b}_i^2})\right).$$

The solution of $\dfrac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \boldsymbol{\beta}} = 0$ is

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\mathbf{X}_i^\top\widehat{u}_i\mathbf{X}_i\right)^{-1}\left[\sum_{i=1}^{n}\mathbf{X}_i(\widehat{u\mathbf{y}_i} - \mathbf{Z}_i\widehat{u\mathbf{b}_i})\right].$$

The solution of $\dfrac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \sigma^2} = 0$ is

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} n_i}.$$

For unstructured $\mathbf{D}$, the solution of $\dfrac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \mathbf{D}^{-1}} = 0$, for all $\mathbf{D}$, is

$$\widehat{\mathbf{D}} = \frac{\sum_{i=1}^{n}\widehat{u\mathbf{b}_i^2}}{n}.$$

A.2 The expected information matrix of the fixed effects

We developed the derivation of the expected information matrix of the fixed effects. Using the method given from McLachlan and Krishnan (1996) we have

$$I(\boldsymbol{\beta};\mathbf{y}) = I_c(\boldsymbol{\beta};\mathbf{y}) + I_m(\boldsymbol{\beta};\mathbf{y}),$$

where $I(\boldsymbol{\beta};\mathbf{y})$ is the information matrix about $\boldsymbol{\beta}$ in the observed data $\mathbf{y}$, $I_c(\boldsymbol{\beta};\mathbf{y})$ is the conditional expectation of the complete-data information matrix, and $I_m(\boldsymbol{\beta};\mathbf{y})$ is the missing information matrix.

The missing data information $I_m(\boldsymbol{\beta};\mathbf{y})$ can be expressed as

$$I_m(\boldsymbol{\beta};\mathbf{y}) = \sum_{i=1}^{n} Var\left\{S_c(\mathbf{y};\boldsymbol{\beta})|\mathbf{V}_i, \mathbf{C}_i\right\},$$

where $S_c(\mathbf{y};\boldsymbol{\beta}) = \frac{\partial logL(\mathbf{y}_c)}{\partial \boldsymbol{\beta}}$ is the gradient vector of the complete-data log likelihood function. So we have that

$$\begin{aligned} I_m(\boldsymbol{\beta};\mathbf{y}) &= \sum_{i=1}^{n} Var\left(\left(\frac{\nu+n_i}{\nu+\delta}\right)\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})|\mathbf{V}_i,\mathbf{C}_i\right) \\ &= \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{Var\left(\left(\frac{\nu+n_i}{\nu+\delta}\right)(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})|\mathbf{V}_i,\mathbf{C}_i\right)\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i. \end{aligned}$$

Now using Lange et al. (1989), the expected (complete-data) information matrix is given then by

$$I_c(\boldsymbol{\beta};\mathbf{y}) = \sum_{i=1}^{n}\frac{\nu+n_i}{\nu+n_i+2}\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i.$$

Thus the observed information matrix is given by

$$I(\boldsymbol{\beta};\mathbf{y}) = \sum_{i=1}^{n}\frac{\nu+n_i}{\nu+n_i+2}\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i - \sum_{i=1}^{n}\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{B}_i\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i,$$

where $\mathbf{B}_i = Var\left\{\frac{\nu+n_i}{\nu+\delta(\mathbf{y}_i)}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})|\mathbf{V}_i,\mathbf{C}_i\right\}$, with $\mathbf{y}_i \sim Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta},\boldsymbol{\Sigma}_i,\nu;\mathbb{A}_i)$.

A.3 Matrix algebra and vector differential calculus

We present some useful results, which may be helpful in understanding some of the results and derivations presented in this work.

Let $f(\mathbf{X})$ be a scalar function of a vector variable $\mathbf{x} = (\mathbf{x}_1,\ldots,\mathbf{x}_p)^\top$. Let

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}},\ldots,\frac{\partial f(\mathbf{x}_1)}{\partial \mathbf{x}_p}\right)^\top,$$

and let

$$\frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \left(\frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_i \partial \mathbf{x}_j} \right)_{p \times p}$$

be the $p \times p$ matrix with (i, j)-th element being $\frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_i \partial \mathbf{x}_j}$.

