

**Diagnóstico de Influência em Modelos
Elípticos com Efeitos Mistos**

Felipe Alberto Osorio Salgado

Tese apresentada ao
Instituto de Matemática e Estatística
da Universidade de São Paulo
para obtenção do grau de
Doutor em Estatística

Área de concentração: **Estatística**
Orientador: **Prof. Dr. Gilberto A. Paula**

*Durante a elaboração deste trabalho, o autor recebeu
apoio financeiro de MIDEPLAN-Chile*

São Paulo, 24 de agosto de 2006

Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Felipe Alberto Osorio Salgado e aprovada pela comissão julgadora.

São Paulo, agosto de 2006.

Comissão Julgadora:

Prof. Dr. Gilberto Alvarenga Paula (Orientador)	-	IME - USP
Prof. Dr. Julio da Motta Singer	-	IME - USP
Profa. Dra. Beatriz Vaz de Melo Mendes	-	UFRJ
Profa. Dra. Reiko Aoki	-	ICMC - USP
Prof. Dr. Filidor Edilson Vilca Labra	-	UNICAMP

Para Bárbara e Vicente

Agradecimentos

Desejo expressar meus sinceros agradecimentos para meu orientador, o Professor Gilberto A. Paula, pelas suas muitas sugestões, confiança e constante ajuda durante a elaboração deste trabalho. Também sou muito grato ao Professor Manuel Galea pelo seu constante estímulo e guia na minha formação acadêmica.

Às agências financiadoras CNPq-Brasil e [MIDEPLAN](#)-Chile que me auxiliaram durante este período.

Como aluno estrangeiro tenho passado a valorizar muito a compreensão e incentivo por parte dos professores do IME-USP, em particular gostaria de agradecer à Professora Júlia M. Pavan Soler pela sua grande ajuda.

A meus pais, minha irmã Alisú e sua filha Macarena, pela sua paciência apoio e incentivo nesta etapa. Aos meus amigos Juan, Fabiana e seu pequeno filho Fausto, sempre tendo feito sentir-me como parte de sua amorosa família. Aos amigos que fiz aqui no Brasil Thales, Flávio e Adriano, particularmente pelos momentos de descontração. Pelo afeto e os momentos sinceros, em especial, de Silvina, Mariela, Jorge e sua hospitaleira família e Juvêncio. Ao grupo chileno: Germán e sua esposa Bernardita, Pedro, Cristián e Mario pela amizade.

Finalmente, a Bárbara e nosso filho Vicente que de diversas maneiras têm me encorajado para completar esta etapa. Eles têm sido uma grande fonte de apoio com seu amor, afeição e humor. Esta tese está dedicada a eles.

Resumo

Neste trabalho estudamos técnicas de diagnóstico em modelos lineares com efeitos mistos sob distribuições de contornos elípticos. O principal atrativo da classe de distribuições elípticas é que a mesma permite estender os modelos desenvolvidos sob normalidade considerando distribuições simétricas com caudas mais leves ou mais pesadas do que a normal. É conhecido que a modelagem estatística sob erros normais pode ser influenciada por observações aberrantes. Deste modo, usamos modelos baseados em distribuições com caudas mais pesadas do que a normal com o intuito de obter estimativas robustas contra observações aberrantes. Consideramos dois enfoques para introduzir distribuições elípticas no modelo linear com efeitos mistos, para cada uma dessas formulações descrevemos a estimação por máxima verossimilhança. Derivamos as curvaturas requeridas para o procedimento de influência local (Cook, 1986) para o modelo elíptico linear com efeitos mistos sob diferentes esquemas de perturbação e examinamos sua conexão com a matriz de alavancas generalizadas (Wei, Hu e Fung, 1998). Estendemos a definição da matriz de alavancas generalizadas para manipular modelos com dados incompletos. Finalmente, examinamos a habilidade dos métodos propostos para reduzir a influência de observações aberrantes e/ou influentes mediante um estudo de sensibilidade em conjuntos de dados previamente analisados sob normalidade. Disponibilizamos duas bibliotecas para o pacote S-PLUS em que as análises descritas neste trabalho podem ser realizadas. Informações relativas a tais bibliotecas estão disponíveis na página Web <http://www.ime.usp.br/~osorio/software.html>.

Abstract

In this work we consider diagnostic techniques for linear mixed-effects models derived under the class of elliptically contoured distributions. The elliptical class allows extending models developed using the normal distribution by considering symmetric distributions with either lighter or heavier-tails than the normal ones. It is well known that statistical models under normal errors may be highly influenced by outlying observations. Thus, we use models based on longer-than-normal tails with the purpose of attaining robustness aspects of the parameter estimates against outlying observations. We consider two approaches to introduce elliptical distributions in linear mixed-effects models, for each formulation we describe the maximum likelihood estimation procedure. We derive the local influence curvatures (Cook, 1986) for each case under various perturbation schemes and we examine the connection with generalized leverage (Wei, Hu e Fung, 1998). We extend the definition of the generalized leverage matrix to handle models with incomplete data. Finally, we illustrate the ability of the proposed methods to reduce the influence of outlying and influential observations through a sensitivity study in data sets previously analysed under normality. We implement the methods described in this work in two libraries for the S-PLUS package. Informations concerning there libraries are available in the Web page <http://www.ime.usp.br/~osorio/software.html>.

Sumário

Sumário	iv
Lista de Tabelas	vii
Lista de Figuras	viii
Introdução	1
1 Modelo Elíptico Linear Misto	12
1.1 Introdução	12
1.2 Modelo Linear Misto Marginal	14
1.3 Modelo Linear Misto Hierárquico	18
1.3.1 Algoritmo ECM quando \mathbf{b}_i e v_i são não observáveis .	20
1.3.2 Algoritmo ECM quando os \mathbf{b}_i são integrados	22
1.3.3 Esperanças condicionais para os algoritmos tipo-EM .	23
1.4 Considerações Computacionais	25
1.4.1 Implementação	36
2 Diagnósticos de Influência	37
2.1 Introdução	37
2.2 Influência Local	38

2.3	Método de Alavanca Generalizado para Modelos com Dados Incompletos	41
2.4	Influência Local no Modelo Linear Misto Marginal	43
2.4.1	Matriz de Informação Observada	44
2.4.2	Esquemas de Perturbação	45
2.5	Método de Alavanca Generalizado no Modelo Linear Misto Marginal	49
2.5.1	Conexão entre o Método de Alavanca Generalizado e o Procedimento de Influência Local	52
2.6	Influência Local no Modelo Linear Misto Hierárquico	53
2.6.1	Influência Local quando \mathbf{b}_i e v_i são não observáveis	53
2.6.2	Influência Local quando os v_i são não observáveis	61
2.7	Método de Alavanca Generalizado no Modelo Linear Misto Hierárquico	68
2.7.1	Conexão entre o Método de Alavanca Generalizado para Dados Incompletos e o Procedimento de Influência Local	71
2.8	Considerações Computacionais	72
2.8.1	Implementação	74
3	Aplicações	75
3.1	Dados Dentais	75
3.2	Estudo Reprodutivo em Roedores	82
3.3	Capacidade Vital Pulmonar	89
3.4	Comentários Finais	94
4	Considerações Finais	95
4.1	Conclusões	95
4.2	Perspectivas de Trabalhos Futuros	96
4.3	Desenvolvimento de Software	97

A Cálculo dos Diferenciais	98
A.1 Derivação da Curvatura no Modelo Linear Misto Marginal . . .	98
A.2 Método de Alavanca Generalizado no Modelo Linear Misto Marginal	101
A.3 Derivação da Curvatura no Modelo Linear Misto Hierárquico	102
A.3.1 Derivação da Curvatura quando \mathbf{b}_i e v_i são não observáveis	102
A.3.2 Derivação da Curvatura quando os v_i são não observáveis	105
A.4 Método de Alavanca Generalizado no Modelo Linear Misto Hierárquico	109
B Equivalência entre Medidas de Eliminação de Casos e o Método de Influência Local	111
C Diagnóstico de Influência no Modelo de Grubbs	114
Referências Bibliográficas	121

Lista de Tabelas

1.1	Valores de $q(\theta)$ para algumas distribuições elípticas.	16
3.1	Estimativas de máxima verossimilhança para os três modelos ajustados ao conjunto de dados dentais.	77
3.2	Estimativas de máxima verossimilhança para os três modelos ajustados aos dados de peso de roedores.	84
3.3	Estimativas de máxima verossimilhança para os três modelos ajustados ao conjunto de dados de capacidade pulmonar.	90

Lista de Figuras

3.1	Perfis individuais para o conjunto de dados de ortodontia. . .	76
3.2	Gráficos de índices das distâncias de Mahalanobis para os três modelos ajustados.	78
3.3	Pesos estimados para os modelos t de Student e exponencial potência.	79
3.4	Gráficos de distâncias transformadas para os três modelos ajustados.	79
3.5	Gráficos de índices de B_i para $\hat{\beta}$ sob ponderação de casos. . .	80
3.6	Gráficos de índices de B_i para $\hat{\tau}$ sob ponderação de casos. . .	80
3.7	Gráficos de índices de B_i para $\hat{\beta}$ sob perturbação da matriz de escala.	81
3.8	Gráficos de índices de B_i para $\hat{\tau}$ sob perturbação da matriz de escala.	81
3.9	Gráfico de índices de B_i para $\hat{\theta}$ sob perturbação da resposta. . .	82
3.10	Peso das crias para cada ninhada.	83
3.11	Gráficos de índices das distâncias de Mahalanobis para os três modelos ajustados.	85
3.12	Pesos estimados para os modelos t de Student e exponencial potência.	85
3.13	Gráficos de distâncias transformadas para os três modelos ajustados.	86
3.14	Gráficos de resíduos condicionais e b -distâncias para o modelo normal.	86

3.15	Gráficos de índices de B_i para $\hat{\beta}$ sob ponderação de casos. . .	87
3.16	Gráficos de índices de B_i para $\hat{\tau}$ sob ponderação de casos. . .	87
3.17	Gráficos de índices de B_i para $\hat{\psi}$ sob ponderação da matriz de escala dos efeitos aleatórios.	88
3.18	Gráfico de índices de B_i para $\hat{\beta}$ sob perturbação da resposta.	88
3.19	Gráficos de índices das distâncias de Mahalanobis para os três modelos ajustados.	91
3.20	Pesos estimados para os modelos t de Student e exponencial potência.	91
3.21	Gráficos de distâncias transformadas para os três modelos ajustados.	91
3.22	Gráficos de índices de B_i para $\hat{\mu}$ sob ponderação de casos. . .	92
3.23	Gráficos de índices de B_i para $\hat{\phi}$ sob ponderação de casos. . .	93
3.24	Gráfico de índices de B_i para $\hat{\theta}_1$ sob perturbação da resposta. . .	93

Introdução

Diagnóstico de influência constitui uma das técnicas mais úteis para modelagem estatística, em que se procura avaliar a robustez das estimativas fornecidas pelo modelo quando introduzimos perturbações no modelo e/ou nos dados, bem como avaliar afastamentos importantes das principais suposições feitas para os erros. O tipo de perturbação mais conhecido é a eliminação de casos.

O objetivo principal deste trabalho é estudar diagnóstico de influência em modelos lineares mistos sob distribuições de contornos elípticos. Realizamos, em particular, análises de sensibilidade no modelo elíptico com efeitos mistos usando a técnica de influência local (Cook, 1986) e estudamos a sua conexão com o método de alavanca generalizado (Wei, Hu e Fung, 1998).

Motivação

Um dos principais objetivos da modelagem estatística é avaliar a qualidade do modelo postulado para representar o fenômeno sob estudo. Desse modo, uma das principais preocupações é a escolha adequada do modelo. Contudo, freqüentemente existem observações que podem alterar substancialmente a inferência estatística quando são retiradas do conjunto de dados. Na literatura estatística tem havido, portanto, grande interesse na detecção de tais observações com o intuito de avaliar o impacto das mesmas no modelo ajustado, assim como considerar possíveis medidas corretivas.

Tais fatos têm motivado o desenvolvimento de duas principais linhas de pesquisa em diagnóstico de influência: *eliminação de pontos* (veja, por exemplo, Belsley, Kuh e Welsh, 1980; Cook e Weisberg, 1982 e Chatterjee e Hadi, 1988) para identificação de observações influentes, mediante a eliminação de casos ou *influência global* e a técnica de *influência local* (Cook, 1986) que consiste em estudar o efeito de introduzir pequenas perturbações

no modelo (ou dados) usando uma medida de influência apropriada. Esta metodologia tem recebido uma crescente atenção nos últimos 20 anos principalmente devido a sua flexibilidade, permitindo sua aplicação em diversas situações (veja discussão em Cook, 1997).

Um outro enfoque corresponde à acomodação de observações aberrantes (resíduo alto) através de *procedimentos robustos*, que têm sido desenvolvidos modificando o método de máxima verossimilhança sob normalidade com o intuito de atenuar o efeito de observações aberrantes nos resultados da estimação (veja, por exemplo, Huber, 1981 e Rousseeuw e Leroy, 1987). Sob esse enfoque também podemos considerar distribuições simétricas com caudas mais pesadas do que a distribuição normal. Uma escolha interessante neste sentido corresponde à classe de distribuições de contornos elípticos. O principal atrativo desta classe é que permite estender os modelos desenvolvidos sob erro normal considerando distribuições simétricas com caudas mais leves ou mais pesadas do que a normal. Alguns exemplos de distribuições nesta classe são à t de Student, exponencial potência e normal contaminada, modelos conhecidos por acomodar observações aberrantes. Uma descrição detalhada da classe de distribuições elípticas multivariadas é apresentada, por exemplo, em Fang e Zhang (1990), Fang, Kotz e Ng (1990) e Arellano (1994).

Vários autores têm considerado o uso de distribuições na classe elíptica para modelagem robusta. Por exemplo, Dempster, Laird e Rubin (1980) consideram a estimação em modelos de regressão linear através de mínimos quadrados iterativamente ponderados supondo uma subclasse das elípticas conhecida como família de misturas de escala normal; Little (1988) aborda a estimação em modelos com dados incompletos usando a normal contaminada, Lange, Little e Taylor (1989) propõem a distribuição t para modelagem robusta em modelos de regressão tanto univariados como multivariados, Lange e Sinsheimer (1993) consideram distribuições na classe das misturas de escala normal para estimação robusta em regressão linear e não-linear, Bolfarine e Galea (1996) discutem a estimação de máxima verossimilhança no modelo de calibração comparativa estrutural usando a distribuição t enquanto Fernández e Steel (1999) estudam a estimação de parâmetros em modelos lineares multivariados via máxima verossimilhança usando a distribuição t e alertam com relação a problemas na estimação dos graus de liberdade. Cysneiros e Paula (2005) consideram métodos restritos em modelos de regressão simétricos univariados e, em particular para modelos lineares com efeitos mistos Huggins (1993a, 1993b) propõe M -estimadores baseados em distribuições de contornos elípticos, com ênfase em modelos para

componentes de variância, Lindsey (1999) e Lindsey e Lindsey (2005) consideram a distribuição exponencial potência para problemas com medidas repetidas, Pinheiro, Liu e Wu (2001) propõem algoritmos para estimação por máxima verossimilhança em modelos lineares com efeitos mistos considerando a distribuição t multivariada e mais recentemente, Savalli, Paula e Cysneiros (2006) discutem testes para componentes de variância em modelos elípticos lineares mistos. Num contexto bayesiano e principalmente usando distribuições de mistura de escala normal, Liu (1996) estuda o modelo de regressão linear com dados incompletos, Chen e Dey (1998) consideram a modelagem de funções de ligação para modelos com resposta binária e Choy e Smith (1997), Rosa, Padovani e Gianola (2003, 2004) e Portela e Gómez-Villegas (2004) desenvolvem implementações MCMC para estimação em modelos lineares com efeitos mistos.

Contudo, esses modelos desenvolvidos sob distribuições com caudas mais pesadas do que a normal, ainda podem estar vulneráveis a observações influentes como é notado, por exemplo, em Galea, Paula e Bolfarine (1997), Galea, Riquelme e Paula (2000), Liu (2000, 2002), Díaz-García, Galea e Leiva-Sánchez (2003), Galea, Paula e Uribe-Opazo (2003), Galea, Paula e Cysneiros (2005) e Savalli, Paula e Cysneiros (2006). Portanto, existe a necessidade de realizar estudos de sensibilidade nesta classe de modelos.

O objetivo deste trabalho é estudar diagnóstico de influência em modelos lineares com efeitos mistos considerando distribuições de contornos elípticos. Na seção a seguir apresentamos um resumo de métodos de diagnóstico aplicados ao modelo linear misto, posteriormente fazemos uma breve descrição da classe de distribuições elípticas e do algoritmo de maximização tipo EM. Finalmente, indicamos os objetivos assim como a organização do trabalho.

Diagnósticos em Modelos com Efeitos Mistos

Métodos de diagnóstico e validação aplicados em modelos lineares com efeitos mistos (Laird e Ware, 1982) têm sido desenvolvidos principalmente sob a suposição de normalidade. Baseados em procedimentos de eliminação de observações, Christensen, Pearson e Johnson (1992) estudam a influência de observações num modelo de efeitos aleatórios para análise de variância enquanto Tan, Ouwens e Berger (2001) discutem a detecção de observações influentes em modelos lineares mistos com ênfase em estudos longitudinais. Diagnósticos de influência global em modelos de curvas de crescimento são considerados por Pan (2002) e Zewotir e Galpin (2005) que estendem os

resultados de Christensen et al. (1992). Em parte devido a complexidade do modelo linear com efeitos mistos, alguns autores têm aplicado o método de influência local para realizar estudos de sensibilidade. Por exemplo, Beckman, Nachtsheim e Cook (1987) estudam diagnóstico de influência num modelo misto para análise de variância. Em modelos de curvas de crescimento a metodologia de influência local é investigada por Pan, Fang e von Rosen (1997), Pan e Bai (2003) e Shi e Ojeda (2004). Lesaffre e Verbeke (1998) consideram o esquema de ponderação de casos no modelo linear com efeitos mistos usando uma perspectiva marginal.

Atualmente, grande parte das metodologias desenvolvidas sob normalidade estão sendo estendidas para a classe de distribuições elípticas. Contudo, pouco tem sido investigado em modelos lineares com efeitos mistos sob distribuições de contornos elípticos. Este trabalho é motivado pelo fato de tais procedimentos de diagnóstico, além de identificar observações influentes, também permite selecionar modelos paramétricos dentro da classe de distribuições elípticas que representam boas alternativas à distribuição normal.

Distribuições de Contornos Elípticos

A classe de distribuições elípticas tem recebido crescente atenção na literatura estatística nos últimos 15 anos (veja, Fang e Zhang, 1990; Fang, Kotz e Ng, 1990 e Arellano, 1994), particularmente devido ao fato de incluir distribuições com caudas mais pesadas do que a normal, tais como a distribuição t de Student, exponencial potência e normal contaminada, dentre outras. Uma subclasse importante na família de distribuições de contornos elípticos corresponde à classe das misturas de escala normal (veja, por exemplo, Andrews e Mallows, 1974; Dempster, Laird e Rubin, 1980; Lange e Sinsheimer, 1993; Chen e Dey, 1998 e Gómez, Gómez-Villegas e Marín, 2006) que apresenta boas propriedades como a normal, é relativamente simples de trabalhar e permite robustificar procedimentos estatísticos. A seguir introduzimos as definições de distribuições elípticas que serão utilizadas neste trabalho.

Definição 1 (distribuição elíptica). Dizemos que o vetor aleatório m -dimensional \mathbf{Y} tem *distribuição elíptica multivariada* com vetor de posição $\boldsymbol{\mu} \in \mathbb{R}^m$ e matriz de escala positiva definida $\boldsymbol{\Lambda}$ se sua função de densidade assume a forma

$$f(\mathbf{y}) = |\boldsymbol{\Lambda}|^{-1/2} g[(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{y} - \boldsymbol{\mu})], \quad (1)$$

em que $g : \mathbb{R} \rightarrow [0, \infty)$ satisfaz a condição $\int_0^\infty u^{m/2-1}g(u) du < \infty$. A função g é tipicamente conhecida como função geradora de densidade. Quando um vetor aleatório tem densidade dada por (1) usamos a notação $\mathbf{Y} \sim EC_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}; g)$.

Um estudo completo das propriedades da classe de distribuições elípticas assim como alguns exemplos podem ser encontrados em Fang e Zhang (1990), Fang, Kotz e Ng (1990) e Arellano (1994). A seguir apresentamos uma propriedade que estabelece que a classe de distribuições de contornos elípticos é fechada sob transformações lineares.

Proposição 1. *Seja $\mathbf{Y} \sim EC_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}; g)$. Se $\mathbf{a} \in \mathbb{R}^n$ e \mathbf{A} é uma matriz $n \times m$ com $\text{posto}(\mathbf{A}) = n \leq m$, então*

$$\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{a} \sim EC_n(\mathbf{A}\boldsymbol{\mu} + \mathbf{a}, \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}^T; g). \quad (2)$$

A partir de (2) podemos notar que qualquer distribuição marginal de $\mathbf{Y} \sim EC_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}; g)$ também é elíptica.

Definição 2 (distribuição de mistura de escala normal). Seja $\boldsymbol{\mu} \in \mathbb{R}^m$, $\boldsymbol{\Lambda}$ matriz positiva definida $m \times m$, $H(v; \boldsymbol{\nu})$ função de distribuição de probabilidade (unidimensional) e $\kappa(\cdot)$ uma função estritamente positiva, então a densidade f definida como

$$f(\mathbf{y}) = |2\pi\boldsymbol{\Lambda}|^{-1/2} \int_0^\infty \{\kappa(v)\}^{-m/2} \exp\{-\frac{1}{2}\kappa^{-1}(v)u\} dH(v), \quad (3)$$

é dita uma *densidade de mistura de escala normal*, em que $u = (\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$ e H é conhecida como a distribuição da variável de mistura V . Se um vetor aleatório \mathbf{Y} tem densidade dada por (3) dizemos que \mathbf{Y} tem distribuição de mistura de escala normal e escrevemos $\mathbf{Y} \sim SMN_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}; H)$.

Um vetor aleatório m -dimensional \mathbf{Y} tem distribuição $SMN_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}; H)$ se admite a representação

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \kappa^{1/2}(V)\mathbf{Z}, \quad (4)$$

em que $\mathbf{Z} \sim N_m(\mathbf{0}, \boldsymbol{\Lambda})$, V é variável aleatória positiva com distribuição H independente de \mathbf{Z} . Aqui $\stackrel{d}{=}$ denota igualdade em distribuição. Logo, a partir de (4) obtemos que a distribuição de \mathbf{Y} condicional a $V = v$ é $(\mathbf{Y}|V = v) \sim N_m(\boldsymbol{\mu}, \kappa(v)\boldsymbol{\Lambda})$.

Alguns exemplos de distribuições elípticas são dados a seguir (veja também, Fang, Kotz e Ng, 1990 e Arellano, 1994). Para cada caso considere $u = (\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$.

Distribuição t de Student generalizada. Seja $\mathbf{Y} \sim Gt_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu, \gamma)$ a distribuição t de Student generalizada com $\nu, \gamma > 0$. Isto é, \mathbf{Y} tem densidade dada por

$$f(\mathbf{y}) = \frac{\Gamma(\frac{m+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{m/2}} \gamma^{-m/2} |\boldsymbol{\Lambda}|^{-1/2} \left(1 + \frac{u}{\gamma}\right)^{-(m+\nu)/2}. \quad (5)$$

Note que, quando $\gamma = \nu$ obtemos a distribuição t de Student, $t_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ com $\nu > 0$ graus de liberdade, e a distribuição de Cauchy quando $\gamma = \nu = 1$.

Distribuição Exponencial Potência. Se $\mathbf{Y} \sim PE_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ com parâmetro de forma $\nu > 0$ (Gómez, Gómez-Villegas e Marín, 1998) então sua densidade é da forma

$$f(\mathbf{y}) = \frac{\nu \Gamma(\frac{m}{2})}{\pi^{m/2} \Gamma(\frac{m}{2\nu}) 2^{m/2\nu}} |\boldsymbol{\Lambda}|^{-1/2} \exp(-u^\nu/2). \quad (6)$$

Esta família de distribuições apresenta ambos tipos de caudas, mais leves ($\nu > 1$) assim como mais pesadas do que a normal ($\nu < 1$) e inclui a normal como um caso particular ($\nu = 1$).

Distribuição Normal Contaminada. Considere $\mathbf{Y} \sim NC_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \delta, \gamma)$ a distribuição normal contaminada m-variada (Little, 1988) em que $0 \leq \delta \leq 1$ denota a percentagem de “contaminação” e $\gamma > 0$ corresponde a um fator de inflação de escala. Neste caso a densidade de \mathbf{Y} é dada por

$$f(\mathbf{y}) = (2\pi)^{-m/2} |\boldsymbol{\Lambda}|^{-1/2} \{\delta \gamma^{m/2} e^{-\gamma u/2} + (1 - \delta) e^{-u/2}\}. \quad (7)$$

Distribuição Slash. Seja $\mathbf{Y} \sim Slash_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ com parâmetro de forma $\nu > 0$ (Rogers e Tukey, 1972), então sua densidade fica dada por

$$f(\mathbf{y}) = \nu (2\pi)^{-m/2} |\boldsymbol{\Lambda}|^{-1/2} \int_0^1 v^{m/2+\nu-1} \exp(-vu/2) dv. \quad (8)$$

Distribuição Logística tipo-II. Considere $\mathbf{Y} \sim L_m^{II}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, então a densidade do vetor aleatório m -dimensional \mathbf{Y} assume a forma

$$f(\mathbf{y}) = c_m |\boldsymbol{\Lambda}|^{-1/2} \frac{e^{-\sqrt{u}}}{(1 + e^{-\sqrt{u}})^2}, \quad (9)$$

em que a constante de padronização c_m é dada por

$$c_m = \frac{\pi^{m/2}}{\Gamma(\frac{m}{2})} \int_0^\infty t^{m/2-1} \frac{e^{-\sqrt{t}}}{(1 + e^{-\sqrt{t}})^2} dt.$$

A propriedade (4) para a classe de distribuições de mistura de escala normal é particularmente útil para a estimação de parâmetros por máxima verossimilhança mediante o algoritmo de maximização tipo EM. A seguir apresentamos uma descrição deste algoritmo.

Algoritmo EM

O algoritmo EM (Dempster, Laird e Rubin, 1977) é um enfoque amplamente aplicável para o cálculo iterativo de estimativas de máxima verossimilhança, sendo bastante útil para problemas com dados incompletos.

Muitos problemas em Estatística podem ser encarados utilizando uma *formulação de dados aumentados* permitindo assim simplificar a obtenção das estimativas de máxima verossimilhança. Os dados aumentados, também chamados *dados completos*, correspondem aos dados observados, referidos nesta formulação como *dados incompletos*, e dados adicionais conhecidos como *dados perdidos*. Neste contexto, as funções de verossimilhança baseadas nos dados completos e observados são denominadas verossimilhança de dados completos e dados incompletos, respectivamente. É importante salientar que a parte aumentada dos dados não requer que eles sejam “perdidos” no sentido estrito da palavra, pois somente representam um mecanismo técnico. De fato, esta idéia é usada para descrever uma variedade de modelos estatísticos, tais como: misturas, efeitos aleatórios, agrupamentos, censura e observações incompletas (veja, por exemplo, McLachlan e Krishnan, 1997. Cap. 2).

No algoritmo EM, cada iteração é formada pelos passos: Esperança (passo E) e Maximização (passo M). Para introduzir idéias, sejam \mathbf{Y} o vetor de dados observados e \mathbf{z} o vetor de dados perdidos; desse modo o vetor de dados

completos \mathbf{Y}_c aumenta \mathbf{Y} mediante \mathbf{z} como $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{z}^T)^T$. Considere $L(\boldsymbol{\theta}|\mathbf{Y}_c)$ o logaritmo da função de verossimilhança de dados completos para o vetor de parâmetros $\boldsymbol{\theta} \in \Theta$. O algoritmo EM aborda problemas com dados incompletos indiretamente mediante a substituição da parte não observável em $L(\boldsymbol{\theta}|\mathbf{Y}_c)$ por suas esperanças condicionais dado \mathbf{Y} , usando o ajuste atual para $\boldsymbol{\theta}$. Isto é, considera a função Q definida como

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \text{E}\{L(\boldsymbol{\theta}|\mathbf{Y}_c)|\mathbf{Y}, \hat{\boldsymbol{\theta}}\}. \quad (10)$$

A $(r+1)$ -ésima iteração do algoritmo EM é definida como

Passo E: para $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(r)}$, calcular $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ como

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \text{E}\{L(\boldsymbol{\theta}|\mathbf{Y}_c)|\mathbf{Y}, \hat{\boldsymbol{\theta}}\}.$$

Passo M: escolher $\boldsymbol{\theta}^{(r+1)}$ que maximize $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ tal que

$$Q(\boldsymbol{\theta}^{(r+1)}|\hat{\boldsymbol{\theta}}) \geq Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}), \quad \forall \boldsymbol{\theta} \in \Theta.$$

Deve-se alternar os passos E e M repetidamente até atingir a convergência. Cada iteração do algoritmo EM incrementa a função de verossimilhança de dados observados $L(\boldsymbol{\theta}|\mathbf{Y})$ e sob condições apropriadas o algoritmo EM apresenta convergência monótona ao máximo global ou local de $L(\boldsymbol{\theta}|\mathbf{Y})$ (Wu, 1983 e McLachlan e Krishnan, 1997, Sec. 3.5).

Quando o passo M no algoritmo EM é complicado, este pode ser amenizado realizando o processo de maximização condicional a alguma função dos parâmetros que estão sendo estimados. Este algoritmo EM generalizado proposto por Meng e Rubin (1993) é denominado algoritmo de maximização condicional de esperança (ECM). A idéia neste caso é substituir o passo M no algoritmo EM por uma seqüência de passos de maximização condicional (CM) computacionalmente mais simples, em que cada um deles maximiza a função Q sujeita a restrições em $\boldsymbol{\theta}$, de modo que essas restrições garantam que a maximização é realizada sobre todo o espaço paramétrico para $\boldsymbol{\theta}$. Como as maximizações CM são efetuadas sobre espaços de dimensão menor, freqüentemente tais etapas são mais simples, velozes e estáveis que suas contrapartidas realizadas sobre todo o espaço paramétrico Θ . O algoritmo ECM tipicamente converge num número maior de iterações que o algoritmo EM, porém pode ser mais veloz em tempo de computação.

Em muitas situações, o cálculo do passo E pode ser de custo computacional muito menor do que os passos CM. Desse modo podemos ter interesse em executar um passo E antes de cada passo CM, o que é conhecido como um ciclo. Tal algoritmo é denominado algoritmo ECM multiciclo (Meng e Rubin, 1993). É importante notar que as propriedades de simplicidade, estabilidade e convergência monótona do algoritmo EM são compartilhadas pelos algoritmos ECM e ECM multiciclo, porém com taxas de convergência mais velozes.

Definição dos Objetivos e Organização do Trabalho

O objetivo geral deste trabalho é desenvolver técnicas para análise de sensibilidade em modelos lineares com efeitos mistos sob distribuições de contornos elípticos considerando os métodos de influência local (Cook, 1986) e de alavanca generalizado (Wei, Hu e Fung, 1998). É um fato conhecido que este tipo de distribuição com caudas mais pesadas permite acomodar observações aberrantes. Descrevemos dois enfoques para introduzir distribuições elípticas no modelo linear misto, uma formulação *marginal* e uma *hierárquica* baseadas nas propostas de Lange, Little e Taylor (1989) e Pinheiro, Liu e Wu (2001), respectivamente. Ambas abordagens podem ser consideradas como procedimentos para acomodação de observações aberrantes. Finalmente, aplicamos os resultados desenvolvidos a três conjuntos de dados. Deste modo, os objetivos específicos deste trabalho são:

- 1) Abordar a estimação por máxima verossimilhança no modelo linear com efeitos mistos tanto na sua formulação marginal como hierárquica. No caso do modelo em sua formulação hierárquica discutir a estimação mediante algoritmos tipo-EM.
- 2) Derivar as curvaturas requeridas para o procedimento de influência local para o modelo elíptico linear com efeitos mistos sob diferentes esquemas de perturbação.
- 3) Aplicar o método de alavanca generalizado ao modelo linear com efeitos mistos sob distribuições de contornos elípticos e examinar sua conexão com o procedimento de influência local.
- 4) Examinar a capacidade dos métodos propostos para reduzir a influência de observações aberrantes mediante um estudo de sensibilidade em conjuntos de dados previamente analisados sob normalidade.

O trabalho contido nesta tese é organizado em quatro capítulos e três apêndices.

O Capítulo 1 é dedicado ao estudo do modelo linear com efeitos mistos considerando dois enfoques para a introdução de distribuições de contornos elípticos, denominados neste trabalho como formulações marginal e hierárquica. Para o caso da formulação hierárquica consideramos uma subclasse das elípticas conhecida como família de misturas de escala normal (Andrews e Mallows, 1974). Mostramos que a distribuição marginal para as respostas observadas também pertence à classe das misturas de escala normal, simplificando desse modo a inferência estatística uma vez que as densidades envolvidas podem ser avaliadas de maneira simples, fato que não é compartilhado por muitos modelos com efeitos mistos desenvolvidos para dados não normais. Na atualidade, o sucesso de qualquer técnica estatística está intimamente relacionado à disponibilidade de software confiável e eficiente para o ajuste dos modelos por parte de usuários de estatística (veja Stromberg, 2004). Devido a tal fato, desenvolvemos métodos para a estimação de parâmetros no modelo linear com efeitos mistos em sua formulação hierárquica. Disponibilizamos uma biblioteca escrita em S e C para o pacote S-PLUS em que tais análises podem ser realizados.

No Capítulo 2 desenvolvemos diagnósticos de influência para o modelo elíptico linear misto mediante o método de influência local (Cook, 1986) e de alavanca generalizado (Wei et al., 1998). Derivamos a curvatura normal em modelos lineares mistos sob distribuições elípticas considerando diversos esquemas de perturbação e examinamos a sua conexão com o método de alavanca generalizado. Estendemos a definição da matriz de alavancas generalizadas para manipular modelos com dados incompletos. As quantidades de interesse para tais procedimentos de diagnóstico são calculadas eficientemente usando o método de diferenciação matricial descrito em Magnus e Neudecker (1988). Os principais resultados neste tópico são apresentados no Apêndice A. No Apêndice B examinamos a equivalência entre alguns métodos de diagnóstico por eliminação de casos e o procedimento de influência local.

No Capítulo 3 aplicamos a metodologia desenvolvida em alguns conjuntos de dados provenientes de estudos em biometria. O exemplo final corresponde a uma aplicação do modelo de Grubbs (Grubbs, 1948, 1973, 1983). Tal modelo pode ser visto como um caso particular do modelo linear com efeitos mistos. Contudo, desenvolvemos a influência para este modelo considerando-o como um caso particular do modelo de Barnett (Barnett, 1969). Esses resultados são resumidos no Apêndice C. Mediante a análise de influência para cada

um desses conjuntos de dados, notamos que os modelos desenvolvidos sob distribuições de contornos elípticos apresentam propriedades de robustez contra observações aberrantes assim como contra os vários esquemas de perturbação considerados.

Finalmente, no Capítulo 4, apresentamos algumas considerações finais e perspectivas de trabalhos futuros.

Capítulo 1

Modelo Elíptico Linear Misto

Neste capítulo apresentamos dois enfoques para introduzir distribuições de contornos elípticos no modelo linear com efeitos mistos. Descrevemos a estimação por máxima verossimilhança para cada um dos enfoques, finalmente abordamos alguns aspectos computacionais.

1.1 Introdução

Considere o modelo linear com efeitos mistos (Laird e Ware, 1982) dado por

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M, \quad (1.1)$$

em que \mathbf{Y}_i representa o vetor aleatório n_i -dimensional das respostas observadas para o i -ésimo indivíduo ou grupo, \mathbf{X}_i e \mathbf{Z}_i são matrizes de planejamento $n_i \times p$ e $n_i \times q$, respectivamente, $\boldsymbol{\beta}$ é um vetor p -dimensional de efeitos fixos, \mathbf{b}_i denota um vetor q -dimensional de efeitos aleatórios e $\boldsymbol{\epsilon}_i$ indica um vetor de erros.

Um dos principais motivos pelos quais os modelos com efeitos mistos têm-se tornado populares deriva da sua flexibilidade para modelar a estrutura de correlação intra-indivíduos, a capacidade de lidar tanto com dados balanceados como desbalanceados, assim como a disponibilidade de pacotes confiáveis para o ajuste de tais modelos, dentre os quais podemos destacar a biblioteca para S-PLUS/R, [nlme](#) (Pinheiro e Bates, 2000) o procedimento SAS, Proc MIXED (Little et al., 1988) e em particular para modelos longitudinais a biblioteca para S-PLUS [OSWALD](#) (Smith, Robertson e Diggle, 1996).

Normalidade é assumida freqüentemente tanto para os efeitos aleatórios \mathbf{b}_i quanto para os erros ϵ_i , e as inferências clássicas são baseadas no modelo marginal para \mathbf{Y}_i (veja por exemplo, Pinheiro e Bates, 2000 e Verbeke e Molenberghs, 2001). Contudo, a suposição de normalidade tem sido criticada pelo fato de não ser realista, além de ser bem conhecido que a presença de observações aberrantes pode distorcer as estimativas de máxima verossimilhança assim como a inferência estatística. Isto tem motivado o desenvolvimento de metodologias robustas com o interesse de atenuar o efeito de tais observações.

Para acomodação de observações aberrantes em modelos lineares com efeitos mistos, diversos procedimentos têm sido propostos. Por exemplo, Richardson e Welsh (1995) discutem a aplicação de métodos robustos para máxima verossimilhança restrita, Richardson (1997) considera a estimação de influência limitada para reduzir a ação de observações aberrantes tanto na resposta quanto nas variáveis explicativas, Gill (2000) sugere robustificar o logaritmo da função de verossimilhança e aplica o procedimento proposto para estimação robusta em dados longitudinais e He, Cui e Simpson (2004) propõem um método robusto para a análise de dados longitudinais considerando que a função que limita a influência de observações aberrantes, a função ρ na literatura de métodos robustos (veja Huber, 1981), corresponde ao logaritmo da densidade t de Student univariada.

Neste trabalho consideramos dois enfoques para introduzir distribuições de contornos elípticos no modelo linear com efeitos mistos. Tais enfoques são baseados no modelo marginal descrito por Lange, Little e Taylor (1989) e a representação hierárquica dada em Pinheiro, Liu e Wu (2001). Ambos podem ser considerados métodos para acomodação de dados aberrantes e têm a possibilidade de trazer informações úteis para a detecção de tais observações.

Primeiramente, consideramos o modelo (1.1) em que \mathbf{b}_i e ϵ_i são tais que $(\mathbf{b}_i^T, \epsilon_i^T)^T$ segue uma distribuição de contornos elípticos com parâmetro de posição $\mathbf{0}$, matriz de escala $\mathbf{\Lambda}_i = \mathbf{\Psi} \oplus \mathbf{\Sigma}_i$ e função geradora de densidade g , em que $\mathbf{\Psi}$ é uma matriz simétrica $q \times q$ e $\mathbf{\Sigma}_i$ é uma matriz simétrica $n_i \times n_i$. Deste modo, temos que

$$\begin{pmatrix} \mathbf{Y}_i \\ \mathbf{b}_i \end{pmatrix} \stackrel{ind}{\sim} EC_{n_i+q} \left(\begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i \mathbf{\Psi} \mathbf{Z}_i^T + \mathbf{\Sigma}_i & \mathbf{Z}_i \mathbf{\Psi} \\ \mathbf{\Psi} \mathbf{Z}_i^T & \mathbf{\Psi} \end{pmatrix}; g \right), \quad i = 1, \dots, M.$$

Usualmente a inferência estatística é baseada no modelo marginal para o vetor de respostas. Usando as propriedades da classe elíptica, temos que os

\mathbf{Y}_i são independentes e marginalmente distribuídos como

$$\mathbf{Y}_i \stackrel{ind}{\sim} EC_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i; g), \quad i = 1, \dots, M, \quad (1.2)$$

em que $\mathbf{V}_i = \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$. Consideramos que $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$ e $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\lambda})$ são parametrizadas por $\boldsymbol{\alpha}$ e $\boldsymbol{\lambda}$ vetores k e d dimensionais, respectivamente. O modelo dado por (1.2) será chamado *modelo linear misto marginal*. Uma aplicação deste modelo ao estudo de dados longitudinais é apresentado em Osorio, Paula e Galea (2006).

Usando a classe de distribuições de mistura de escala normal podemos expressar o *modelo linear misto hierárquico* (veja Pinheiro, Liu e Wu, 2001) como

$$\begin{aligned} \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{b}_i \end{pmatrix} \Big| v_i \stackrel{ind}{\sim} N_{n_i+q} \left(\begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \kappa(v_i) \begin{pmatrix} \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i & \mathbf{Z}_i\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\mathbf{Z}_i^T & \boldsymbol{\Psi} \end{pmatrix} \right) \text{ e} \\ v_i \stackrel{ind}{\sim} H(v_i; \boldsymbol{\nu}), \quad i = 1, \dots, M, \end{aligned} \quad (1.3)$$

ou, alternativamente, como

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b}_i, v_i \stackrel{ind}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \kappa(v_i)\boldsymbol{\Sigma}_i), \quad \mathbf{b}_i | v_i \stackrel{ind}{\sim} N_q(\mathbf{0}, \kappa(v_i)\boldsymbol{\Psi}) \text{ e} \\ v_i \stackrel{ind}{\sim} H(v_i; \boldsymbol{\nu}), \quad i = 1, \dots, M, \end{aligned} \quad (1.4)$$

neste caso $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$ e $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\lambda})$ representam as matrizes de escala dos erros $\boldsymbol{\epsilon}_i$ e dos efeitos aleatórios \mathbf{b}_i , respectivamente, $H(v_i, \boldsymbol{\nu})$ denota a função de distribuição da variável de mistura e $\kappa(\cdot)$ é uma função positiva. Devido às recomendações dadas em Fernández e Steel (1999), neste trabalho consideramos fixados os parâmetros de forma da distribuição elíptica para o enfoque marginal, assim como os parâmetros relativos à variável de mistura v_i no modelo hierárquico.

1.2 Modelo Linear Misto Marginal

Considere $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$, $i = 1, \dots, M$, vetores aleatórios independentes seguindo o modelo marginal dado em (1.2), isto é, $\mathbf{Y}_i \sim EC_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i; g)$ para $i = 1, \dots, M$. O logaritmo da função de verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ é dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^M L_i(\boldsymbol{\theta}) = \sum_{i=1}^M \left\{ -\frac{1}{2} \log |\mathbf{V}_i| + \log g(u_i) \right\}, \quad (1.5)$$

em que $u_i = (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})$, $i = 1, \dots, M$, representa uma medida de distância ponderada pela matriz \mathbf{V}_i . Assumindo que $g(\cdot)$ é contínua e diferenciável, podemos definir as quantidades

$$W_g(u) = \frac{d}{du} \log g(u) = \frac{g'(u)}{g(u)} \quad \text{e} \quad W'_g(u) = \frac{d}{du} W_g(u).$$

Seja $\boldsymbol{\tau} = (\boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$; as funções escore para $\boldsymbol{\theta}$ são dadas por

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^M q_i(\boldsymbol{\theta}) \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \quad \text{e} \quad (1.6)$$

$$\mathbf{U}(\boldsymbol{\tau}) = (U(\tau_1), \dots, U(\tau_l))^T, \quad (1.7)$$

em que

$$U(\tau_j) = -\frac{1}{2} \sum_{i=1}^M \{\text{tr} \mathbf{V}_i^{-1} \dot{\mathbf{V}}_i(j) - q_i(\boldsymbol{\theta}) \mathbf{r}_i^T \mathbf{V}_i^{-1} \dot{\mathbf{V}}_i(j) \mathbf{V}_i^{-1} \mathbf{r}_i\},$$

para $j = 1, \dots, l$; $l = k+d$, $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ e $q_i(\boldsymbol{\theta}) = -2W_g(u_i)$, $i = 1, \dots, M$. Usamos a notação $\dot{\mathbf{V}}_i(j) = \partial \mathbf{V}_i / \partial \tau_j$.

Considere o seguinte procedimento iterativo para obter a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$:

Etapa 1. Avaliar as quantidades $q_i^{(r)} = q_i(\boldsymbol{\theta}^{(r)})$, $i = 1, \dots, M$ e para $\boldsymbol{\tau}$ fixado, atualizar $\boldsymbol{\beta}^{(r+1)}$ como

$$\boldsymbol{\beta}^{(r+1)}(\boldsymbol{\tau}) = \left(\sum_{i=1}^M q_i^{(r)} \mathbf{X}_i^T \mathbf{V}_i^{-1}(\boldsymbol{\tau}) \mathbf{X}_i \right)^{-1} \sum_{i=1}^M q_i^{(r)} \mathbf{X}_i^T \mathbf{V}_i^{-1}(\boldsymbol{\tau}) \mathbf{Y}_i, \quad (1.8)$$

em que $\mathbf{V}_i(\boldsymbol{\tau}) = \mathbf{Z}_i \boldsymbol{\Psi}(\boldsymbol{\lambda}) \mathbf{Z}_i^T + \boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$, $i = 1, \dots, M$.