Let $\mathbf{A}$ and $\mathbf{B}$ are matrices. We first present the following rules for matrix algebra, which are often useful,

$$\begin{aligned}
\det(\mathbf{AB}) &= \det(\mathbf{A}) \det(\mathbf{B}), \\
\det(\mathbf{A}^{-1}) &= \det(\mathbf{A})^{-1}, \\
\det(I + \mathbf{AB}^\top) &= \det(I + \mathbf{B}^\top \mathbf{A}), \\
\mathrm{tr}(\mathbf{AB}) &= \mathrm{tr}(\mathbf{BA}),
\end{aligned}$$

where $\det(\mathbf{A})$ is the determinant of matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ is the trace of $\mathbf{A}$, and I denote the identity matrix.

Let $\mathbf{x}$ and $\mathbf{y}$ are vector functions and $\mathbf{X}$ and $\mathbf{Y}$ be matrix functions. Let $\mathbf{A}$ and $\mathbf{B}$ be constant matrices. The following are some useful differentials results

$$\begin{aligned}
d(\mathrm{tr}(\mathbf{X})) &= \mathrm{tr}(d(\mathbf{X})), \\
d(\mathbf{XY}) &= (d\mathbf{X})\mathbf{Y} + \mathbf{X}(d\mathbf{Y}), \\
d\mathbf{X}^{-1} &= -\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}, \\
d \det(\mathbf{A}) &= \det(\mathbf{A}) \mathrm{tr}(\mathbf{A}^{-1} d\mathbf{A}),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial(\mathbf{A}^\top \mathbf{x})}{\partial \mathbf{x}} &= \mathbf{A}, \\
\frac{\partial(\mathbf{x}^\top \mathbf{A} \mathbf{y})}{\partial \mathbf{x}} &= \mathbf{A}\mathbf{y}, \\
\frac{\partial(\mathbf{x}^\top \mathbf{x})}{\partial \mathbf{x}} &= 2\mathbf{x}, \\
\frac{\partial(\mathbf{x}^\top \mathbf{A} \mathbf{x})}{\partial \mathbf{x}} &= 2\mathbf{A}\mathbf{x}, \\
\frac{\partial \log(\det(\mathbf{A}(\mathbf{x})))}{\partial \mathbf{x}} &= \mathrm{tr}\left(\mathbf{A}^{-1}(\mathbf{x}) \frac{\partial \mathbf{A}(\mathbf{x})}{\partial \mathbf{x}} \right),
\end{aligned}$$

where $\mathbf{A}(\mathbf{x})$, means that matrix $\mathbf{A}$ is a function of variable $\mathbf{x}$.

Let $f(\mathbf{A})$ be a scalar function of a matrix $\mathbf{A} = (a_{ij})_{p \times p}$, and let $\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}$ be the matrix with

(i,j)-th element being $\frac{\partial f(\mathbf{A})}{\partial a_{ij}}$. We have

$$\begin{aligned}
\frac{\partial(\mathbf{x}^\top \mathbf{A}\mathbf{y})}{\partial \mathbf{A}} &= \mathbf{x}\mathbf{y}^\top, \\
\frac{\partial(\mathrm{tr}(\mathbf{A}))}{\partial \mathbf{A}} &= I, \\
\frac{\partial(\mathbf{x}^\top \mathbf{A}\mathbf{x})}{\partial \mathbf{A}} &= 2\mathbf{x}\mathbf{x}^\top - diag(\mathbf{x}\mathbf{x}^\top), \quad \text{if } \mathbf{A} \text{ is symmetric}, \\
\frac{\partial|\mathbf{A}|}{\partial \mathbf{A}} &= |\mathbf{A}|(2\mathbf{A}^{-1} - diag(\mathbf{A}^{-1})), \quad \text{if } \mathbf{A} \text{ is symmetric}, \\
\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{B})}{\partial \mathbf{A}} &= \mathbf{B} + \mathbf{B}^\top - diag(\mathbf{B}), \quad \text{if } \mathbf{A} \text{ is symmetric}, \\
\frac{\partial \log(|\mathbf{A}|)}{\partial \mathbf{A}^{-1}} &= -2\mathbf{A} + diag(\mathbf{A}), \quad \text{if } \mathbf{A} \text{ is symmetric},
\end{aligned}$$

where $diag(\mathbf{A})$ is the diagonal matrix of $\mathbf{A}$.