Etapa 2. Atualizar $\boldsymbol{\tau}^{(r+1)}$ mediante maximizar o logaritmo da verossimilhança concentrada, $L(\boldsymbol{\beta}^{(r+1)}(\boldsymbol{\tau}), \boldsymbol{\tau}) = \sum_{i=1}^M L_i(\boldsymbol{\beta}^{(r+1)}(\boldsymbol{\tau}), \boldsymbol{\tau})$, em que

$$L_i(\boldsymbol{\beta}^{(r+1)}(\boldsymbol{\tau}), \boldsymbol{\tau}) = -\frac{1}{2} \log |\mathbf{V}_i| + \log g(\tilde{u}_i)$$

com $\tilde{u}_i = \hat{\mathbf{r}}_i^T \mathbf{V}_i^{-1}(\boldsymbol{\tau}) \hat{\mathbf{r}}_i$, em que $\hat{\mathbf{r}}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(r+1)}(\boldsymbol{\tau})$. Desse modo atualizamos $\boldsymbol{\tau}^{(r+1)}$ como

$$\boldsymbol{\tau}^{(r+1)} = \arg \max_{\boldsymbol{\tau}} \{L(\boldsymbol{\beta}^{(r+1)}(\boldsymbol{\tau}), \boldsymbol{\tau})\}. \quad (1.9)$$

Deve-se repetir as etapas 1 e 2 até atingir convergência. Para efetuar a maximização em (1.9) podemos considerar um algoritmo secante multivariado (veja, por exemplo, Dennis e Schnabel, 1996) em que as funções gradientes requeridas pelo método são dadas pelas funções escore em (1.7).

Na Tabela 1.1 apresentamos o valor de $q(\boldsymbol{\theta})$ para algumas distribuições de contornos elípticos.

Tabela 1.1: Valores de $q(\boldsymbol{\theta})$ para algumas distribuições elípticas.

Distribuição	$q(\boldsymbol{\theta}) = -2W_g(u)$
Normal	1
t de Student	$(\nu + n)/(\nu + u)$
t de Student generalizada	$(\nu + n)/(\gamma + u)$
Pearson VII	$2s/(r + u)$
Logística I	$2 \tanh(u/2)$
Logística II	$u^{1/2} \tanh(u^{1/2}/2)$
Exponencial potência	$\nu u^{\nu-1}, \nu \neq \frac{1}{2}$
Normal contaminada	$h_1(u)/h_0(u)^a$

$$^a h_j(u) = 1 - \delta + \delta \gamma^{n/2+j} e^{(1-\gamma)u/2}, j = 0, 1$$

A quantidade $q_i(\boldsymbol{\theta}) = -2W_g(u_i)$ que aparece nas equações (1.6)-(1.7) pode ser interpretada como um peso e, dado que $g(u_i)$ é em geral uma função decrescente temos que $q_i(\boldsymbol{\theta}) > 0$ para grande parte dos modelos elípticos. Em particular, para as distribuições t de Student e exponencial potência ($\nu < 1$), $q_i(\boldsymbol{\theta})$ é inversamente proporcional à distância de Mahalanobis. Desse modo, observações com distâncias u_i grandes receberão pesos $q_i(\boldsymbol{\theta})$ pequenos, e portanto o procedimento de estimação em (1.8)-(1.9) tende a produzir estimativas de parâmetros mais robustas contra observações aberrantes do que o modelo normal. Esta influência é controlada pelos parâmetros associados à função geradora, g . Tais parâmetros atuam como constantes reguladoras (ou afinadoras) do procedimento de estimação, que por sua vez corresponde a um tipo de estimação M (Maronna, 1976; Lange et al., 1989 e Kowalski et al., 1999).

A estimação dos efeitos aleatórios \mathbf{b}_i através do modelo linear misto marginal bem como testes para os componentes de variância são discutidos em Savalli, Paula e Cysneiros (2006).

Estimação do Erro Padrão das Estimativas dos Parâmetros

Vários procedimentos podem ser considerados para estimar o erro padrão dos estimadores de máxima verossimilhança no modelo elíptico linear misto. Neste trabalho estimamos o erro padrão das estimativas de máxima verossimilhança mediante a matriz de informação de Fisher, que assume a forma bloco diagonal (Lange et al., 1989; Mitchell, 1989)

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\boldsymbol{\tau}) \end{pmatrix},$$

em que

$$\mathbf{K}(\boldsymbol{\beta}) = \sum_{i=1}^M \frac{4d_{gi}}{n_i} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \quad \text{e} \quad \mathbf{K}(\boldsymbol{\tau}) = \sum_{i=1}^M \mathbf{K}_i(\boldsymbol{\tau}). \quad (1.10)$$

O elemento (r, s) da matriz $\mathbf{K}_i(\boldsymbol{\tau})$ é dado por

$$K_{i,rs}(\boldsymbol{\tau}) = \frac{b_{rsi}}{4} \left(\frac{4f_{gi}}{n_i(n_i+2)} - 1 \right) + \frac{2f_{gi}}{n_i(n_i+2)} \text{tr}\{\mathbf{V}_i^{-1} \dot{\mathbf{V}}_i(r) \mathbf{V}_i^{-1} \dot{\mathbf{V}}_i(s)\}$$

em que $d_{gi} = E\{W_g^2(U_i)U_i\}$, $f_{gi} = E\{W_g^2(U_i)U_i^2\}$ com $U_i = \|\mathbf{Z}_i\|^2$, $\mathbf{Z}_i \sim EC_{n_i}(\mathbf{0}, \mathbf{I}_{n_i}; g)$ e $b_{rsi} = \text{tr}\{\mathbf{V}_i^{-1} \dot{\mathbf{V}}_i(r)\} \text{tr}\{\mathbf{V}_i^{-1} \dot{\mathbf{V}}_i(s)\}$. É possível obter expressões em forma fechada das esperanças d_{gi} e f_{gi} para algumas distribuições de contornos elípticos.

Exemplos de d_{gi} e f_{gi} para algumas distribuições de contornos elípticos são:

- t de Student com $\nu > 0$ graus de liberdade, $t_{n_i}(\boldsymbol{\mu}_i, \mathbf{V}_i, \nu)$

$$d_{gi} = \frac{n_i}{4} \left(\frac{\nu + n_i}{\nu + n_i + 2} \right) \quad \text{e} \quad f_{gi} = \frac{n_i(n_i + 2)}{4} \left(\frac{\nu + n_i}{\nu + n_i + 2} \right),$$

para $\nu = 1$ obtemos d_{gi} e f_{gi} para a distribuição de Cauchy, $C_{n_i}(\boldsymbol{\mu}_i, \mathbf{V}_i)$.

- Exponencial potência, $PE_{n_i}(\boldsymbol{\mu}_i, \mathbf{V}_i, \nu)$ com parâmetro de forma $\nu > 0$. Temos que

$$d_{gi} = \frac{\nu^2}{2^{1/\nu}} \frac{\Gamma(\frac{n_i-2}{2\nu} + 2)}{\Gamma(\frac{n_i}{2\nu})} \quad \text{e} \quad f_{gi} = \frac{n_i(n_i + 2\nu)}{4},$$

obtemos d_{gi} e f_{gi} para a distribuição normal multivariada $N_{n_i}(\boldsymbol{\mu}_i, \mathbf{V}_i)$, fazendo $\nu = 1$.

Porém, para distribuições tais como a normal contaminada ou logísticas tipo I e II o cálculo das integrais d_{gi} e f_{gi} pode ser mais desafiante, sugerimos então aproximar tais integrais usando técnicas de Monte Carlo ou aproximações de Laplace.

Neste trabalho, estimamos as matrizes de dispersão para os estimadores de máxima verossimilhança $\hat{\beta}$ e $\hat{\tau}$ usando $\mathbf{K}^{-1}(\beta)$ e $\mathbf{K}^{-1}(\tau)$, respectivamente.

1.3 Modelo Linear Misto Hierárquico

Nesta seção descrevemos o uso de algoritmos tipo-EM para estimação por máxima verossimilhança no modelo linear misto usando distribuições na classe das misturas de escala normal.

O algoritmo EM (Dempster et al., 1977) é um método bastante popular para estimação por máxima verossimilhança em modelos com dados incompletos. Uma característica de interesse do algoritmo EM é que os dados não observáveis podem ser de natureza teórica, fato que tem sido explorado para a obtenção de algoritmos numericamente estáveis para estimação por máxima verossimilhança usando a classe das misturas de escala normal (Lange e Sinsheimer, 1993; Jamshidian, 1999).

Segundo Pinheiro et al. (2001) consideramos algoritmos tipo-EM para estimação de parâmetros no modelo linear misto hierárquico definido por (1.3)-(1.4). Primeiramente, descrevemos um algoritmo ECM (Meng e Rubin, 1993) considerando \mathbf{b}_i e v_i como não observáveis. Posteriormente, apresentamos um algoritmo ECM no modelo marginal obtido *a partir* da formulação hierárquica dada em (1.3) ou (1.4).

A densidade marginal de \mathbf{Y}_i no modelo linear misto hierárquico definido em (1.4) é dada por

$$f(\mathbf{y}_i) = \int \int f(\mathbf{y}_i | \mathbf{b}_i, v_i) f(\mathbf{b}_i | v_i) d\mathbf{b}_i dH(v_i),$$

em que $f(\mathbf{y}_i | \mathbf{b}_i, v_i)$ e $f(\mathbf{b}_i | v_i)$ denotam as densidades condicionais associadas a (1.4). Como

$$\begin{aligned} f(\mathbf{y}_i | v_i) &= \int f(\mathbf{y}_i | \mathbf{b}_i, v_i) f(\mathbf{b}_i | v_i) d\mathbf{b}_i \\ &= |2\pi\kappa(v_i)\mathbf{V}_i|^{-1/2} \exp\{-\frac{1}{2}\kappa^{-1}(v_i)u_i\} \end{aligned}$$

com $u_i = (\mathbf{Y}_i - \mathbf{X}_i\beta)^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\beta)$ e $\mathbf{V}_i = \mathbf{Z}_i \Psi \mathbf{Z}_i^T + \Sigma_i$,

a densidade marginal de \mathbf{Y}_i é

$$f(\mathbf{y}_i) = |2\pi\mathbf{V}_i|^{-1/2} \int_0^\infty \{\kappa(v_i)\}^{-n_i/2} \exp\{-\frac{1}{2} \kappa^{-1}(v_i)u_i\} dH(v_i). \quad (1.11)$$

Então $\mathbf{Y}_i \sim SMN_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i; H)$, $i = 1, \dots, M$. Este resultado é importante pois indica que a partir da formulação hierárquica em (1.4) obtemos que a distribuição marginal \mathbf{Y}_i também pertence à classe das misturas de escala normal. Além disso, este resultado permite em algum sentido, justificar a modelagem de modelos lineares com efeitos mistos usando (1.2). Por outro lado, usando (1.11) podemos expressar a densidade marginal de \mathbf{Y}_i empregando uma formulação hierárquica como

$$\begin{aligned} \mathbf{Y}_i | v_i &\stackrel{ind}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \kappa(v_i)\mathbf{V}_i) \quad e \\ v_i &\stackrel{ind}{\sim} H(v_i; \boldsymbol{\nu}), \quad i = 1, \dots, M. \end{aligned} \quad (1.12)$$

O logaritmo da função de verossimilhança de dados observados $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ para o modelo linear misto hierárquico é dado por $L(\boldsymbol{\theta}|\mathbf{Y}) = \sum_{i=1}^M L_i(\boldsymbol{\theta}|\mathbf{Y}_i)$, com

$$\begin{aligned} L_i(\boldsymbol{\theta}|\mathbf{Y}_i) &= \log \int_0^\infty \{\kappa(v_i)\}^{-n_i/2} \exp\{-\frac{1}{2} \kappa^{-1}(v_i)u_i\} dH(v_i) \\ &\quad - \frac{1}{2} \log |\mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i| + c. \end{aligned} \quad (1.13)$$

Notamos que a densidade (1.13) pode ser avaliada de forma relativamente simples, uma vez que somente é requerido resolver uma integral unidimensional. Fato que não é compartilhado por outros modelos lineares com efeitos mistos para dados não normais. Por exemplo, no caso de modelos lineares generalizados mistos (veja McCulloch e Searle, 2001. Cap. 8), a densidade marginal é o resultado de uma integração q -dimensional.

Por outro lado, também é possível utilizar procedimentos de aproximação tais como o método de Laplace (veja Kass, 1997) para resolver a integral em (1.13). Veremos na Seção 1.3.3 que para muitos membros da família de distribuições de mistura de escala normal a integral em (1.13) pode ser resolvida analiticamente.

Neste trabalho optamos por abordar a estimação de parâmetros no modelo linear misto hierárquico usando algoritmos tipo-EM. Nas subseções a seguir consideramos algoritmos ECM baseados nas equações (1.4) e (1.12).

1.3.1 Algoritmo ECM quando \mathbf{b}_i e v_i são não observáveis

Seja $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{b}^T, \mathbf{v}^T)^T$ o vetor de dados completos, em que $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ representa o vetor de respostas observadas para os M indivíduos, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_M^T)^T$ denotam os efeitos aleatórios e $\mathbf{v} = (v_1, \dots, v_M)^T$ correspondem às variáveis de mistura. Aqui consideramos que $(\mathbf{b}^T, \mathbf{v}^T)^T$ é não observável. O logaritmo da função de verossimilhança para os dados completos \mathbf{Y}_c é dado por $L(\boldsymbol{\theta}|\mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}|\mathbf{Y}_c)$, com

$$\begin{aligned} L_i(\boldsymbol{\theta}|\mathbf{Y}_c) &= -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{\kappa^{-1}(v_i)}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ &\quad - \frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c, \end{aligned} \quad (1.14)$$

em que $h(v_i; \boldsymbol{\nu})$ é a função de densidade da variável de mistura com vetor de parâmetros $\boldsymbol{\nu}$ e c denota uma constante.

Usando a representação hierárquica dada em (1.3) temos que a distribuição de \mathbf{b}_i condicional a (\mathbf{Y}_i, v_i) segue uma distribuição normal com parâmetros

$$\begin{aligned} \mathbb{E}(\mathbf{b}_i|\mathbf{Y}_i, v_i) &= \boldsymbol{\Psi} \mathbf{Z}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \quad \text{e} \\ \text{Cov}(\mathbf{b}_i|\mathbf{Y}_i, v_i) &= \kappa(v_i) (\boldsymbol{\Psi} - \boldsymbol{\Psi} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \boldsymbol{\Psi}). \end{aligned}$$

Deste modo, a esperança condicional do logaritmo da função de verossimilhança para dados completos, também chamada de função Q , para o modelo (1.4) é

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \mathbb{E}\{L(\boldsymbol{\theta}|\mathbf{Y}_c)|\mathbf{Y}, \hat{\boldsymbol{\theta}}\},$$

em que $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, com $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) + Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$, dadas por

$$\begin{aligned} Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) &= -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \hat{\boldsymbol{\Omega}}_i \mathbf{Z}_i^T \\ &\quad - \frac{1}{2} \hat{\kappa}_i (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\hat{\mathbf{b}}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\hat{\mathbf{b}}_i), \end{aligned} \quad (1.15)$$

$$Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}}) = -\frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{1}{2} \text{tr} \boldsymbol{\Psi}^{-1} (\hat{\kappa}_i \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \hat{\boldsymbol{\Omega}}_i), \quad (1.16)$$

em que

$$\hat{\kappa}_i = \mathbb{E}(\kappa^{-1}(v_i)|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}), \quad \hat{\mathbf{b}}_i = \mathbb{E}(\mathbf{b}_i|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}) \quad \text{e} \quad \hat{\boldsymbol{\Omega}}_i = \kappa^{-1}(v_i) \text{Cov}(\mathbf{b}_i|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}).$$

Deste modo, obtemos

$$\begin{aligned} \hat{\mathbf{b}}_i &= \hat{\boldsymbol{\Psi}} \mathbf{Z}_i^T (\mathbf{Z}_i \hat{\boldsymbol{\Psi}} \mathbf{Z}_i^T + \hat{\boldsymbol{\Sigma}}_i)^{-1} (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) \\ &= (\hat{\boldsymbol{\Psi}}^{-1} + \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) \quad \text{e} \end{aligned} \quad (1.17)$$

$$\hat{\boldsymbol{\Omega}}_i = \hat{\boldsymbol{\Psi}} - \hat{\boldsymbol{\Psi}} \mathbf{Z}_i^T (\mathbf{Z}_i \hat{\boldsymbol{\Psi}} \mathbf{Z}_i^T + \hat{\boldsymbol{\Sigma}}_i)^{-1} \mathbf{Z}_i \hat{\boldsymbol{\Psi}} = (\hat{\boldsymbol{\Psi}}^{-1} + \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i)^{-1}. \quad (1.18)$$

Tipicamente as esperanças condicionais $\hat{\kappa}_i$ dependem das distâncias u_i . Note que

$$\begin{aligned} u_i(\hat{\boldsymbol{\theta}}) &= (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T (\mathbf{Z}_i \hat{\boldsymbol{\Psi}} \mathbf{Z}_i^T + \hat{\boldsymbol{\Sigma}}_i)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \\ &= \hat{\mathbf{b}}_i^T \hat{\boldsymbol{\Psi}}^{-1} \hat{\mathbf{b}}_i + (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i)^T \hat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i). \end{aligned} \quad (1.19)$$

Temos então que a distância $u_i(\hat{\boldsymbol{\theta}})$ pode ser decomposta em duas outras distâncias devidas aos efeitos aleatórios estimados $\hat{\mathbf{b}}_i$ e aos resíduos $\mathbf{e}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i$, respectivamente.

Considere o seguinte algoritmo ECM, que maximiza o logaritmo da função de verossimilhança iterativamente mediante as etapas resumidas a seguir:

Passo E: a partir de estimativas iniciais para $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e usando (1.17) e (1.18), calcular $\hat{\mathbf{b}}_i$, $\hat{\boldsymbol{\Omega}}_i$ e $\hat{\kappa}_i$ para $i = 1, \dots, M$.

Passo CM: maximização da esperança do logaritmo da função de verossimilhança de dados completos, $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$.

Passo CM-1: fixando $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$ atualizar $\hat{\boldsymbol{\beta}}$ como:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^M \hat{\kappa}_i \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right)^{-1} \sum_{i=1}^M \hat{\kappa}_i \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\alpha}) (\mathbf{Y}_i - \mathbf{Z}_i \hat{\mathbf{b}}_i).$$

Passo CM-2: fixar $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ e atualizar $\hat{\boldsymbol{\alpha}}$ mediante maximizar $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com respeito a $\boldsymbol{\alpha}$

$$\hat{\boldsymbol{\alpha}} = \arg \max_{\boldsymbol{\alpha}} \{Q_1(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}), \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})\}.$$

Passo CM-3: atualizar $\hat{\boldsymbol{\lambda}}$ mediante maximizar $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com respeito a $\boldsymbol{\lambda}$

$$\hat{\boldsymbol{\lambda}} = \arg \max_{\boldsymbol{\lambda}} \{Q_2(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})\},$$

em que $Q_1(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})$ e $Q_2(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$ com $Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$ dadas em (1.15) e (1.16).

As etapas CM-2 e CM-3 deste algoritmo ECM devem ser resolvidas usando métodos iterativos. Com este objetivo é possível considerar, por exemplo, procedimentos quase-Newton.

1.3.2 Algoritmo ECM quando os b_i são integrados

Integrando os efeitos aleatórios em (1.4) ou, equivalentemente, considerando a formulação hierárquica dada em (1.12), podemos abordar a estimação de parâmetros no modelo marginal para \mathbf{Y}_i via um algoritmo tipo-EM. Neste caso, o vector de dados completos é $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{v}^T)^T$ em que as variáveis de mistura $\mathbf{v} = (v_1, \dots, v_M)^T$ são consideradas não observáveis.

O logaritmo da função de verossimilhança para os dados completos \mathbf{Y}_c assume a forma $L(\boldsymbol{\theta}|\mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}|\mathbf{Y}_c)$, com

$$L_i(\boldsymbol{\theta}|\mathbf{Y}_c) = -\frac{1}{2} \log |\mathbf{V}_i| - \frac{\kappa^{-1}(v_i)}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) + \log h(v_i; \boldsymbol{\nu}) + c, \quad (1.20)$$

em que, como anteriormente, $h(v_i; \boldsymbol{\nu})$ é a função de densidade da variável de mistura, c representa uma constante e $\mathbf{V}_i = \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$.

A Q -função no modelo definido por (1.12) assume a forma $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com

$$Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = -\frac{1}{2} \log |\mathbf{V}_i| - \frac{1}{2} \hat{\kappa}_i (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}), \quad (1.21)$$

em que $\hat{\kappa}_i = E(\kappa^{-1}(v_i)|\mathbf{Y}_i, \hat{\boldsymbol{\theta}})$, para $i = 1, \dots, M$. Desde que

$$\hat{\mathbf{b}}_i = E(\mathbf{b}_i|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Psi}}\mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}), \quad i = 1, \dots, M,$$

podemos calcular eficientemente as distâncias $u_i(\hat{\boldsymbol{\theta}})$ mediante (1.19) que por sua vez permite avaliar os pesos $\hat{\kappa}_i$ de forma simples.

Deste modo, um algoritmo ECM para estimação de parâmetros no modelo definido por (1.12) é dado pelas duas etapas resumidas a seguir:

Passo E: usando estimativas iniciais para $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, obter $\hat{\kappa}_i$, $i = 1, \dots, M$.

Passo CM: atualizar $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\lambda}}^T)^T$ mediante maximizar $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com respeito a $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ e $\boldsymbol{\lambda}$

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\}.$$

Observamos de (1.21) que a etapa CM do algoritmo ECM dado acima é equivalente à estimação por máxima verossimilhança no modelo linear misto sob normalidade com a inclusão de pesos $\hat{\kappa}_i$ para cada indivíduo. Pinheiro et al. (2001) notam que este algoritmo ECM pode ser implementado usando as facilidades de softwares tal como o procedimento Proc MIXED do SAS ou a função para S-PLUS `lme` da biblioteca `nlme`.

1.3.3 Esperanças condicionais para os algoritmos tipo-EM

No passo E dos algoritmos ECM considerados nas Seções 1.3.1 e 1.3.2 requer-se o cálculo dos pesos $\hat{\kappa}_i = E(\kappa^{-1}(v_i)|\mathbf{Y}_i, \hat{\boldsymbol{\theta}})$. A seguir apresentamos fórmulas de $\hat{\kappa}_i$ para algumas distribuições de mistura de escala normal freqüentemente utilizadas.

Usando (1.5), a forma geral para a esperança condicional, $E(\kappa^{-1}(v)|\mathbf{Y})$ é dada por

$$E(\kappa^{-1}(v)|\mathbf{Y}) = \frac{\int_0^\infty \{\kappa(v)\}^{-(n/2+1)} \exp\{-\kappa^{-1}(v)u/2\} dH(v)}{\int_0^\infty \{\kappa(v)\}^{-n/2} \exp\{-\kappa^{-1}(v)u/2\} dH(v)}. \quad (1.22)$$

Esta esperança da distribuição a posteriori da função $\kappa^{-1}(v)$ em (1.22) pode ser avaliada usando, por exemplo, o método de Laplace (veja Kass, 1997). Contudo, para muitas distribuições na classe das misturas de escala normal é possível obter expressões em forma fechada para (1.22).

Distribuição t de Student generalizada. A distribuição t de Student generalizada, $Gt_n(\boldsymbol{\mu}, \mathbf{V}, \nu, \gamma)$, $\nu, \gamma > 0$ pertence à classe das misturas de escala normal com $\kappa(v) = 1/v$ e $V \sim Gama(\nu/2, \gamma/2)$, em que $H(v)$ tem densidade

$$h(v; \boldsymbol{\nu}) = \frac{(\gamma/2)^{\nu/2} v^{\nu/2-1}}{\Gamma(\nu/2)} \exp(-\frac{1}{2}\gamma v), \quad \boldsymbol{\nu} = (\nu, \gamma)^T,$$

deste modo a esperança condicional em (1.22) fica dada por

$$E(\kappa^{-1}(v)|\mathbf{Y}) = \frac{\nu + n}{\gamma + u}.$$

Note que a distribuição t de Student é recuperada quando $\gamma = \nu$. Pinheiro et al. (2001) utilizam a distribuição t de Student para modelagem robusta no modelo linear com efeitos mistos atribuindo pesos às observações mediante a esperança condicional $E(\kappa^{-1}(v)|\mathbf{Y}) = (\nu + n)/(\nu + u)$.

Distribuição Slash. Para a distribuição slash (Rogers e Tukey, 1972; Lange e Sinsheimer, 1993), temos que $\kappa(v) = 1/v$ e a variável de mistura tem densidade

$$h(v; \nu) = \nu v^{\nu-1}, \quad 0 < v \leq 1 \text{ e } \nu > 0.$$

Considere a seguinte função gama incompleta:

$$G(a, b) = \int_0^1 r^{a-1} e^{-br} dr.$$

Logo, a esperança $E(\kappa^{-1}(v)|\mathbf{Y})$ para a distribuição slash fica dada por

$$E(\kappa^{-1}(v)|\mathbf{Y}) = \frac{G(n/2 + \nu + 1, u/2)}{G(n/2 + \nu, u/2)}.$$

Seja $P_x(a, b)$ a função de distribuição de uma variável aleatória $Gama(a, b)$, isto é

$$P_x(a, b) = \frac{b^a}{\Gamma(a)} \int_0^x r^{a-1} e^{-br} dr,$$

deste modo é possível avaliar a esperança condicional em (1.22) usando as facilidades dos pacotes estatísticos (veja Jamshidian, 1999) como

$$E(\kappa^{-1}(v)|\mathbf{Y}) = \left(\frac{n + 2\nu}{u} \right) \frac{P_1(n/2 + \nu + 1, u/2)}{P_1(n/2 + \nu, u/2)}.$$

Distribuição Normal Contaminada. No caso da distribuição normal contaminada multivariada, $CN_n(\boldsymbol{\mu}, \mathbf{V}, \delta, \gamma)$, $0 \leq \delta \leq 1$ e $0 < \gamma < 1$ (Little, 1988), temos que $\kappa(v) = 1/v$ e a variável de mistura discreta V tem função de probabilidade

$$h(v; \boldsymbol{\nu}) = \begin{cases} \delta, & \text{se } v = \gamma, \\ 1 - \delta, & \text{se } v = 1, \end{cases}$$

em que $\boldsymbol{\nu} = (\delta, \gamma)^T$. A esperança em (1.22) para esta distribuição resulta em

$$E(\kappa^{-1}(v)|\mathbf{Y}) = \frac{1 - \delta + \delta\gamma^{n/2+1}e^{(1-\gamma)u/2}}{1 - \delta + \delta\gamma^{n/2}e^{(1-\gamma)u/2}}.$$

Distribuição Exponencial Potência. É possível caracterizar a distribuição exponencial potência na classe de misturas de escala normal (West, 1987; Lange e Sinsheimer, 1993) para $0 < \nu \leq 1$, em que a variável de mistura corresponde a uma distribuição estável positiva, $S^P(a)$ de índice a , $0 < a < 1$ (veja West, 1987 e Chen e Dey, 1998). Deste modo, em lugar de avaliar (1.22) adotamos o enfoque sugerido por Lange e Sinsheimer (1993, Sec. 3).

Escrevemos a densidade exponencial potência como

$$|2\pi\mathbf{V}|^{-1/2} \exp\{-\eta(u)/2\},$$

em que $\eta(u) = u^\nu + n \log 2\pi c_n$ e c_n é a constante de padronização. Logo $\eta'(u)$ corresponde a $E(\kappa^{-1}(u)|\mathbf{Y})$, obtendo-se

$$\eta'(u) = \nu u^{\nu-1}, \quad \text{em que } u \neq 0 \text{ e } \nu \neq \frac{1}{2}.$$

Estimação do Erro Padrão das Estimativas de Parâmetros

A partir da equação (1.11) e usando a representação das misturas de escala normal dada em Lange e Sinsheimer (1993) podemos considerar

$$t(u_i) = -2 \log \int_0^\infty \{\kappa(v_i)\}^{-n_i/2} \exp\{-\frac{\kappa^{-1}(v_i)}{2}u_i\} dH(v_i).$$

Agora, analogamente ao caso marginal definimos $W_t(u) = -\frac{1}{2}t'(u)$ e fazemos $d_{ti} = E\{W_t^2(U_i)U_i\}$, $f_{ti} = E\{W_t^2(U_i)U_i^2\}$ com $U_i = \|\mathbf{Z}_i\|^2$, $\mathbf{Z}_i \sim EC_{n_i}(\mathbf{0}, \mathbf{I}_{n_i}; g^*)$ em que $g^*(u) = (2\pi)^{-n_i/2} \exp(-\frac{1}{2}t(u))$. Desse modo podemos utilizar a expressão da matriz de informação de Fisher dada em (1.10) para obter uma estimação da matriz de dispersão para o estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}$.

1.4 Considerações Computacionais

Nesta seção revisamos alguns elementos computacionais relativos aos algoritmos tipo-EM considerados previamente. O principal interesse é apresentar formas eficientes para a estimação de parâmetros no modelo linear misto hierárquico. Este objetivo é alcançado mediante utilizar decomposições matriciais. Nossa exposição segue algumas idéias e notação usadas em Bates e Pinheiro (1998) e Bates e DebRoy (2004).

Para eficiência computacional consideramos o modelo hierárquico dado em (1.3)-(1.4) tal que $\mathbf{Y}_i | \mathbf{b}_i, v_i \sim N_{n_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \kappa(v_i) \phi \mathbf{I})$ e fatoramos $\phi > 0$ como $\boldsymbol{\Psi} = \phi \mathbf{D}$. A seguir decomponemos a matriz de precisão $\mathbf{D}^{-1}(\boldsymbol{\lambda})$ usando um fator raiz-quadrada, como

$$\mathbf{D}^{-1} = \boldsymbol{\Delta}^T \boldsymbol{\Delta}. \quad (1.23)$$

O fator $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\lambda})$ pode ser obtido usando, por exemplo, a decomposição Cholesky (Golub e van Loan, 1996. Cap. 4). Notamos que no modelo hierárquico (1.3)-(1.4) a matriz $\boldsymbol{\Psi}(\boldsymbol{\lambda})$ deve ser positiva definida. De fato a decomposição em (1.23) permite estimar $\boldsymbol{\lambda}$ assegurando tal restrição. Usando as parametrizações discutidas em Pinheiro e Bates (1996) podemos expressar $\boldsymbol{\Delta}$ como função de $\boldsymbol{\lambda}$ e deste modo simplificar o procedimento de otimização.

Note que, quando os efeitos aleatórios \mathbf{b}_i e as variáveis de mistura v_i são considerados não observáveis, o logaritmo da função de verossimilhança de dados completos para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi, \boldsymbol{\lambda}^T)^T$, resulta em $L(\boldsymbol{\theta} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta} | \mathbf{Y}_c)$,

em que

$$L_i(\boldsymbol{\theta}|\mathbf{Y}_c) = -\frac{n_i+q}{2} \log \phi - \frac{\kappa^{-1}(v_i)}{2\phi} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^T (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ - \frac{1}{2} \log |\mathbf{D}| - \frac{\kappa^{-1}(v_i)}{2\phi} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c,$$

agora, como $|\mathbf{D}|^{-1} = |\boldsymbol{\Delta}|^2$, temos que

$$L_i(\boldsymbol{\theta}|\mathbf{Y}_c) = -\frac{n_i+q}{2} \log \phi - \frac{\kappa^{-1}(v_i)}{2\phi} \|\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 \\ + \log |\boldsymbol{\Delta}| - \frac{\kappa^{-1}(v_i)}{2\phi} \|\boldsymbol{\Delta}\mathbf{b}_i\|^2 + \log h(v_i; \boldsymbol{\nu}) + c.$$

Deste modo $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, em que $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = Q_{1i}(\boldsymbol{\beta}, \phi|\hat{\boldsymbol{\theta}}) + Q_{2i}(\boldsymbol{\lambda}, \phi|\hat{\boldsymbol{\theta}})$, com

$$Q_{1i}(\boldsymbol{\beta}, \phi|\hat{\boldsymbol{\theta}}) = -\frac{n_i+q}{2} \log \phi - \frac{1}{2\phi} \text{tr} \mathbf{Z}_i \hat{\boldsymbol{\Omega}}_i \mathbf{Z}_i^T - \frac{1}{2\phi} \hat{\kappa}_i \|\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\hat{\mathbf{b}}_i\|^2 \text{ e} \\ Q_{2i}(\boldsymbol{\lambda}, \phi|\hat{\boldsymbol{\theta}}) = \log |\boldsymbol{\Delta}| - \frac{1}{2\phi} \text{tr} \boldsymbol{\Delta} (\hat{\kappa}_i \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \hat{\boldsymbol{\Omega}}_i) \boldsymbol{\Delta}^T.$$

Analogamente a (1.17) e (1.18), temos que

$$\hat{\mathbf{b}}_i = (\hat{\mathbf{D}}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) \\ = (\hat{\boldsymbol{\Delta}}^T \hat{\boldsymbol{\Delta}} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})$$

e

$$\hat{\boldsymbol{\Omega}}_i = \hat{\phi} (\hat{\mathbf{D}}^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} = \hat{\phi} (\hat{\boldsymbol{\Delta}}^T \hat{\boldsymbol{\Delta}} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1}.$$

Podemos tornar eficiente o procedimento de estimação mediante avaliar as expressões anteriores usando uma decomposição ortogonal-triangular, frequentemente conhecida como decomposição QR (Golub e van Loan, 1996. Cap. 5), da forma

$$(\mathbf{Z}_i, \mathbf{X}_i) = \mathbf{Q}_{(i)} \begin{pmatrix} \mathbf{R}_{11(i)} & \mathbf{R}_{12(i)} \\ \mathbf{0} & \mathbf{R}_{22(i)} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (1.24)$$

para $n_i \geq p + q$, em que $\mathbf{Q}_{(i)}$ é matriz ortogonal $n_i \times n_i$, $\mathbf{R}_{11(i)}$ e $\mathbf{R}_{22(i)}$ são matrizes triangulares superiores $q \times q$ e $p \times p$, respectivamente e $\mathbf{R}_{12(i)}$ é matriz $q \times p$. Em (1.24) também podemos supor que $\text{posto}(\mathbf{Z}_i) = q$, em cujo caso $\mathbf{R}_{11(i)}$ é matriz não-singular. Aplicando o fator ortogonal a \mathbf{Y}_i obtemos

$$\mathbf{Q}_{(i)}^T \mathbf{Y}_i = (\mathbf{c}_{1(i)}^T, \mathbf{c}_{2(i)}^T, \mathbf{c}_{3(i)}^T)^T.$$

A decomposição anterior pode ser agrupada na matriz aumentada

$$(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Y}_i) = \mathbf{Q}_{(i)} \begin{pmatrix} \mathbf{R}_{11(i)} & \mathbf{R}_{12(i)} & \mathbf{c}_{1(i)} \\ \mathbf{0} & \mathbf{R}_{22(i)} & \mathbf{c}_{2(i)} \\ \mathbf{0} & \mathbf{0} & \mathbf{c}_{3(i)} \end{pmatrix}. \quad (1.25)$$

Para introduzir idéias, considere decompor a matriz de produtos cruzados $(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Y}_i)^T (\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Y}_i)$ como

$$\begin{pmatrix} \mathbf{Z}_i^T \mathbf{Z}_i & \mathbf{Z}_i^T \mathbf{X}_i & \mathbf{Z}_i^T \mathbf{Y}_i \\ \mathbf{X}_i^T \mathbf{Z}_i & \mathbf{X}_i^T \mathbf{X}_i & \mathbf{X}_i^T \mathbf{Y}_i \\ \mathbf{Y}_i^T \mathbf{Z}_i & \mathbf{Y}_i^T \mathbf{X}_i & \mathbf{Y}_i^T \mathbf{Y}_i \end{pmatrix} = \mathbf{G}_{(i)}^T \mathbf{G}_{(i)},$$

em que $\mathbf{G}_{(i)}$ corresponde ao fator triangular dado em (1.25). Logo, podemos notar que o fator Cholesky $\mathbf{T}_{(i)}$ da seguinte decomposição:

$$\begin{pmatrix} \mathbf{Z}_i^T \mathbf{Z}_i + \mathbf{D}^{-1} & \mathbf{Z}_i^T \mathbf{X}_i & \mathbf{Z}_i^T \mathbf{Y}_i \\ \mathbf{X}_i^T \mathbf{Z}_i & \mathbf{X}_i^T \mathbf{X}_i & \mathbf{X}_i^T \mathbf{Y}_i \\ \mathbf{Y}_i^T \mathbf{Z}_i & \mathbf{Y}_i^T \mathbf{X}_i & \mathbf{Y}_i^T \mathbf{Y}_i \end{pmatrix} = \mathbf{T}_{(i)}^T \mathbf{T}_{(i)}$$

é equivalente ao fator triangular da decomposição ortogonal-triangular,

$$\begin{aligned} \begin{pmatrix} \Delta(\lambda) & \mathbf{0} & \mathbf{0} \\ \mathbf{R}_{11(i)} & \mathbf{R}_{12(i)} & \mathbf{c}_{1(i)} \\ \mathbf{0} & \mathbf{R}_{22(i)} & \mathbf{c}_{2(i)} \\ \mathbf{0} & \mathbf{0} & \mathbf{c}_{3(i)} \end{pmatrix} &= \mathbf{S}_{(i)} \begin{pmatrix} \mathbf{T}_{11(i)} & \mathbf{T}_{12(i)} & \mathbf{t}_{13(i)} \\ \mathbf{0} & \mathbf{T}_{22(i)} & \mathbf{t}_{23(i)} \\ \mathbf{0} & \mathbf{0} & \mathbf{t}_{33(i)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= \mathbf{S}_{(i)} \begin{pmatrix} \mathbf{T}_{(i)} \\ \mathbf{0} \end{pmatrix}. \end{aligned} \quad (1.26)$$

Portanto, temos que

$$\begin{aligned} \|\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i\|^2 + \|\Delta \mathbf{b}_i\|^2 &= \mathbf{a}^T \mathbf{T}_{(i)}^T \mathbf{T}_{(i)} \mathbf{a} = \|\mathbf{T}_{(i)} \mathbf{a}\|^2 \\ &= \|\mathbf{t}_{13(i)} - \mathbf{T}_{12(i)} \boldsymbol{\beta} - \mathbf{T}_{11(i)} \mathbf{b}_i\|^2 + \|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)} \boldsymbol{\beta}\|^2 + \mathbf{t}_{33(i)}^2 \end{aligned}$$

para $\mathbf{a} = (-\mathbf{b}_i^T, -\boldsymbol{\beta}^T, 1)^T$.

A partir de estimativas iniciais para $\boldsymbol{\theta}$, e usando a decomposição em (1.26) podemos avaliar as distâncias $u_i(\hat{\boldsymbol{\theta}})$ dadas em (1.19), como

$$\begin{aligned} u_i(\hat{\boldsymbol{\theta}}) &= \frac{1}{\hat{\phi}} \{ \|\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i\|^2 + \|\hat{\Delta} \hat{\mathbf{b}}_i\|^2 \} \\ &= \frac{1}{\hat{\phi}} \{ \|\mathbf{t}_{13(i)} - \mathbf{T}_{12(i)} \hat{\boldsymbol{\beta}} - \mathbf{T}_{11(i)} \hat{\mathbf{b}}_i\|^2 + \|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)} \hat{\boldsymbol{\beta}}\|^2 + \mathbf{t}_{33(i)}^2 \}. \end{aligned} \quad (1.27)$$

Por outro lado, temos que

$$\begin{aligned}\widehat{\mathbf{b}}_i &= (\widehat{\Delta}^T \widehat{\Delta} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T (\mathbf{Y}_i - \mathbf{X}_i \widehat{\beta}) \\ &= (\mathbf{T}_{11(i)}^T \mathbf{T}_{11(i)})^{-1} \mathbf{T}_{11(i)}^T (\mathbf{t}_{13(i)} - \mathbf{T}_{12(i)} \widehat{\beta}) \\ &= \mathbf{T}_{11(i)}^{-1} (\mathbf{t}_{13(i)} - \mathbf{T}_{12(i)} \widehat{\beta}),\end{aligned}$$

ou seja, $\widehat{\mathbf{b}}_i = \mathbb{E}(\mathbf{b}_i | \mathbf{Y}_i, \widehat{\theta})$ pode ser obtido eficientemente como a solução do sistema triangular

$$\mathbf{T}_{11(i)} \widehat{\mathbf{b}}_i = \mathbf{t}_{13(i)} - \mathbf{T}_{12(i)} \widehat{\beta}, \quad (1.28)$$

também temos que

$$\widehat{\Omega}_i = \widehat{\phi} (\widehat{\Delta}^T \widehat{\Delta} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} = \widehat{\phi} \mathbf{T}_{11(i)}^{-1} \mathbf{T}_{11(i)}^{-T}. \quad (1.29)$$

Seja

$$\widetilde{\mathbf{R}}_{11(i)} = \begin{pmatrix} \mathbf{R}_{11(i)} \\ \mathbf{0} \end{pmatrix}, \quad \widetilde{\mathbf{R}}_{10(i)} = \begin{pmatrix} \mathbf{R}_{12(i)} \\ \mathbf{R}_{22(i)} \end{pmatrix}$$

e $\widetilde{\mathbf{c}}_{1(i)} = (\mathbf{c}_{1(i)}^T, \mathbf{c}_{2(i)}^T)^T$, $\widetilde{\mathbf{c}}_{0(i)} = \mathbf{c}_{3(i)}$. Deste modo a decomposição em (1.25), resulta em

$$(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Y}_i) = \mathbf{Q}_{(i)} \begin{pmatrix} \widetilde{\mathbf{R}}_{11(i)} & \widetilde{\mathbf{R}}_{10(i)} & \widetilde{\mathbf{c}}_{1(i)} \\ \mathbf{0} & \mathbf{0} & \widetilde{\mathbf{c}}_{0(i)} \end{pmatrix}.$$

Defina a resposta de trabalho como $\mathbf{w}_{(i)} = \widetilde{\mathbf{c}}_{1(i)} - \widetilde{\mathbf{R}}_{11(i)} \widehat{\mathbf{b}}_i$, para $i = 1, \dots, M$. Note que a estimativa $\widehat{\beta}$ da etapa CM-1 do algoritmo ECM dado na Seção 1.3.1 é equivalente à solução do sistema triangular

$$\mathbf{R}_{00} \widehat{\beta} = \mathbf{c}_0, \quad (1.30)$$

em que \mathbf{R}_{00} e \mathbf{c}_0 são obtidos a partir da decomposição QR,

$$\begin{pmatrix} \sqrt{\widehat{\kappa}_1} \widetilde{\mathbf{R}}_{10(1)} & \sqrt{\widehat{\kappa}_1} \mathbf{w}_{(1)} \\ \vdots & \vdots \\ \sqrt{\widehat{\kappa}_M} \widetilde{\mathbf{R}}_{10(M)} & \sqrt{\widehat{\kappa}_M} \mathbf{w}_{(M)} \end{pmatrix} = \mathbf{Q}_0 \begin{pmatrix} \mathbf{R}_{00} & \mathbf{c}_0 \\ \mathbf{0} & \mathbf{c}_{-1} \end{pmatrix}. \quad (1.31)$$

Para $\beta = \widehat{\beta}$ e $\lambda = \widehat{\lambda}$ fixados, temos que o estimador de ϕ é dado por

$$\widehat{\phi} = \frac{1}{N + Mq} \sum_{i=1}^M \{ \text{tr}(\mathbf{Z}_i \widehat{\Omega}_i \mathbf{Z}_i^T + \widehat{\Delta} \widehat{\Omega}_i \widehat{\Delta}^T) + \widehat{\kappa}_i (\|\mathbf{Y}_i - \mathbf{X}_i \widehat{\beta} - \mathbf{Z}_i \widehat{\mathbf{b}}_i\|^2 + \|\widehat{\Delta} \widehat{\mathbf{b}}_i\|^2) \},$$

em que $N = \sum_{i=1}^M n_i$ denota o número total de observações.

Na etapa CM-3 do algoritmo ECM da Seção 1.3.1 é requerido maximizar $Q_2(\boldsymbol{\lambda}, \phi | \hat{\boldsymbol{\theta}})$ com respeito a $\boldsymbol{\lambda}$ para $\phi = \hat{\phi}$ fixado. Temos que

$$Q_2(\boldsymbol{\lambda}, \hat{\phi} | \hat{\boldsymbol{\theta}}) = M \log |\boldsymbol{\Delta}| - \frac{1}{2\hat{\phi}} \text{tr} \boldsymbol{\Delta} \{ \sum_{i=1}^M (\hat{\kappa}_i \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \hat{\boldsymbol{\Omega}}_i) \} \boldsymbol{\Delta}^T$$

com $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\lambda})$. A soma $\sum_{i=1}^M (\hat{\kappa}_i \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \hat{\boldsymbol{\Omega}}_i)$ é fixa na etapa CM-3 neste algoritmo ECM e pode ser avaliada de forma eficiente usando a seguinte decomposição ortogonal-triangular:

$$\begin{pmatrix} \sqrt{\hat{\kappa}_1} \hat{\mathbf{b}}_1^T \\ \sqrt{\hat{\phi}} \mathbf{T}_{11(1)}^{-T} \\ \vdots \\ \sqrt{\hat{\kappa}_M} \hat{\mathbf{b}}_M^T \\ \sqrt{\hat{\phi}} \mathbf{T}_{11(M)}^{-T} \end{pmatrix} = \mathbf{Q}_b \begin{pmatrix} \mathbf{R}_b \\ \mathbf{0} \end{pmatrix}$$

com \mathbf{Q}_b matriz ortogonal de orden $M(q+1)$ e \mathbf{R}_b matriz triangular superior $q \times q$. Deste modo obtemos,

$$Q_2(\boldsymbol{\lambda}, \hat{\phi} | \hat{\boldsymbol{\theta}}) = M \log |\boldsymbol{\Delta}| - \frac{1}{2\hat{\phi}} \text{tr} \boldsymbol{\Delta} \mathbf{R}_b^T \mathbf{R}_b \boldsymbol{\Delta}^T.$$

Sugerimos otimizar $Q_2(\boldsymbol{\lambda}, \hat{\phi} | \hat{\boldsymbol{\theta}})$ usando um método secante multivariado (Dennis e Schnabel, 1996. Cap. 9). Este algoritmo $\min_{\boldsymbol{\lambda}} \{-Q_2(\boldsymbol{\lambda}, \hat{\phi} | \hat{\boldsymbol{\theta}})\}$ mediante a iteração

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \delta_k \mathbf{p}_k, \quad k = 0, 1, \dots, \quad (1.32)$$

em que \mathbf{p}_k representa uma direção de busca, definida como

$$\mathbf{p}_k = -\mathbf{H}_k \dot{\mathbf{Q}}_2(\boldsymbol{\lambda}^{(k)})$$

com \mathbf{H}_k sendo uma aproximação da inversa da matriz hessiana $\ddot{\mathbf{Q}}_2(\boldsymbol{\lambda})$ dada pela atualização BFGS, $\dot{\mathbf{Q}}_2(\boldsymbol{\lambda}') = \partial Q_2(\boldsymbol{\lambda}, \hat{\phi} | \hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\lambda} |_{\boldsymbol{\lambda}=\boldsymbol{\lambda}'}$ e δ_k denota o tamanho do passo. O vetor gradiente $\dot{\mathbf{Q}}_2(\boldsymbol{\lambda})$ requerido para o método tem elementos dados por

$$\dot{\mathbf{Q}}_2(j) = M \text{tr} \boldsymbol{\Delta}^{-1} \dot{\boldsymbol{\Delta}}(j) - \frac{1}{2\hat{\phi}} \text{tr} \dot{\boldsymbol{\Delta}}(j) \mathbf{R}_b^T \mathbf{R}_b \boldsymbol{\Delta}^T - \frac{1}{2\hat{\phi}} \text{tr} \boldsymbol{\Delta} \mathbf{R}_b^T \mathbf{R}_b \dot{\boldsymbol{\Delta}}^T(j),$$

em que $\dot{\boldsymbol{\Delta}}(j) = \partial \boldsymbol{\Delta}(\boldsymbol{\lambda}) / \partial \lambda_j$, para $j = 1, \dots, d$.

As equações (1.27) a (1.29) apresentam maneiras extremamente eficientes para calcular a etapa E para os algoritmos ECM das Seções 1.3.1 e 1.3.2. De fato, a etapa E apresenta-se de baixo custo computacional e, deste modo, podemos executar a etapa E antes de *cada* etapa CM obtendo um algoritmo ECM multiciclo (Meng e Rubin, 1993).

A seguir apresentamos a $(r+1)$ -ésima etapa da versão multiciclo do algoritmo ECM dado na Seção 1.3.1, usando os métodos computacionais descritos acima.

Algoritmo 1

Aplicar a decomposição (1.25) a $(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Y}_i)$ para $i = 1, \dots, M$.

Passo E: usando estimativas da r -ésima etapa, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(r)}$ calcular $\hat{\mathbf{b}}_i, \hat{\Omega}_i$ usando (1.28) e (1.29), obter $u_i(\hat{\boldsymbol{\theta}})$ como

$$u_i(\hat{\boldsymbol{\theta}}) = \frac{1}{\phi} \{ \|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)}\hat{\boldsymbol{\beta}}\|^2 + t_{33(i)}^2 \}$$

e avaliar $\hat{\kappa}_i = E(\kappa^{-1}(v_i) | \mathbf{Y}_i, \hat{\boldsymbol{\theta}})$ usando as expressões desenvolvidas na Seção 1.3.3 para alguma mistura de escala normal particular.

Passo CM multiciclo: maximização da esperança do logaritmo da função de verossimilhança de dados completos, $Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})$ usando uma seqüência de passos CM.

Passo CM-1: atualizar $\boldsymbol{\beta}^{(r+1)}$ como a solução do sistema

$$\mathbf{R}_{00}\boldsymbol{\beta}^{(r+1)} = \mathbf{c}_0,$$

em que \mathbf{R}_{00} e \mathbf{c}_0 são obtidos da decomposição em (1.31).

Passo E: executar o passo E com $\hat{\boldsymbol{\theta}} = (\boldsymbol{\beta}^{(r+1)T}, \phi^{(r)}, \boldsymbol{\lambda}^{(r)T})^T$.

Passo CM-2: fixar $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(r+1)}$, $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{(r)}$ e atualizar $\phi^{(r+1)}$ como

$$\phi^{(r+1)} = \frac{1}{N + Mq} \left\{ Mq \phi^{(r)} + \sum_{i=1}^M \hat{\kappa}_i (\|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)}\boldsymbol{\beta}^{(r+1)}\|^2 + t_{33(i)}^2) \right\}.$$

Passo E: executar o passo E com $\hat{\boldsymbol{\theta}} = (\boldsymbol{\beta}^{(r+1)T}, \phi^{(r+1)}, \boldsymbol{\lambda}^{(r)T})^T$.

Passo CM-3: fixar $\hat{\phi} = \phi^{(r+1)}$ e atualizar $\boldsymbol{\lambda}^{(r+1)}$ mediante

$$\boldsymbol{\lambda}^{(r+1)} = \arg \min_{\boldsymbol{\lambda}} \{-Q_2(\boldsymbol{\lambda}, \phi^{(r+1)} | \hat{\boldsymbol{\theta}})\}. \quad (1.33)$$

Resolvemos o subproblema (1.33) usando (1.32). Inicializamos o algoritmo dado em (1.32) por meio de $\boldsymbol{\lambda}^{(r)}$. Seja $\boldsymbol{\lambda}^*$ o valor de convergência da seqüência definida por (1.32), deste modo tomamos $\boldsymbol{\lambda}^{(r+1)} = \boldsymbol{\lambda}^*$.

Então, fazer $\boldsymbol{\theta}^{(r)} = (\boldsymbol{\beta}^{(r+1)T}, \phi^{(r+1)}, \boldsymbol{\lambda}^{(r+1)T})^T$ e voltar à etapa E. O algoritmo ECM itera entre os passos E e CM até que a seqüência $\boldsymbol{\theta}^{(r)}$ atinja a convergência.

A seguir integramos os efeitos aleatórios a fim de obter a densidade marginal $f(\mathbf{y}_i)$. O objetivo agora é expressar esta densidade marginal usando uma formulação hierárquica como em (1.12), permitindo assim apresentar um enfoque eficiente para estimação de parâmetros usando uma versão multiciclo do algoritmo ECM considerado na Seção 1.3.2. A partir dos procedimentos computacionais discutidos anteriormente temos que

$$\begin{aligned} f(\mathbf{y}_i | v_i) &= \frac{|\Delta|}{\{2\pi\phi\kappa(v_i)\}^{-n_i/2}} \exp\left\{-\frac{\kappa^{-1}(v_i)}{2\phi} (\|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)}\boldsymbol{\beta}\|^2 + t_{33(i)}^2)\right\} \\ &\times \int \{2\pi\phi\kappa(v_i)\}^{-q/2} \exp\left\{-\frac{\kappa^{-1}(v_i)}{2\phi} \|\mathbf{t}_{13(i)} - \mathbf{T}_{12(i)}\boldsymbol{\beta} - \mathbf{T}_{11(i)}\mathbf{b}_i\|^2\right\} d\mathbf{b}_i. \end{aligned} \quad (1.34)$$

Notando que $\mathbf{T}_{11(i)}$ é uma matriz não singular, podemos considerar a seguinte mudança de variáveis:

$$\mathbf{g}_i = \frac{\kappa^{-1/2}(v_i)}{\sqrt{\phi}} (\mathbf{t}_{13(i)} - \mathbf{T}_{12(i)}\boldsymbol{\beta} - \mathbf{T}_{11(i)}\mathbf{b}_i)$$

cujos diferencial é dado por $d\mathbf{g}_i = -\{\kappa^{-1}(v_i)/\phi\}^{1/2}\mathbf{T}_{11(i)}d\mathbf{b}_i$, agora como $|\mathbf{T}_{11(i)}|^2 = |\mathbf{Z}_i^T\mathbf{Z}_i + \Delta^T\Delta|$ e devido a que $\text{posto}(\mathbf{Z}_i) = q$, temos que a matriz $\mathbf{Z}_i^T\mathbf{Z}_i + \Delta^T\Delta$ é positiva definida, e portanto o Jacobiano da transformação resulta em

$$|J(\mathbf{b}_i \rightarrow \mathbf{g}_i)| = \left\{\frac{\kappa^{-1}(v_i)}{\phi}\right\}^{q/2} |\mathbf{T}_{11(i)}|.$$

Logo, a integral em (1.34) é dada por

$$\begin{aligned} & \int \{2\pi\phi\kappa(v_i)\}^{-q/2} \exp\left\{-\frac{\kappa^{-1}(v_i)}{2\phi} \|\mathbf{t}_{13(i)} - \mathbf{T}_{12(i)}\boldsymbol{\beta} - \mathbf{T}_{11(i)}\mathbf{b}_i\|^2\right\} d\mathbf{b}_i \\ &= \frac{1}{|\mathbf{T}_{11(i)}|} \int (2\pi)^{-q/2} \exp\left(-\frac{1}{2} \|\mathbf{g}_i\|^2\right) d\mathbf{g}_i \\ &= 1/|\mathbf{T}_{11(i)}|. \end{aligned}$$

Finalmente, obtemos que a densidade marginal $f(\mathbf{y}_i)$ assume a forma

$$\begin{aligned} f(\mathbf{y}_i) &= (2\pi\phi)^{-n_i/2} \frac{|\boldsymbol{\Delta}|}{|\mathbf{T}_{11(i)}|} \\ &\times \int_0^\infty \{\kappa(v_i)\}^{-n_i/2} \exp\left\{-\frac{\kappa^{-1}(v_i)}{2\phi} (\|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)}\boldsymbol{\beta}\|^2 + t_{33(i)}^2)\right\} dH(v_i). \end{aligned} \quad (1.35)$$

Como $|\mathbf{T}_{11(i)}|^2 = |\mathbf{Z}_i^T \mathbf{Z}_i + \boldsymbol{\Delta}^T \boldsymbol{\Delta}|$, então o fator em (1.35) resulta em $|\boldsymbol{\Delta}|/|\mathbf{T}_{11(i)}| = |\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{I}_{n_i}|^{-1/2}$, e portanto a densidade marginal derivada acima é equivalente à apresentada em (1.11).

Seja $u_i^*(\boldsymbol{\theta}) = (\|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)}\boldsymbol{\beta}\|^2 + t_{33(i)}^2)/\phi$ e $\mathbf{V}_i^* = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{I}_{n_i}$, para $i = 1, \dots, M$, então (1.35) pode ser escrito como

$$f(\mathbf{y}_i) = |2\pi\phi\mathbf{V}_i^*|^{-1/2} \int_0^\infty \{\kappa(v_i)\}^{-n_i/2} \exp\left\{-\frac{1}{2} \kappa^{-1}(v_i) u_i^*\right\} dH(v_i).$$

Considere agora

$$\mathbf{t}_i = \begin{pmatrix} \mathbf{t}_{23(i)} \\ t_{33(i)} \end{pmatrix} \quad \text{e} \quad \mathbf{U}_i = \begin{pmatrix} \mathbf{T}_{22(i)} \\ \mathbf{0} \end{pmatrix}, \quad i = 1, \dots, M.$$

Logo, temos que $u_i^*(\boldsymbol{\theta}) = \|\mathbf{t}_i - \mathbf{U}_i\boldsymbol{\beta}\|^2/\phi$, para $i = 1, \dots, M$. Deste modo a densidade marginal em (1.35) pode ser formulada usando a seguinte estrutura hierárquica:

$$\begin{aligned} \mathbf{t}_i | v_i &\stackrel{ind}{\sim} N_{n_i}(\mathbf{U}_i\boldsymbol{\beta}, \phi\kappa(v_i)\mathbf{V}_i^*) \quad \text{e} \\ v_i &\stackrel{ind}{\sim} H(v_i; \boldsymbol{\nu}), \quad i = 1, \dots, M. \end{aligned}$$

A função Q associada assume a forma $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, em que

$$Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = -\frac{n_i}{2} \log \phi - \frac{1}{2} \log |\mathbf{V}_i^*| - \frac{1}{2\phi} \hat{\kappa}_i \|\mathbf{t}_i - \mathbf{U}_i\boldsymbol{\beta}\|^2 \quad (1.36)$$

e $\hat{\kappa}_i = E(\kappa^{-1}(v_i)|\mathbf{t}_i, \hat{\boldsymbol{\theta}})$, para $i = 1, \dots, M$. Não obstante, computacionalmente é preferível escrever $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ como

$$Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = -\frac{n_i}{2} \log \phi + \log \left(\frac{|\Delta|}{|\mathbf{T}_{11(i)}|} \right) - \frac{1}{2\phi} \hat{\kappa}_i (\|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)}\boldsymbol{\beta}\|^2 + t_{33(i)}^2).$$

Considere a seguinte decomposição QR:

$$\begin{pmatrix} \sqrt{\hat{\kappa}_1} \mathbf{T}_{22(1)} & \sqrt{\hat{\kappa}_1} \mathbf{t}_{23(1)} \\ \vdots & \vdots \\ \sqrt{\hat{\kappa}_M} \mathbf{T}_{22(M)} & \sqrt{\hat{\kappa}_M} \mathbf{t}_{23(M)} \end{pmatrix} = \mathbf{Q}_0^* \begin{pmatrix} \mathbf{T}_{00} & \mathbf{t}_0 \\ \mathbf{0} & \mathbf{t}_{-1} \end{pmatrix}, \quad (1.37)$$

logo a função Q em (1.36) pode ser expressa como

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= -\frac{N}{2} \log \phi - \frac{1}{2\phi} \{ \|\mathbf{t}_{-1}\|^2 + \|\mathbf{t}_0 - \mathbf{T}_{00}\boldsymbol{\beta}\|^2 + \sum_{i=1}^M \hat{\kappa}_i t_{33(i)}^2 \} \\ &\quad + \sum_{i=1}^M \log \left(\frac{|\Delta|}{|\mathbf{T}_{11(i)}|} \right). \end{aligned} \quad (1.38)$$

Podemos perfilar a função Q obtendo expressões explícitas para $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ condicionais ao valor de $\boldsymbol{\lambda}$ e deste modo tornar mais simples o passo CM do algoritmo ECM apresentado na Seção 1.3.2. Adicionalmente, isto permite escrever uma versão multiciclo de tal algoritmo.

Para \mathbf{T}_{00} não singular, o valor para $\boldsymbol{\beta}$ que maximiza $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ em (1.38) satisfaz o sistema triangular

$$\mathbf{T}_{00}\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = \mathbf{t}_0.$$

Substituindo $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ em (1.38) obtemos que $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ é maximizada com relação a ϕ para

$$\hat{\phi}(\boldsymbol{\lambda}) = \frac{1}{N} \left\{ \|\mathbf{t}_{-1}\|^2 + \sum_{i=1}^M \hat{\kappa}_i t_{33(i)}^2 \right\}.$$

Deste modo a função Q perfilada assume a forma

$$Q(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}}) = Q(\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \hat{\phi}(\boldsymbol{\lambda}), \boldsymbol{\lambda}|\hat{\boldsymbol{\theta}}) = c^* - \frac{N}{2} \log \hat{\phi}(\boldsymbol{\lambda}) + \sum_{i=1}^M \log \left(\frac{|\Delta|}{|\mathbf{T}_{11(i)}|} \right), \quad (1.39)$$

em que c^* representa uma constante.

Finalmente, apresentamos a $(r+1)$ -ésima etapa da versão multiciclo do algoritmo ECM dado na Seção 1.3.2 usando os procedimentos descritos anteriormente.

Algoritmo 2

Aplicar a decomposição (1.25) a $(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Y}_i)$, para $i = 1, \dots, M$.

Passo E: usando estimativas da r -ésima etapa, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(r)}$ calcular $u_i^*(\hat{\boldsymbol{\theta}})$ como

$$u_i^*(\hat{\boldsymbol{\theta}}) = \frac{1}{\hat{\phi}} \{ \|\mathbf{t}_{23(i)} - \mathbf{T}_{22(i)}\hat{\boldsymbol{\beta}}\|^2 + t_{33(i)}^2 \}$$

e avaliar $\hat{\kappa}_i = \mathbb{E}(\kappa^{-1}(v_i)|\mathbf{t}_i, \hat{\boldsymbol{\theta}})$, para $i = 1, \dots, M$, usando as expressões desenvolvidas na Seção 1.3.3 para alguma mistura de escala normal particular.

Passo CM multiciclo: maximização da esperança do logaritmo da função de verossimilhança de dados completos, $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ usando uma seqüência de passos CM.

Passo CM-1: atualizar $\boldsymbol{\beta}^{(r+1)}$ como a solução do sistema

$$\mathbf{T}_{00}\boldsymbol{\beta}^{(r+1)} = \mathbf{t}_0,$$

em que \mathbf{T}_{00} e \mathbf{t}_0 são obtidos da decomposição em (1.37).

Passo E: executar o passo E com $\hat{\boldsymbol{\theta}} = (\boldsymbol{\beta}^{(r+1)T}, \phi^{(r)}, \boldsymbol{\lambda}^{(r)T})^T$.

Passo CM-2: fixar $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(r+1)}$, $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{(r)}$ e atualizar $\phi^{(r+1)}$ como

$$\phi^{(r+1)} = \frac{1}{N} \left\{ \|\mathbf{t}_{-1}\|^2 + \sum_{i=1}^M \hat{\kappa}_i t_{33(i)}^2 \right\}.$$

Passo E: executar o passo E com $\hat{\boldsymbol{\theta}} = (\boldsymbol{\beta}^{(r+1)T}, \phi^{(r+1)}, \boldsymbol{\lambda}^{(r)T})^T$.

Passo CM-3: fixar $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(r+1)}$, $\hat{\phi} = \phi^{(r+1)}$ e atualizar $\boldsymbol{\lambda}^{(r+1)}$ como

$$\boldsymbol{\lambda}^{(r+1)} = \arg \min_{\boldsymbol{\lambda}} \{-Q(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})\},$$

em que $Q(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$ é dada em (1.39).

Então, fazer $\boldsymbol{\theta}^{(r)} = (\boldsymbol{\beta}^{(r+1)T}, \phi^{(r+1)}, \boldsymbol{\lambda}^{(r+1)T})^T$ e voltar à etapa E. Este algoritmo itera entre os passos E e CM até que a seqüência $\boldsymbol{\theta}^{(r)}$ atinja convergência.

As decomposições matriciais consideradas nesta seção provêm de métodos eficientes para estimação de parâmetros no modelo linear misto hierárquico, permitindo avaliar de forma simples e rápida as expressões envolvidas nos algoritmos tipo-EM considerados. Além de reduzir a dimensão das quantidades que devem ser manipuladas e armazenadas.

Com efeito, (1.25) corresponde a uma pré-decomposição que é executada somente em uma ocasião e pode ser armazenada na matriz de dados original

$$\begin{pmatrix} \mathbf{Z}_1 & \mathbf{X}_1 & \mathbf{Y}_1 \\ \vdots & \vdots & \vdots \\ \mathbf{Z}_M & \mathbf{X}_M & \mathbf{Y}_M \end{pmatrix}.$$

Por outro lado, a decomposição (1.26) é calculada no início de cada passo E, note porém que não é requerido armazenar muitas das matrizes envolvidas nesta decomposição.

O algoritmo 1 desta seção requer o armazenamento de $u_i(\hat{\boldsymbol{\theta}})$ e $\hat{\kappa}_i$ assim como de $\hat{\mathbf{b}}_i$ e $\mathbf{T}_{11(i)}$ para $i = 1, \dots, M$, com motivo de avaliar $Q_2(\boldsymbol{\lambda}, \hat{\phi}|\hat{\boldsymbol{\theta}})$ na etapa CM-3. Adicionalmente, esta versão multiciclo requer o armazenamento de $\mathbf{t}_{23(i)}$, $\mathbf{T}_{22(i)}$ e $t_{33(i)}$, $i = 1, \dots, M$, que pode ser liberado antes da etapa CM-3. Notamos que este algoritmo sem a modificação multiciclo não precisa de armazenamento para $\mathbf{t}_{23(i)}$, $\mathbf{T}_{22(i)}$ e $t_{33(i)}$.

No caso do algoritmo 2 apresentado nesta seção somente é requerido armazenar $\mathbf{t}_{23(i)}$, $\mathbf{T}_{22(i)}$ e $t_{33(i)}$, $i = 1, \dots, M$, para a atualização de $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ nas etapas CM-1 e CM-2, respectivamente, assim como para os passos E na versão multiciclo. Apreciamos também que o espaço armazenado pode ser liberado antes da etapa CM-3. Note que fica fácil avaliar a função $Q(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$ dada em (1.39), uma vez que $\boldsymbol{\Delta}$ e $\mathbf{T}_{11(i)}$ são matrizes triangulares e portanto seu determinante é o produto de seus elementos diagonais, permitindo deste modo simplificar o procedimento de otimização requerido na etapa CM-3.

Finalmente, a estrutura geral apresentada em (1.3)-(1.4) pode ser recuperada usando um fator raiz-quadrada de $\boldsymbol{\Sigma}_i^*$ como $\boldsymbol{\Sigma}_i^* = \boldsymbol{\Sigma}_i^{*1/2} \boldsymbol{\Sigma}_i^{*1/2}$ com $\boldsymbol{\Sigma}_i = \phi \boldsymbol{\Sigma}_i^*$ em que $\boldsymbol{\Sigma}_i^*$ representa uma versão *escalada* de $\boldsymbol{\Sigma}_i$ e deve-se considerar

$$\mathbf{Y}_i^* = \boldsymbol{\Sigma}_i^{*-1/2} \mathbf{Y}_i, \quad \mathbf{X}_i^* = \boldsymbol{\Sigma}_i^{*-1/2} \mathbf{X}_i \quad \text{e} \quad \mathbf{Z}_i^* = \boldsymbol{\Sigma}_i^{*-1/2} \mathbf{Z}_i,$$

deste modo, obtemos que

$$\begin{aligned} \mathbf{Y}_i^* | \mathbf{b}_i, v_i &\stackrel{ind}{\sim} N_{n_i}(\mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{Z}_i^* \mathbf{b}_i, \phi \kappa(v_i) \mathbf{I}), & \mathbf{b}_i | v_i &\stackrel{ind}{\sim} N_q(\mathbf{0}, \phi \kappa(v_i) \mathbf{D}) \quad e \\ v_i &\stackrel{ind}{\sim} H(v_i; \boldsymbol{\nu}), & i &= 1, \dots, M, \end{aligned}$$

seguem a formulação hierárquica utilizada no desenvolvimento das expressões computacionais derivadas nesta seção. Note que o Jacobiano da transformação resulta em $|J(\mathbf{Y}_i \rightarrow \mathbf{Y}_i^*)| = |\boldsymbol{\Sigma}_i^{*-1/2}|$. Logo, considerando o seguinte vetor de dados completos $\mathbf{Y}_c^* = (\mathbf{Y}^{*T}, \mathbf{b}^T, \mathbf{v}^T)^T$ com $\mathbf{Y}^{*T} = (\mathbf{Y}_1^{*T}, \dots, \mathbf{Y}_M^{*T})^T$, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_M^T)^T$ e $\mathbf{v} = (v_1, \dots, v_M)^T$, obtemos

$$L(\boldsymbol{\theta} | \mathbf{Y}_c) = - \sum_{i=1}^M \log |\boldsymbol{\Sigma}_i^{*-1/2}| + L(\boldsymbol{\theta} | \mathbf{Y}_c^*)$$

que é equivalente à equação apresentada em (1.14).

1.4.1 Implementação

Os procedimentos computacionais descritos neste capítulo formam a base de uma biblioteca para S-PLUS, `roblme` para o ajuste de modelos lineares com efeitos mistos hierárquico usando a classe das misturas de escala normal.

A biblioteca `roblme` utiliza as facilidades das rotinas numéricas disponíveis em BLAS (Lawson, Hanson, Kincaid e Krogh, 1979; Dongarra, DuCroz, Hammarling e Hanson, 1988 e Dongarra, DuCroz, Hammarling e Duff, 1990) para operações matriciais básicas e LINPACK (Dongarra, Bunch, Moler e Stewart, 1979) para as decomposições matriciais e solução de equações lineares. O código fonte em S, C e FORTRAN da biblioteca `roblme` está disponível em <http://www.ime.usp.br/~osorio/roblme/>.

Capítulo 2

Diagnósticos de Influência

O objetivo deste capítulo é desenvolver diagnósticos de influência para o modelo elíptico linear misto apresentado no capítulo anterior. Aplicamos o método de influência local proposto por Cook (1986) e Zhu e Lee (2001) sob diversos esquemas de perturbação no modelo linear com efeitos mistos considerando distribuições de contornos elípticos seguindo as formulações marginal e hierárquica discutidas no Capítulo 1. A partir da proposta de Wei, Hu e Fung (1998) aplicamos o método de alavanca para modelos com dados incompletos. As quantidades requeridas nos procedimentos de diagnóstico considerados neste capítulo são calculadas eficientemente usando a metodologia de diferenciação matricial descrita em Magnus e Neudecker (1988).

2.1 Introdução

Em estudos de modelagem estatística, uma etapa de reconhecida importância corresponde à validação das suposições do modelo mediante estudos de sensibilidade. Cook (1986) propôs um enfoque geométrico para a verificação das suposições do modelo assim como a identificação de dados aberrantes e/ou influentes, por meio de estudar o efeito de introduzir pequenas perturbações no modelo (ou dados) usando uma medida de influência apropriada. Esta etapa permite sugerir modelos que reagem de forma adequada ao tipo de perturbação introduzida, assim como considerar algum procedimento alternativo de estimação.

A técnica de influência local tem se constituído numa ferramenta útil para realizar análises de sensibilidade de modelos estatísticos e tem sido sido

amplamente utilizada em modelos de regressão linear e não-linear. Em particular, para modelos lineares com efeitos mistos, Beckman, Nachtsheim e Cook (1987) aplicam este método para detectar observações influentes num modelo linear misto com erros normais e ênfase na influência de observações individuais. Por outro lado, Lesaffre e Verbeke (1998) consideram a influência local em modelos lineares mistos normais sob o esquema de ponderação de casos e mais recentemente Lee e Xu (2004) estudam a influência local em modelos com efeitos mistos não-lineares sob normalidade. Exemplos da aplicação do método de influência local em modelos lineares mistos não normais são descritos em Ouwens, Tan e Berger (2001) e Zhu e Lee (2003) que estudam a influência em modelos lineares generalizados mistos.

Diagnóstico de influência em modelos elípticos lineares e não-lineares tem sido discutido por diversos autores. Por exemplo, Galea, Paula e Bolfarine (1997), Liu (2000) e Galea, Paula e Uribe-Opazo (2003) consideram estudos de influência em modelos de regressão linear univariados. Galea, Paula e Cysneiros (2005) estendem influência local para modelos não-lineares univariados considerando vários esquemas de perturbação e Liu (2002) e Díaz-García, Galea e Leiva-Sánchez (2003) discutem diagnósticos de influência em modelos elípticos lineares multivariados. Contudo, pouco tem sido investigado a respeito da obtenção de medidas de influência local em modelos elípticos lineares com efeitos mistos.

O principal interesse de estudar diagnóstico de influência em modelos elípticos com efeitos mistos vem do fato de que a classe elíptica é rica em distribuições com caudas mais pesadas do que a normal e deste modo pode acomodar melhor observações aberrantes. Contudo, ainda esses modelos podem sofrer o efeito de observações influentes, havendo portanto a necessidade de realizar estudos de sensibilidade nesta classe. Por outro lado, esses procedimentos também permitem selecionar modelos dentro da classe elíptica que se comportam adequadamente ante o tipo de perturbação considerada.

Consideramos o procedimento de influência local (Cook, 1986; Zhu e Lee, 2001) sob diferentes esquemas de perturbação assim como o método de alavancas generalizadas (Wei, Hu e Fung, 1998). Nas seções a seguir apresentamos um breve resumo de cada uma dessas metodologias.

2.2 Influência Local

O objetivo do método de influência local é investigar o comportamento de alguma medida de influência $T(\omega)$ quando pequenas perturbações são in-

troduzidas no modelo (ou dados) por meio de um vetor de perturbação $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$. Neste trabalho, usamos medidas de influência baseadas na função de verossimilhança, como: o afastamento da verossimilhança e a função Q -afastamento (Zhu e Lee, 2001) para estudar a influência no modelo linear misto marginal e hierárquico, respectivamente.

Seja $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ o logaritmo da função de verossimilhança perturbada. Assume-se que o modelo não perturbado é encaixado no modelo perturbado, isto é, existe $\boldsymbol{\omega}_0 \in \Omega$ (vetor de não perturbação) tal que $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$. Considere $\hat{\boldsymbol{\theta}}_\omega$ a estimativa de máxima verossimilhança sob o modelo perturbado. Uma maneira de determinar a influência de uma perturbação particular na estimativa de máxima verossimilhança é comparar $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_\omega$ segundo alguma medida de influência $T(\boldsymbol{\omega})$ conforme $\boldsymbol{\omega}$ varia em Ω .

Para modelos baseados na função de verossimilhança, Cook (1986) sugere considerar o *afastamento da verossimilhança*

$$LD(\boldsymbol{\omega}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_\omega)\}, \quad (2.1)$$

a fim de comparar as diferenças entre $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_\omega$ relativas aos contornos do logaritmo da função de verossimilhança não perturbada $L(\boldsymbol{\theta})$. Cook (1986) propõe estudar o comportamento local de $LD(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0$, usando a curvatura normal na direção \mathbf{h} , dada por

$$C_h(\boldsymbol{\theta}) = -2\mathbf{h}^T \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}^T \ddot{\mathbf{L}}^{-1}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Delta}_{\boldsymbol{\omega}_0} \mathbf{h}, \quad (2.2)$$

em que \mathbf{h} corresponde a uma direção unitária, isto é, $\|\mathbf{h}\| = 1$ e

$$\ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}}) = \left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad \boldsymbol{\Delta}_\omega = \left. \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})} \quad (2.3)$$

com $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$ (a matriz de informação observada) e $\boldsymbol{\Delta}_\omega$ sendo avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Sugere-se considerar o diagrama de índices da direção \mathbf{h}_{\max} associada à maior curvatura $C_{h_{\max}}(\boldsymbol{\theta})$ (Cook, 1986). Por outro lado, Lesaffre e Verbeke (1998) propuseram estudar a direção associada à i -ésima observação, ou seja, o diagrama de índices de $C_{h_i}(\boldsymbol{\theta})$ em que \mathbf{h}_i denota um vetor $q \times 1$ com um 1 na i -ésima posição e zeros nas restantes. Tais diagramas podem indicar aquelas observações que exercem notável influência sobre $LD(\boldsymbol{\omega})$.

Para dados incompletos Zhu e Lee (2001) propuseram um enfoque para realizar estudos de influência considerando

$$f_Q(\boldsymbol{\omega}) = 2\{Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_\omega|\hat{\boldsymbol{\theta}})\}, \quad (2.4)$$

como uma medida da diferença entre $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_\omega$, em que $\hat{\boldsymbol{\theta}}_\omega$ denota a estimativa de $\boldsymbol{\theta}$ que maximiza

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}) = E\{L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) | \mathbf{Y}, \hat{\boldsymbol{\theta}}\}.$$

Os logaritmos das verossimilhanças para dados completos e dados observados do modelo perturbado são denotados por $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ e $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y})$, respectivamente. Assume-se que existe um vetor de não perturbação $\boldsymbol{\omega}_0$ tal que $L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c) = L(\boldsymbol{\theta} | \mathbf{Y}_c)$ e $L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}) = L(\boldsymbol{\theta} | \mathbf{Y})$ para todo $\boldsymbol{\theta}$.

Analogamente a Cook (1986), Zhu e Lee (2001) estudam o comportamento da superfície $\boldsymbol{\gamma}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^T, f_Q(\boldsymbol{\omega}))^T$, mediante a curvatura normal $C_{f_Q, h}$ na direção do vetor unitário $\mathbf{h} \in \mathbb{R}^q$, dada por

$$C_{f_Q, h}(\boldsymbol{\theta}) = -2\mathbf{h}^T \boldsymbol{\Delta}_{\omega_0}^T \ddot{Q}^{-1}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Delta}_{\omega_0} \mathbf{h}, \quad (2.5)$$

em que

$$\ddot{Q}(\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{e} \quad \boldsymbol{\Delta}_\omega = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})} \quad (2.6)$$

são avaliadas em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$. Como no caso da curvatura dada em (2.2) sugere-se o exame do gráfico de índices do autovetor associado ao maior autovalor de $\ddot{\mathbf{T}} = -\boldsymbol{\Delta}^T \ddot{Q}^{-1} \boldsymbol{\Delta}$. Usando a proposta de Lesaffre e Verbeke (1998) também é possível calcular a influência local total $C_{f_Q, i}(\boldsymbol{\theta})$.

No estudo dos gráficos de \mathbf{h}_{\max} ou de alguma outra direção de interesse é importante determinar quais observações de fato são influentes. Com esse objetivo alguns autores têm desenvolvido critérios para julgar a magnitude das medidas de influência local (veja Lesaffre e Verbeke, 1998 e Zhu e Zhang, 2004) assim como a construção de gráficos de estatísticas relevantes contra a perturbação sob estudo. Por exemplo, o gráfico de $LD(\boldsymbol{\omega}(a))$ versus a em que $\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a\mathbf{h}$, com $a \in \mathbb{R}$ e \mathbf{h} alguma direção selecionada (Wu e Luo, 1993 e Cook, 1997).

Em certas ocasiões podemos estar interessados num subconjunto $\boldsymbol{\theta}_1$ de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$; neste caso o afastamento da verossimilhança é dado por

$$LD_1(\boldsymbol{\omega}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{1\omega}, \hat{\boldsymbol{\theta}}_2(\hat{\boldsymbol{\theta}}_{1\omega}))\},$$

em que $\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1)$ representa a estimativa de máxima verossimilhança no modelo não perturbado de $\boldsymbol{\theta}_2$ para $\boldsymbol{\theta}_1$ fixado.

Considere então a partição de $\ddot{\mathbf{L}}$,

$$\ddot{\mathbf{L}} = \begin{pmatrix} \ddot{\mathbf{L}}_{11} & \ddot{\mathbf{L}}_{12} \\ \ddot{\mathbf{L}}_{21} & \ddot{\mathbf{L}}_{22} \end{pmatrix}$$

de acordo com a partição dada por $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ e seja

$$\mathbf{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix}.$$

Então, a curvatura normal na direção \mathbf{h} para $\boldsymbol{\theta}_1$ assume a forma

$$C_h(\boldsymbol{\theta}_1) = -2\mathbf{h}^T \boldsymbol{\Delta}_{\omega_0}^T \{ \ddot{\mathbf{L}}^{-1}(\hat{\boldsymbol{\theta}}) - \mathbf{B}_{22} \} \boldsymbol{\Delta}_{\omega_0} \mathbf{h},$$

e a direção de maior curvatura corresponde ao autovetor associado ao maior autovalor de $\boldsymbol{\Delta}_{\omega_0}^T \{ \ddot{\mathbf{L}}^{-1}(\hat{\boldsymbol{\theta}}) - \mathbf{B}_{22} \} \boldsymbol{\Delta}_{\omega_0}$. Expressões equivalentes também podem ser derivadas para $f_{1Q}(\boldsymbol{\omega})$.

Com o intuito de obter uma curvatura invariante a mudanças uniformes na escala, Poon e Poon (1999) propuseram a curvatura conformal $B_h(\boldsymbol{\theta}) = C_h(\boldsymbol{\theta}) / \| 2\boldsymbol{\Delta}_{\omega_0}^T \ddot{\mathbf{L}}^{-1}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Delta}_{\omega_0} \|_F$, em que $\|\cdot\|_F$ denota a norma Frobenius definida como $\|\mathbf{A}\|_F = \{\text{tr}(\mathbf{A}^T \mathbf{A})\}^{1/2}$ com \mathbf{A} sendo uma matriz $r \times s$. Uma propriedade interessante da curvatura conformal é que para qualquer direção unitária \mathbf{h} temos que $0 \leq B_h(\boldsymbol{\theta}) \leq 1$. Isto permite, por exemplo, a comparação de curvaturas entre diferentes modelos sob erros elípticos.

2.3 Método de Alavanca Generalizado para Modelos com Dados Incompletos

Vários autores têm destacado a importância do conceito de ponto de alavanca para diagnósticos em modelos de regressão (Hoaglin e Welsh, 1978; Cook e Weisberg, 1982; Chatterjee e Hadi 1986 e Andrade, 2004, dentre outros). A idéia do ponto de alavanca tem sido generalizada para modelos mais complexos. Por exemplo, Ross (1987), St. Laurent e Cook (1992) e Wei et al. (1998) estenderam o método de alavanca para modelos não-lineares; Banerjee e Frees (1997) definem matrizes de alavanca para cada indivíduo no contexto de medidas repetidas; Christensen, Pearson e Johnson (1992) sugerem uma medida de alavanca para modelos lineares mistos sob normalidade; Paula (1999) considera o método de alavanca em modelos de regressão linear quando o vetor de parâmetros $\boldsymbol{\theta}$ é restrito em desigualdades

lineares e, mais recentemente, Nobre e Singer (2006) propõem incorporar a informação dos efeitos aleatórios ajustados $\hat{\mathbf{b}}_i$ e consideram a matriz de alavancas generalizadas em modelos lineares com efeitos mistos baseados em $\hat{\mathbf{Y}}_i = \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\mathbf{b}}_i$.

O objetivo principal do método de alavanca é medir a influência da resposta observada no seu próprio valor ajustado. Tipicamente esta medida tem sido definida como uma taxa de mudança instantânea no valor predito relativo ao valor da variável resposta (veja, por exemplo, St. Laurent e Cook, 1992 e Wei et al., 1998). Baseados na proposta de Wei et al. (1998) assim como nos trabalhos de Zhu e Lee (2001) e Zhu et al. (2001) estendemos o método de alavanca para dados incompletos.

Para fixar idéias, seja \mathbf{Y} o vetor de respostas observadas com densidade $f(\mathbf{Y}; \boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ representa um vetor de parâmetros p^* -dimensional. Considere $\boldsymbol{\mu} = E(\mathbf{Y})$ tal que $\boldsymbol{\mu}$ pode ser expresso como $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$ e seja $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y})$ a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$. Deste modo $\hat{\mathbf{Y}} = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$ representa o vetor de respostas preditas. A matriz de alavancas generalizadas é definida como $\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \partial\hat{\mathbf{Y}}/\partial\mathbf{Y}^T$. Segue de Wei et al. (1998) que para modelos com função de verossimilhança a matriz de alavancas generalizadas assume a forma

$$\mathbf{GL}(\hat{\boldsymbol{\theta}}) = [\mathbf{D}_\theta\{-\ddot{\mathbf{L}}(\boldsymbol{\theta})\}^{-1}\ddot{\mathbf{L}}(\boldsymbol{\theta}, \mathbf{Y})]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}, \quad (2.7)$$

em que $\mathbf{D}_\theta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^T$, $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$ representa a matriz de informação observada e $\ddot{\mathbf{L}}(\boldsymbol{\theta}, \mathbf{Y}) = \partial^2L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\mathbf{Y}^T$. Alguns exemplos da aplicação do método de alavanca para modelos lineares e não-lineares são dados em Wei et al. (1998), Cysneiros e Paula (2005) e Galea, Paula e Cysneiros (2005).

Para modelos de maior complexidade assim como modelos com dados incompletos, avaliar a expressão da matriz de alavancas generalizadas dada em (2.7) pode ser mais difícil. Baseados na função Q do algoritmo EM, vários autores têm proposto medidas para realizar análises de influência. Por exemplo, Zhu et al. (2001) discutem diagnósticos de influência mediante a eliminação de casos no contexto de dados incompletos e Zhu e Lee (2001) propõem a influência local para modelos com dados incompletos. Esses enfoques têm sido aplicados na análise de influência em modelos com efeitos mistos, tais como: diagnóstico por eliminação de casos em modelos lineares mistos sob normalidade (Fei e Pan, 2003), influência local em modelos lineares mistos generalizados (Zhu e Lee, 2003) e recentemente o estudo de influência local em modelos mistos não-lineares (Lee e Xu, 2004). Contudo,

o conceito de alavanca em modelos com dados incompletos aparece como um t3pico ainda n3o tratado.

Considere a fun33o Q proveniente da esperana condicional do logaritmo da verossimilhan3a de dados completos no algoritmo EM (veja equa33o (10)). Usando o lema em Wei et al. (1998), definimos a *matriz de alavancas generalizadas para modelos com dados incompletos* como

$$\mathbf{GL}_Q(\hat{\boldsymbol{\theta}}) = [\mathbf{D}_\theta \{-\ddot{\mathbf{Q}}(\boldsymbol{\theta})\}^{-1} \ddot{\mathbf{Q}}(\boldsymbol{\theta}, \mathbf{Y})] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(Y)}, \quad (2.8)$$

em que $\mathbf{D}_\theta = \partial \boldsymbol{\mu} / \partial \boldsymbol{\theta}^T$, $\ddot{\mathbf{Q}}(\boldsymbol{\theta}) = \partial^2 Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T$ e $\ddot{\mathbf{Q}}(\boldsymbol{\theta}, \mathbf{Y}) = \partial^2 Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta} \partial \mathbf{Y}^T$.

A ampla aplicabilidade do algoritmo EM permite utilizar a defini33o dada por (2.8) para diversas situa33es n3o ficando restrito somente a manipular modelos com dados incompletos. Note tamb3m que na aus3ncia de dados incompletos (2.8) reduz-se a $\mathbf{GL}(\hat{\boldsymbol{\theta}})$ dado por (2.7). Neste cap3tulo derivamos a matriz de alavancas generalizadas para o modelo linear misto em suas formula33es marginal assim como hier3rquica a partir das equa33es (2.7) e (2.8), respectivamente.

Seja $p_0 = \sum_{j=1}^K \mathbf{GL}_{jj}(\hat{\boldsymbol{\theta}}) = \text{tr } \mathbf{GL}(\hat{\boldsymbol{\theta}})$, logo podemos utilizar $c_0 p_0 / K$ como um instrumento para comparar os elementos diagonais de $\mathbf{GL}(\hat{\boldsymbol{\theta}})$, em que K denota a dimens3o do vetor de respostas observadas \mathbf{Y} . Portanto, podemos olhar com aten33o aquelas observa33es tais que $\mathbf{GL}_{jj}(\hat{\boldsymbol{\theta}}) > c_0 p_0 / K$, em que a constante positiva c_0 pode ser escolhida, por exemplo, como $c_0 = 2$. Contudo, crit3rios mais objetivos para julgar observa33es com alta alavancagem ainda s3o objeto de investiga33o.

Nas se33es a seguir derivamos a curvatura normal assim como a matriz de alavancas generalizadas para o modelo el3ptico linear misto em suas formula33es marginal e hier3rquica. Obtemos as quantidades requeridas para os procedimentos de diagn3stico de maneira eficiente usando a metodologia de diferencia33o matricial descrita em Nel (1980) e Magnus e Neudecker (1998). Exemplos da aplica33o deste procedimento na deriva33o da curvatura normal para o m3todo de influ3ncia local t3m sido considerados por Liu (2000, 2002), D3az-Garc3a et al. (2003) e Osorio, Paula e Galea (2006).

2.4 Influ3ncia Local no Modelo Linear Misto Marginal

Considere o modelo discutido na Se33o 1.2, isto 3, $\mathbf{Y}_i \sim EC_{n_i}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i; g)$ para $i = 1, \dots, M$, em que \mathbf{X}_i 3 uma matriz de planejamento $n_i \times p$ e

$V_i = Z_i \Psi Z_i^T + \Sigma_i$. Nesta seção derivamos a matriz de informação observada $-\ddot{L}$ e a matriz Δ segundo vários esquemas de perturbação.

O logaritmo da função de verossimilhança para o modelo não perturbado é dado por $L(\theta) = \sum_{i=1}^M L_i(\theta)$, com

$$L_i(\theta) = -\frac{1}{2} \log |V_i| + \log g(u_i), \quad (2.9)$$

em que $u_i = (Y_i - X_i \beta)^T V_i^{-1} (Y_i - X_i \beta)$, para $i = 1, \dots, M$.

2.4.1 Matriz de Informação Observada

Seja $\theta = (\beta^T, \alpha^T, \lambda^T)^T$ o vetor de parâmetros de interesse. A matriz de informação observada $-\ddot{L}(\theta)$ (veja Apêndice A.1) avaliada em $\theta = \hat{\theta}$ é dada por $-\ddot{L}(\hat{\theta}) = -\sum_{i=1}^M \ddot{L}_i(\hat{\theta})$, em que $\ddot{L}_i(\hat{\theta})$ tem forma particionada:

$$\ddot{L}_i(\hat{\theta}) = \frac{\partial^2 L_i(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}} = \begin{pmatrix} \ddot{L}_{11,i} & \ddot{L}_{12,i} & \ddot{L}_{13,i} \\ \ddot{L}_{12,i}^T & \ddot{L}_{22,i} & \ddot{L}_{23,i} \\ \ddot{L}_{13,i}^T & \ddot{L}_{23,i}^T & \ddot{L}_{33,i} \end{pmatrix}, \quad (2.10)$$

em que

$$\ddot{L}_{11,i} = 2X_i^T \hat{V}_i^{-1} \{W_g(\hat{u}_i) \hat{V}_i + 2W_g'(\hat{u}_i) \hat{r}_i \hat{r}_i^T\} \hat{V}_i^{-1} X_i,$$

$$\ddot{L}_{12,i} = \frac{\partial^2 L_i(\theta)}{\partial \beta \partial \alpha^T} \Big|_{\theta=\hat{\theta}} \quad \text{e} \quad \ddot{L}_{13,i} = \frac{\partial^2 L_i(\theta)}{\partial \beta \partial \lambda^T} \Big|_{\theta=\hat{\theta}},$$

com

$$\begin{aligned} \frac{\partial^2 L_i(\theta)}{\partial \beta \partial \alpha_r} \Big|_{\theta=\hat{\theta}} &= 2X_i \hat{V}_i^{-1} \{W_g(\hat{u}_i) \hat{V}_i + W_g'(\hat{u}_i) \hat{r}_i \hat{r}_i^T\} \hat{V}_i^{-1} \dot{\Sigma}_i(r) \hat{V}_i^{-1} \hat{r}_i \quad \text{e} \\ \frac{\partial^2 L_i(\theta)}{\partial \beta \partial \lambda_s} \Big|_{\theta=\hat{\theta}} &= 2X_i \hat{V}_i^{-1} \{W_g(\hat{u}_i) \hat{V}_i + W_g'(\hat{u}_i) \hat{r}_i \hat{r}_i^T\} \hat{V}_i^{-1} Z_i \dot{\Psi}(s) Z_i^T \hat{V}_i^{-1} \hat{r}_i, \end{aligned}$$

para $r = 1, \dots, k$; $s = 1, \dots, d$, e

$$\ddot{L}_{22,i} = \frac{\partial^2 L_i(\theta)}{\partial \alpha \partial \alpha^T} \Big|_{\theta=\hat{\theta}} \quad \text{e} \quad \ddot{L}_{23,i} = \frac{\partial^2 L_i(\theta)}{\partial \alpha \partial \lambda^T} \Big|_{\theta=\hat{\theta}},$$

cujos elementos são dados por

$$\begin{aligned}
\left. \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \alpha_r \partial \alpha_s} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \widehat{\mathbf{V}}_i^{-1} \{ \dot{\boldsymbol{\Sigma}}_i(r) \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) - \ddot{\boldsymbol{\Sigma}}_i(r, s) \} \\
&+ \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \{ W_g'(\hat{u}_i) \dot{\boldsymbol{\Sigma}}_i(r) \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) - W_g(\hat{u}_i) \ddot{\boldsymbol{\Sigma}}_i(r, s) \\
&+ W_g(\hat{u}_i) \dot{\boldsymbol{\Sigma}}_i(r) \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) + W_g(\hat{u}_i) \dot{\boldsymbol{\Sigma}}_i(s) \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \} \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i \quad \text{e} \\
\left. \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \alpha_s \partial \lambda_t} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \\
&+ \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \{ W_g'(\hat{u}_i) \dot{\boldsymbol{\Sigma}}_i(s) \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \\
&+ W_g(\hat{u}_i) \dot{\boldsymbol{\Sigma}}_i(s) \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T + W_g(\hat{u}_i) \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) \} \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i,
\end{aligned}$$

para $r, s = 1, \dots, k$; $t = 1, \dots, d$. Finalmente

$$\ddot{\mathbf{L}}_{33,i} = \left. \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

em que o elemento (t, u) de $\ddot{\mathbf{L}}_{33,i}$ fica dado por

$$\begin{aligned}
\left. \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \lambda_t \partial \lambda_u} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \{ \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(u) - \ddot{\boldsymbol{\Psi}}(t, u) \} \mathbf{Z}_i^T \\
&+ \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \{ W_g'(\hat{u}_i) \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(u) - W_g(\hat{u}_i) \ddot{\boldsymbol{\Psi}}(t, u) \\
&+ W_g(\hat{u}_i) \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(u) + W_g(\hat{u}_i) \dot{\boldsymbol{\Psi}}(u) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \} \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i,
\end{aligned}$$

para $t, u = 1, \dots, d$. Temos que $\hat{u}_i = \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i$, $\widehat{\mathbf{r}}_i = \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{V}}_i = \mathbf{Z}_i \boldsymbol{\Psi}(\widehat{\boldsymbol{\lambda}}) \mathbf{Z}_i^T + \boldsymbol{\Sigma}_i(\widehat{\boldsymbol{\alpha}})$, para $i = 1, \dots, M$, com $\dot{\boldsymbol{\Sigma}}_i(r) = \partial \boldsymbol{\Sigma}_i / \partial \alpha_r$, $\ddot{\boldsymbol{\Sigma}}_i(r, s) = \partial^2 \boldsymbol{\Sigma}_i / \partial \alpha_r \partial \alpha_s$, $\dot{\boldsymbol{\Psi}}(t) = \partial \boldsymbol{\Psi} / \partial \lambda_t$ e $\ddot{\boldsymbol{\Psi}}(t, u) = \partial^2 \boldsymbol{\Psi} / \partial \lambda_t \partial \lambda_u$ avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

2.4.2 Esquemas de Perturbação

Para o modelo linear misto marginal definido na Seção 1.2, consideramos vários esquemas de perturbação, a saber: ponderação de casos, da matriz de escala, das variáveis explicativas e da resposta. No contexto de dados longitudinais esses esquemas têm sido considerados por Osorio, Paula e Galea (2006). Através desses esquemas de perturbação é possível verificar a sensibilidade de algumas suposições do modelo, assim como compreender melhor o processo de modelagem.

Para cada um dos esquemas de perturbação considerados a matriz Δ assume a forma

$$\Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \end{pmatrix},$$

em que $\Delta_1 = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^T \in \mathbb{R}^{p \times q}$, $\Delta_2 = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\omega}^T \in \mathbb{R}^{k \times q}$, $\Delta_3 = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\lambda}\partial\boldsymbol{\omega}^T \in \mathbb{R}^{d \times q}$ e q denota a dimensão do vetor de perturbação $\boldsymbol{\omega}$. Avaliando $d_{\beta\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})$, $d_{\alpha\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})$ e $d_{\lambda\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})$ descritos no Apêndice A.1 em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ obtemos Δ_1 , Δ_2 e Δ_3 , respectivamente.

Ponderação de Casos

Considere atribuir uma ponderação arbitrária para verossimilhança individual como

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

em que $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$ denota o vetor de ponderações, $0 \leq \omega_i \leq 1$, para $i = 1, \dots, M$ e $L_i(\boldsymbol{\theta})$ é dada por (2.9). Podemos notar que, para $\omega_i = 0$ e $\omega_j = 1$ para $j \neq i$ excluimos o i -ésimo indivíduo da expressão do logaritmo da verossimilhança, e que o vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^T \in \mathbb{R}^M$. Mediante diferenciação, obtemos

$$\begin{aligned} \left. \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{q}_i \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}), \\ \left. \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\alpha_r \partial\omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{2} \{ \text{tr} \hat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) - \hat{q}_i \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i \} \quad \text{e} \\ \left. \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\lambda_s \partial\omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{2} \{ \text{tr} \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T - \hat{q}_i \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i \}, \end{aligned}$$

para $r = 1, \dots, k$; $s = 1, \dots, d$ e $\hat{q}_i = q_i(\hat{\boldsymbol{\theta}})$, $i = 1, \dots, M$.

Esse esquema de perturbação permite identificar aqueles indivíduos que exercem um grande impacto no processo de estimação. No Apêndice B mostramos a equivalência entre a influência local sob esse esquema de perturbação e uma medida de influência por eliminação de casos. Um fato bem conhecido é que as medidas de influência global podem sofrer do fenômeno de mascarar observações influentes (Cook e Weisberg, 1982), assim como freqüentemente essas medidas de diagnóstico podem ser difíceis de serem

obtidas em modelos mais complexos. Contornamos tais problemas considerando o procedimento de influência local.

Para o modelo linear misto sob normalidade ($v_i(\boldsymbol{\theta}) = 1$) esse esquema de perturbação foi considerado por Lesaffre e Verbeke (1998) usando $\mathbf{V}_i = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \phi \mathbf{I}_{n_i}$, em que $\boldsymbol{\Psi}$ representa uma matriz não estruturada. Notamos que o modelo considerado nesta seção permite maior flexibilidade na estrutura de efeitos aleatórios para \mathbf{V}_i (veja Jennrich e Schluchter, 1986), assim como permite manipular membros da classe de distribuições elípticas fazendo $q(\boldsymbol{\theta}) = -2W_g(u)$.

Perturbação da Matriz de Escala

Este esquema de perturbação é introduzido mediante considerar o modelo $\mathbf{Y}_i \sim EC_{n_i}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i / \omega_i; g)$, para $i = 1, \dots, M$, em que $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$, $\omega_i > 0$, para $i = 1, \dots, M$ e $\boldsymbol{\omega}_0 = \mathbf{1}_M$. Neste caso o logaritmo da função de verossimilhança resulta em $L(\boldsymbol{\theta} | \boldsymbol{\omega}) = \sum_{i=1}^M L_i(\boldsymbol{\theta} | \boldsymbol{\omega})$, em que

$$L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) = -\frac{1}{2} \log |\mathbf{V}_i| + \frac{1}{2} n_i \log \omega_i + \log g(u_{i\omega})$$

com $u_{i\omega} = \omega_i u_i$ e $u_i = \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i$, $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, para $i = 1, \dots, M$. Diferenciando $L(\boldsymbol{\theta} | \boldsymbol{\omega})$ com relação a $\boldsymbol{\theta}$ e ω_i e avaliando em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, obtemos

$$\begin{aligned} \left. \frac{\partial^2 L(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \boldsymbol{\omega} = \boldsymbol{\omega}_0} &= -2 \{W_g(\hat{u}_i) + \hat{u}_i W'_g(\hat{u}_i)\} \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i, \\ \left. \frac{\partial^2 L(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \alpha_r \partial \omega_i} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \boldsymbol{\omega} = \boldsymbol{\omega}_0} &= -\{W_g(\hat{u}_i) + \hat{u}_i W'_g(\hat{u}_i)\} \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \dot{\Sigma}_i(r) \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i \quad \text{e} \\ \left. \frac{\partial^2 L(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \lambda_s \partial \omega_i} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \boldsymbol{\omega} = \boldsymbol{\omega}_0} &= -\{W_g(\hat{u}_i) + \hat{u}_i W'_g(\hat{u}_i)\} \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i, \end{aligned}$$

para $r = 1, \dots, k$; $s = 1, \dots, d$ em que $W_g(u)$ e $W'_g(u)$ são avaliadas em $\hat{u}_i = \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i$.

O interesse do esquema de perturbação da matriz de escala é que o mesmo pode revelar indivíduos que interferem principalmente na modelagem da estrutura de escala assim como indicar aqueles indivíduos influentes na estimação dos parâmetros $\boldsymbol{\tau} = (\boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$.

Perturbação das Variáveis Explicativas

Neste caso, o interesse é perturbar uma particular variável explicativa contínua $\mathbf{x}_{it}(\boldsymbol{\omega}) = \mathbf{x}_{it} + \boldsymbol{\omega}_i s_i$, em que $\mathbf{x}_{it} \in \mathbb{R}^{n_i}$ representa a t -ésima coluna da matriz \mathbf{X}_i para $t = 1, \dots, p$, $\boldsymbol{\omega}_i \in \mathbb{R}^{n_i}$ e s_i é um fator de escala. Deste modo $\mathbf{X}_{i\omega} = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{it} + \boldsymbol{\omega}_i s_i, \dots, \mathbf{x}_{ip}) = \mathbf{X}_i + s_i \boldsymbol{\omega}_i \mathbf{c}_t^T$, para $i = 1, \dots, M$, em que \mathbf{c}_t denota um vetor $p \times 1$ com um 1 na t -ésima posição e zeros nas posições restantes, aqui $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^N$ e $N = \sum_{i=1}^M n_i$. Temos que $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^M L_i(\boldsymbol{\theta}|\boldsymbol{\omega})$, com

$$L_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{1}{2} \log |\mathbf{V}_i| + \log g(u_{i\omega}),$$

em que $u_{i\omega} = \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} \mathbf{r}_{i\omega}$, $\mathbf{r}_{i\omega} = \mathbf{r}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$. Usando o método de diferenciação matricial obtemos

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -s_i 2W_g(\hat{u}_i) (\hat{\boldsymbol{\beta}}_t \mathbf{X}_i^T - \mathbf{c}_t \hat{\mathbf{r}}_i^T) \hat{\mathbf{V}}_i^{-1} \\ &\quad + s_i 4W'_g(\hat{u}_i) \hat{\boldsymbol{\beta}}_t \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1}, \\ \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= 2s_i \hat{\boldsymbol{\beta}}_t \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \hat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \hat{\mathbf{V}}_i + W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \hat{\mathbf{V}}_i^{-1} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \lambda_s \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= 2s_i \hat{\boldsymbol{\beta}}_t \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\Psi}(s) \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \hat{\mathbf{V}}_i + W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \hat{\mathbf{V}}_i^{-1}, \end{aligned}$$

para $r = 1, \dots, k$; $s = 1, \dots, d$ e $i = 1, \dots, M$. Neste caso $\hat{\boldsymbol{\beta}}_t$ denota o t -ésimo elemento de $\hat{\boldsymbol{\beta}}$.

Esse esquema de perturbação permite, por exemplo, detectar possíveis maus condicionamentos entre algumas colunas da matriz de desenho \mathbf{X}_i (Belsley, 1991). De fato esse esquema pode ser estendido se o interesse é perturbar várias colunas da matriz de planejamento, assim como perturbar toda a matriz mediante substituir \mathbf{X}_i por $\mathbf{X}_{i\omega} = \mathbf{X}_i + \mathbf{W}_i \mathbf{S}_i$, $i = 1, \dots, M$, em que $\mathbf{W}_i = (\omega_{ij})$ é matriz de perturbações $n_i \times p$, $\mathbf{S}_i = \text{diag}(s_{i1}, \dots, s_{ip})$ e s_{ij} para $j = 1, \dots, p$ representa um fator de escala. O modelo perturbado reduz ao modelo postulado quando $\mathbf{W}_0 = \mathbf{0}$.

Perturbação da Resposta

A perturbação da resposta observada $(\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ é introduzida mediante a substituição de \mathbf{Y}_i por $\mathbf{Y}_{i\omega} = \mathbf{Y}_i + \boldsymbol{\omega}_i$, em que $\boldsymbol{\omega}_i$ é um vetor de

perturbações $n_i \times 1$ para $i = 1, \dots, M$. Neste caso temos que $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^N$ com $N = \sum_{i=1}^M n_i$. O logaritmo da verossimilhança perturbada fica dado por $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^M L_i(\boldsymbol{\theta}|\boldsymbol{\omega})$, com

$$L_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{1}{2} \log |\mathbf{V}_i| + \log g(u_{i\boldsymbol{\omega}}),$$

em que $u_{i\boldsymbol{\omega}} = \mathbf{r}_{i\boldsymbol{\omega}}^T \mathbf{V}_i^{-1} \mathbf{r}_{i\boldsymbol{\omega}}$, $\mathbf{r}_{i\boldsymbol{\omega}} = \mathbf{r}_i + \boldsymbol{\omega}_i$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$. Deste modo,

$$\begin{aligned} \left. \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2 \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \hat{\mathbf{V}}_i + 2W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \hat{\mathbf{V}}_i^{-1}, \\ \left. \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2 \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \hat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \hat{\mathbf{V}}_i + W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \hat{\mathbf{V}}_i^{-1} \quad \text{e} \\ \left. \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \lambda_s \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2 \hat{\mathbf{r}}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\Psi}(s) \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \hat{\mathbf{V}}_i + W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \hat{\mathbf{V}}_i^{-1}, \end{aligned}$$

para $r = 1, \dots, k$; $s = 1, \dots, d$ e $i = 1, \dots, M$.

Uma característica de interesse deste esquema de perturbação é a sua conexão com a matriz de alavancas generalizadas (Wei et al., 1998). Para o modelo elíptico linear misto marginal examinamos esta relação na seção a seguir.

2.5 Método de Alavanca Generalizado no Modelo Linear Misto Marginal

O conceito de ponto de alavanca freqüentemente tem sido associado com influência local usando perturbações aditivas na resposta observada. A conexão entre o método de alavanca generalizado e o procedimento de influência local para modelos de regressão não-linear usando distribuições da família exponencial tem sido examinado por Wei et al. (1998) e para modelos simétricos univariados por Galea et al. (2005). A seguir derivamos a matriz de alavancas generalizadas para o modelo linear misto marginal usando os resultados dados em Wei et al. (1998).

Considere $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$, deste modo podemos escrever o vetor de posição $\boldsymbol{\mu}$ como $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_M^T)^T$. Note que

$$\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^T} = \mathbf{X}_i, \quad \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\alpha}^T} = \mathbf{0} \quad \text{e} \quad \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\lambda}^T} = \mathbf{0}, \quad i = 1, \dots, M.$$

Sejam \mathbf{D} e $\mathbf{\Delta}$ matrizes $N \times p^*$ e $p^* \times N$, respectivamente, em que $N = \sum_{i=1}^M n_i$ representa a dimensão do vetor $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ e $p^* = p+k+d$, dadas por

$$\mathbf{D} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T} = \begin{pmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_M \end{pmatrix} \quad \text{e} \quad \mathbf{\Delta} = \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \mathbf{Y}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(Y)} = \begin{pmatrix} \mathbf{\Delta}_1 \\ \mathbf{\Delta}_2 \\ \mathbf{\Delta}_3 \end{pmatrix}, \quad (2.11)$$

em que $\mathbf{\Delta}_1 \in \mathbb{R}^{p \times N}$, $\mathbf{\Delta}_2 \in \mathbb{R}^{k \times N}$ e $\mathbf{\Delta}_3 \in \mathbb{R}^{d \times N}$ assumem formas particionadas como $\mathbf{\Delta}_1 = (\mathbf{\Delta}_{11}, \dots, \mathbf{\Delta}_{1M})$, $\mathbf{\Delta}_2 = (\mathbf{\Delta}_{21}, \dots, \mathbf{\Delta}_{2M})$ e $\mathbf{\Delta}_3 = (\mathbf{\Delta}_{31}, \dots, \mathbf{\Delta}_{3M})$ com $\mathbf{\Delta}_{1i} \in \mathbb{R}^{p \times n_i}$, $\mathbf{\Delta}_{2i} \in \mathbb{R}^{k \times n_i}$ e $\mathbf{\Delta}_{3i} \in \mathbb{R}^{d \times n_i}$ para $i = 1, \dots, M$. Temos que $\partial \boldsymbol{\mu}_i / \partial \boldsymbol{\theta}^T = \mathbf{D}_i$, com

$$\mathbf{D}_i = (\mathbf{X}_i, \mathbf{0}, \mathbf{0}) \in \mathbb{R}^{n_i \times p^*}, \quad i = 1, \dots, M.$$

Através do método de diferenciação (veja Apêndice A.2) obtemos

$$\mathbf{\Delta}_{1i} = -2\mathbf{X}_i^T \widehat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \widehat{\mathbf{V}}_i + 2W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \widehat{\mathbf{V}}_i^{-1}, \quad i = 1, \dots, M,$$

a r -ésima linha de $\mathbf{\Delta}_{2i}$ fica dada por

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \alpha_r \partial \mathbf{Y}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -2\hat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \dot{\Sigma}_i(r) \widehat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \widehat{\mathbf{V}}_i + W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \widehat{\mathbf{V}}_i^{-1},$$

para $r = 1, \dots, k$ e a s -ésima linha de $\mathbf{\Delta}_{3i}$ assume a forma

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \lambda_s \partial \mathbf{Y}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -2\hat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\Psi}(s) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \{W_g(\hat{u}_i) \widehat{\mathbf{V}}_i + W'_g(\hat{u}_i) \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T\} \widehat{\mathbf{V}}_i^{-1},$$

para $s = 1, \dots, d$ e $i = 1, \dots, M$.

Deste modo a matriz de alavancas generalizadas para o modelo linear misto marginal fica dada por

$$\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \mathbf{D} \{-\ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}})\}^{-1} \mathbf{\Delta}, \quad (2.12)$$

com $\ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}}) = \partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(Y)}$ em que $\ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M \ddot{\mathbf{L}}_i(\hat{\boldsymbol{\theta}})$, $\ddot{\mathbf{L}}_i(\hat{\boldsymbol{\theta}})$ tem forma particionada dada em (2.10), \mathbf{D} e $\mathbf{\Delta}$ são definidas em (2.11). Considere $\mathbf{R} = \{-\ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}})\}^{-1}$ e \mathbf{X} particionadas como

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{R}_{12}^T & \mathbf{R}_{22} & \mathbf{R}_{23} \\ \mathbf{R}_{13}^T & \mathbf{R}_{23}^T & \mathbf{R}_{33} \end{pmatrix} \quad \text{e} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_M \end{pmatrix}, \quad (2.13)$$

deste modo a matriz \mathbf{D} em (2.11) assume a forma $\mathbf{D} = (\mathbf{X}, \mathbf{0}, \mathbf{0})$ e a matriz de alavancas generalizadas fica dada por

$$\begin{aligned} \mathbf{GL}(\hat{\boldsymbol{\theta}}) &= (\mathbf{X}, \mathbf{0}, \mathbf{0}) \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{R}_{12}^T & \mathbf{R}_{22} & \mathbf{R}_{23} \\ \mathbf{R}_{13}^T & \mathbf{R}_{23}^T & \mathbf{R}_{33} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 \\ \boldsymbol{\Delta}_3 \end{pmatrix} \\ &= \mathbf{GL}_\beta(\hat{\boldsymbol{\theta}}) + \mathbf{GL}_\alpha(\hat{\boldsymbol{\theta}}) + \mathbf{GL}_\lambda(\hat{\boldsymbol{\theta}}), \end{aligned}$$

em que

$$\begin{aligned} \mathbf{GL}_\beta(\hat{\boldsymbol{\theta}}) &= \mathbf{X}\mathbf{R}_{11}\boldsymbol{\Delta}_1, & \mathbf{GL}_\alpha(\hat{\boldsymbol{\theta}}) &= \mathbf{X}\mathbf{R}_{12}\boldsymbol{\Delta}_2 & \text{e} \\ \mathbf{GL}_\lambda(\hat{\boldsymbol{\theta}}) &= \mathbf{X}\mathbf{R}_{13}\boldsymbol{\Delta}_3. \end{aligned}$$

As quantidades $\mathbf{GL}_\alpha(\hat{\boldsymbol{\theta}})$ e $\mathbf{GL}_\lambda(\hat{\boldsymbol{\theta}})$ podem ser interpretadas como termos de correção da matriz de alavancas generalizadas devido ao procedimento de estimação.

Note que os blocos diagonais de $\mathbf{GL}(\hat{\boldsymbol{\theta}})$ são dados por

$$\mathbf{GL}_i(\hat{\boldsymbol{\theta}}) = \mathbf{D}_i \left\{ -\sum_{i=1}^M \ddot{\mathbf{L}}_i(\hat{\boldsymbol{\theta}}) \right\}^{-1} \boldsymbol{\Delta}_i, \quad \boldsymbol{\Delta}_i = \begin{pmatrix} \boldsymbol{\Delta}_{1i} \\ \boldsymbol{\Delta}_{2i} \\ \boldsymbol{\Delta}_{3i} \end{pmatrix}, \quad (2.14)$$

para $i = 1, \dots, M$. Logo, esta expressão pode ser considerada uma generalização do resultado apresentado em Banerjee e Frees (1997). Com efeito, para o caso normal e quando somente temos interesse no vetor de parâmetros $\boldsymbol{\beta}$, a matriz de alavancas generalizadas para cada indivíduo fica dada por

$$\mathbf{GL}_i(\hat{\boldsymbol{\beta}}) = \mathbf{X}_i \left(\sum_{i=1}^M \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1}, \quad i = 1, \dots, M, \quad (2.15)$$

que tem uma clara relação com as matrizes de alavancas definidas para modelos de regressão linear (veja Hoaglin e Welsh, 1978 e Galea et al., 2005, dentre outros), assim como para modelos longitudinais (Banerjee e Frees, 1997).

Por outro lado (2.14) permite focar a nossa atenção nas matrizes de alavancas associadas a cada indivíduo ($i = 1, \dots, M$) e deste modo podem resultar em instrumentos úteis para diagnóstico no modelo elíptico linear com efeitos mistos marginal.

2.5.1 Conexão entre o Método de Alavanca Generalizado e o Procedimento de Influência Local

Considere

$$\mathbf{D}_\alpha = \begin{pmatrix} \mathbf{D}_{1\alpha} \\ \vdots \\ \mathbf{D}_{M\alpha} \end{pmatrix} \quad \text{e} \quad \mathbf{D}_\lambda = \begin{pmatrix} \mathbf{D}_{1\lambda} \\ \vdots \\ \mathbf{D}_{M\lambda} \end{pmatrix},$$

em que $\mathbf{D}_{i\alpha}$ tem r -ésima coluna dada por $\dot{\Sigma}_i(r)\widehat{\mathbf{V}}_i^{-1}\widehat{\mathbf{r}}_i$, $r = 1, \dots, k$ e a s -ésima coluna de $\mathbf{D}_{i\lambda}$ é dada por $\mathbf{Z}_i\dot{\Psi}(s)\mathbf{Z}_i^T\widehat{\mathbf{V}}_i^{-1}\widehat{\mathbf{r}}_i$, $s = 1, \dots, d$ para $i = 1, \dots, M$. Então, a matriz $\mathbf{\Delta}$ do esquema de perturbação aditiva no vetor de respostas fica dada por

$$\mathbf{\Delta} = \mathbf{D}^{*T} \mathbf{A},$$

em que $\mathbf{D}^* = (\mathbf{X}, \mathbf{D}_\alpha, \mathbf{D}_\lambda) \in \mathbb{R}^{N \times p^*}$ e $\mathbf{A} = \text{blc diag}(\mathbf{A}_1, \dots, \mathbf{A}_M)$ com $\mathbf{A}_i = -2\widehat{\mathbf{V}}_i^{-1}\{W_g(\widehat{u}_i)\widehat{\mathbf{V}}_i + 2W'_g(\widehat{u}_i)\widehat{\mathbf{r}}_i\widehat{\mathbf{r}}_i^T\}\widehat{\mathbf{V}}_i^{-1}$, para $i = 1, \dots, M$.

A curvatura normal para a perturbação aditiva do vetor de respostas observadas pode ser expressa como

$$\mathbf{B} = -\mathbf{\Delta}^T\{\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})\}^{-1}\mathbf{\Delta} = \mathbf{A}\mathbf{D}^*\{-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})\}^{-1}\mathbf{D}^{*T}\mathbf{A}.$$

Suponha agora que $\boldsymbol{\alpha}$ e $\boldsymbol{\lambda}$ são fixados e considere o esquema de perturbação da resposta $\mathbf{Y}_{i\omega} = \mathbf{Y}_i + \boldsymbol{\omega}_i$ em que $\boldsymbol{\omega}_i \in \mathbb{R}^{n_i}$ para $i = 1, \dots, M$, então temos que a matriz de alavancas generalizadas assume a forma

$$\mathbf{GL}(\widehat{\boldsymbol{\theta}}) = \mathbf{X}\{-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\beta}})\}^{-1}\mathbf{X}^T\mathbf{A} \quad (2.16)$$

com \mathbf{X} definido em (2.13) e $\ddot{\mathbf{L}}(\widehat{\boldsymbol{\beta}})$ dado por $\ddot{\mathbf{L}}(\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^M \ddot{\mathbf{L}}_i(\widehat{\boldsymbol{\beta}})$, em que

$$\ddot{\mathbf{L}}_i(\widehat{\boldsymbol{\beta}}) = -\mathbf{X}_i^T\mathbf{A}_i\mathbf{X}_i, \quad i = 1, \dots, M.$$

Note que os blocos diagonais de $\mathbf{GL}(\widehat{\boldsymbol{\theta}})$ são dados por

$$\mathbf{GL}_i(\widehat{\boldsymbol{\theta}}) = \mathbf{X}_i\left(\sum_{i=1}^M \mathbf{X}_i^T\mathbf{A}_i\mathbf{X}_i\right)^{-1}\mathbf{X}_i^T\mathbf{A}_i, \quad i = 1, \dots, M,$$

que para o caso normal ($W_g(u) = -\frac{1}{2}$ e $W'_g(u) = 0$) reduz-se a (2.15).

Por outro lado, temos que a curvatura normal para esse esquema de perturbação fica dada por

$$\mathbf{B} = -\mathbf{\Delta}_1^T\{\ddot{\mathbf{L}}(\widehat{\boldsymbol{\beta}})\}^{-1}\mathbf{\Delta}_1.$$

Como $\mathbf{\Delta}_1 = \mathbf{X}^T \mathbf{A}$, temos que a relação entre a curvatura normal e a matriz de alavancas generalizadas pode ser escrita como

$$\mathbf{B} = \mathbf{A} \mathbf{X} \left(\sum_{i=1}^M \mathbf{X}_i^T \mathbf{A}_i \mathbf{X}_i \right)^{-1} \mathbf{X}^T \mathbf{A} = \mathbf{A} \mathbf{G} \mathbf{L}(\hat{\boldsymbol{\theta}}),$$

com $\mathbf{G} \mathbf{L}(\hat{\boldsymbol{\theta}})$ dada em (2.16).

A ligação entre o método de influência local e a matriz de alavancas generalizadas é interessante pois oferece uma justificativa para o esquema de perturbação da resposta e permite estudar a influência que respostas aberrantes dentro de cada indivíduo exercem sobre os valores preditos.

2.6 Influência Local no Modelo Linear Misto Hierárquico

Nesta seção derivamos a curvatura normal para o modelo linear com efeitos mistos na sua formulação hierárquica usando a classe das misturas de escala normal quando ambos \mathbf{b}_i e v_i são considerados não observáveis e, posteriormente, para o modelo em que os efeitos aleatórios são integrados, isto é, quando somente as variáveis de mistura v_i são não observáveis. Usamos como medida de influência a função Q -afastamento que é baseada na esperança condicional do logaritmo da verossimilhança de dados completos proveniente do algoritmo EM.

2.6.1 Influência Local quando \mathbf{b}_i e v_i são não observáveis

Considere a formulação hierárquica dada por (1.4) em que $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{b}^T, \mathbf{v}^T)^T$ representa o vetor de dados completos, com $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ o vetor de respostas observadas em que os efeitos aleatórios $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_M^T)^T$ e as variáveis de mistura $\mathbf{v} = (v_1, \dots, v_M)^T$ são não observáveis. Então, o logaritmo da função de verossimilhança para os dados completos \mathbf{Y}_c é dado por $L(\boldsymbol{\theta} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta} | \mathbf{Y}_c)$, com

$$\begin{aligned} L_i(\boldsymbol{\theta} | \mathbf{Y}_c) = & -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{\kappa^{-1}(v_i)}{2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i) \\ & - \frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c, \end{aligned} \quad (2.17)$$

em que $h(v_i; \boldsymbol{\nu})$ é a função de densidade da variável de mistura e c é uma constante.

A partir de (2.17) temos que a Q -função para o modelo linear misto hierárquico quando \mathbf{b}_i e v_i são não observáveis assume a forma $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, com $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) + Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$, para $i = 1, \dots, M$, em que $Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$ são dadas por (1.15) e (1.16), respectivamente.

A seguir obtemos a curvatura normal $C_{f_{Q,h}}$ para o modelo linear misto sob a formulação hierárquica dada em (1.4). Os detalhes do cálculo da hessiana \ddot{Q} e da matriz Δ para diferentes esquemas de perturbação são dados no Apêndice A.3.1.

Matriz Hessiana, \ddot{Q}

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ o vetor de parâmetros de interesse. A matriz hessiana $\ddot{Q}(\boldsymbol{\theta})$ avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ é dada por $\ddot{Q}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M \ddot{Q}_i(\hat{\boldsymbol{\theta}})$ com

$$\ddot{Q}_i(\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \ddot{Q}_i(\hat{\boldsymbol{\eta}}) & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_i(\hat{\boldsymbol{\lambda}}) \end{pmatrix}, \quad (2.18)$$

em que $\ddot{Q}_i(\hat{\boldsymbol{\eta}}) = \partial^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ com $\boldsymbol{\eta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$, e $\ddot{Q}_i(\hat{\boldsymbol{\lambda}}) = \partial^2 Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. Temos que $\ddot{Q}_i(\hat{\boldsymbol{\eta}})$ pode ser escrito em forma particionada como

$$\ddot{Q}_i(\hat{\boldsymbol{\eta}}) = \begin{pmatrix} \ddot{Q}_{11,i} & \ddot{Q}_{12,i} \\ \ddot{Q}_{12,i}^T & \ddot{Q}_{22,i} \end{pmatrix},$$

em que

$$\begin{aligned} \ddot{Q}_{11,i} &= \frac{\partial^2 Q_i(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\hat{\kappa}_i \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i, \\ \ddot{Q}_{12,i} &= \frac{\partial^2 Q_i(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{e} \quad \ddot{Q}_{22,i} = \frac{\partial^2 Q_i(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \end{aligned}$$

com

$$\begin{aligned} \frac{\partial^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \alpha_r} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= -\hat{\kappa}_i \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \hat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i) \quad \text{e} \\ \frac{\partial^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})}{\partial \alpha_r \partial \alpha_s} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \hat{\boldsymbol{\Sigma}}_i^{-1} \{ \dot{\boldsymbol{\Sigma}}_i(r) \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) - \ddot{\boldsymbol{\Sigma}}_i(r, s) \} \\ &\quad - \frac{1}{2} \text{tr} \hat{\boldsymbol{\Sigma}}_i^{-1} \{ \dot{\boldsymbol{\Sigma}}_i(r) \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) + \dot{\boldsymbol{\Sigma}}_i(s) \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) - \ddot{\boldsymbol{\Sigma}}_i(r, s) \} \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{M}}_i, \end{aligned}$$

para $r, s = 1, \dots, k$ e o elemento (t, u) de $\ddot{Q}_i(\hat{\lambda})$ fica dado por

$$\begin{aligned} \frac{\partial^2 Q_{2i}(\lambda|\hat{\theta})}{\partial \lambda_t \partial \lambda_u} \Big|_{\theta=\hat{\theta}} &= \frac{1}{2} \text{tr} \hat{\Psi}^{-1} \{ \dot{\Psi}(t) \hat{\Psi}^{-1} \dot{\Psi}(u) - \ddot{\Psi}(t, u) \} \\ &\quad - \frac{1}{2} \text{tr} \hat{\Psi}^{-1} \{ \dot{\Psi}(t) \hat{\Psi}^{-1} \dot{\Psi}(u) + \dot{\Psi}(u) \hat{\Psi}^{-1} \dot{\Psi}(t) - \ddot{\Psi}(t, u) \} \hat{\Psi}^{-1} \hat{N}_i, \end{aligned}$$

para $t, u = 1, \dots, d$, em que $\hat{M}_i = \hat{\kappa}_i \mathbf{e}_i \mathbf{e}_i^T + \mathbf{Z}_i \hat{\Omega}_i \mathbf{Z}_i^T$, $\hat{N}_i = \hat{\kappa}_i \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + \hat{\Omega}_i$ e $\mathbf{e}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{Z}_i \hat{\mathbf{b}}_i$, para $i = 1, \dots, M$, com $\hat{\kappa}_i$, $\hat{\mathbf{b}}_i$ e $\hat{\Omega}_i$ dados em (1.17) e (1.18), também $\dot{\Sigma}_i(r) = \partial \Sigma_i / \partial \alpha_r$, $\ddot{\Sigma}_i(r, s) = \partial^2 \Sigma_i / \partial \alpha_r \partial \alpha_s$, $\dot{\Psi}(t) = \partial \Psi / \partial \lambda_t$ e $\ddot{\Psi}(t, u) = \partial^2 \Psi / \partial \lambda_t \partial \lambda_u$. Essas expressões são avaliadas em $\theta = \hat{\theta}$.

Esquemas de Perturbação

A seguir obtemos a matriz Δ considerando vários esquemas de perturbação para o modelo linear misto hierárquico definido por (1.4). Consideramos os seguintes esquemas de perturbação: de ponderação de casos, da matriz de escala dos efeitos aleatórios, das variáveis explicativas, da resposta observada e dois esquemas de perturbação da matriz de escala intra-indivíduo.

A matriz Δ para cada esquema de perturbação tem forma particionada

$$\Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \end{pmatrix},$$

em que $\Delta_1 = \partial^2 Q(\theta, \omega|\hat{\theta}) / \partial \beta \partial \omega^T \in \mathbb{R}^{p \times q}$, $\Delta_2 = \partial^2 Q(\theta, \omega|\hat{\theta}) / \partial \alpha \partial \omega^T \in \mathbb{R}^{k \times q}$ e $\Delta_3 = \partial^2 Q(\theta, \omega|\hat{\theta}) / \partial \lambda \partial \omega^T \in \mathbb{R}^{d \times q}$ com q sendo a dimensão do vetor de perturbação ω .

Para o cálculo da matriz Δ assumimos que é possível trocar as operações de integração e diferenciação. Para cada um dos esquemas de perturbação apresentamos $\partial^2 L(\theta, \omega | \mathbf{Y}_c) / \partial \theta \partial \omega^T$, os detalhes dos diferenciais $d^2 L(\theta, \omega | \mathbf{Y}_c)$ são dados no Apêndice A.3.1.

Ponderação de Casos

Para este esquema de perturbação estamos interessados na inclusão de pesos para cada indivíduo no logaritmo da função de verossimilhança de dados completos como

$$L(\theta, \omega | \mathbf{Y}_c) = \sum_{i=1}^M \omega_i L_i(\theta | \mathbf{Y}_c),$$

em que $L_i(\boldsymbol{\theta}|\mathbf{Y}_c)$ é dada por (2.17), $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$ representa o vetor de perturbação com $0 \leq \omega_i \leq 1$ para $i = 1, \dots, M$ e $\boldsymbol{\omega}_0 = \mathbf{1}_M$. Neste caso temos que,

$$\begin{aligned}\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_i} &= \kappa^{-1}(v_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i), \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c)}{\partial \alpha_r \partial \omega_i} &= -\frac{1}{2} \{ \text{tr} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) - \kappa^{-1}(v_i) \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i \} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c)}{\partial \lambda_s \partial \omega_i} &= -\frac{1}{2} \{ \text{tr} \boldsymbol{\Psi}^{-1} \dot{\boldsymbol{\Psi}}(s) - \kappa^{-1}(v_i) \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \dot{\boldsymbol{\Psi}}(s) \boldsymbol{\Psi}^{-1} \mathbf{b}_i \}\end{aligned}$$

para $r = 1, \dots, k$, $s = 1, \dots, d$ e $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, $i = 1, \dots, M$.

Observamos que quando $\omega_i = 0$ e $\omega_j = 1$ para $j = 1, \dots, M$, $j \neq i$, o i -ésimo indivíduo é excluído do logaritmo da verossimilhança de dados completos e também que a influência local para esse esquema de perturbação é equivalente ao diagnóstico por eliminação de casos (veja Apêndice B).

Perturbação da Matriz de Escala dos Efeitos Aleatórios

Uma suposição habitual no modelo linear é que a matriz de escala dos efeitos aleatórios é dada por $\boldsymbol{\Psi}(\boldsymbol{\lambda})$. Para estudar desvios a partir dessa suposição assim como uma má especificação de $\boldsymbol{\Psi}(\boldsymbol{\lambda})$, consideramos que a matriz de escala dos efeitos aleatórios é $\omega_i^{-1} \boldsymbol{\Psi}(\boldsymbol{\lambda})$, com $\omega_i > 0$ para $i = 1, \dots, M$. Aqui $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$, $\boldsymbol{\omega}_0 = \mathbf{1}_M$ e o logaritmo da função de verossimilhança de dados completos do modelo perturbado fica dado por $L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$, em que

$$\begin{aligned}L_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c) &= -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{\kappa^{-1}(v_i)}{2} \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i - \frac{q}{2} \log \omega_i \\ &\quad - \frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{\kappa^{-1}(v_i) \omega_i}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c,\end{aligned}$$

com $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, $i = 1, \dots, M$. Diferenciando $L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e ω_i , obtemos

$$\begin{aligned}\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_i} &= \mathbf{0}, & \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c)}{\partial \boldsymbol{\alpha} \partial \omega_i} &= \mathbf{0} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y}_c)}{\partial \lambda_s \partial \omega_i} &= \frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \dot{\boldsymbol{\Psi}}(s) \boldsymbol{\Psi}^{-1} \mathbf{b}_i,\end{aligned}$$

para $s = 1, \dots, d$.

Perturbação das Variáveis Explicativas

Podemos investigar o mau condicionamento entre algumas colunas da matriz de planejamento \mathbf{X}_i , por exemplo, perturbando uma particular variável explicativa contínua $\mathbf{x}_{it}(\boldsymbol{\omega}) = \mathbf{x}_{it} + \boldsymbol{\omega}_i s_i$ em que $\mathbf{x}_{it} \in \mathbb{R}^{n_i}$ representa a t -ésima coluna da matriz \mathbf{X}_i , $\boldsymbol{\omega}_i \in \mathbb{R}^{n_i}$ e s_i é um fator de escala, aqui $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^N$ com $N = \sum_{i=1}^M n_i$. Temos que $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$, com

$$L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{\kappa^{-1}(v_i)}{2} (\boldsymbol{\epsilon}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\epsilon}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i) \\ - \frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c.$$

Neste caso a matriz de planejamento perturbada para o i -ésimo indivíduo fica dada por $\mathbf{X}_{i\omega} = \mathbf{X}_i + s_i \boldsymbol{\omega}_i \mathbf{c}_t^T$, para $i = 1, \dots, M$, em que \mathbf{c}_t é um vetor $p \times 1$ com 1 na t -ésima posição e zeros nas posições restantes. Portanto,

$$\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} = -\kappa^{-1}(v_i) s_i (\beta_t \mathbf{X}_i^T - \mathbf{c}_t \boldsymbol{\epsilon}_i^T) \boldsymbol{\Sigma}_i^{-1}, \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} = -\kappa^{-1}(v_i) s_i \beta_t \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \boldsymbol{\Sigma}_i^{-1} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \lambda \partial \boldsymbol{\omega}_i^T} = \mathbf{0},$$

para $r = 1, \dots, k$ e $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, $i = 1, \dots, M$. Aqui β_t denota o t -ésimo elemento de $\boldsymbol{\beta}$.

Podemos notar que esse esquema de perturbação pode ser facilmente generalizado para perturbar todas as colunas da matriz de planejamento \mathbf{X}_i como $\mathbf{X}_{i\omega} = \mathbf{X}_i + \mathbf{W}_i \mathbf{S}_i$ em que \mathbf{W}_i é matriz de perturbação $n_i \times p$, $\mathbf{S}_i = \text{diag}(s_{i1}, \dots, s_{ip})$ com s_{ij} , para $j = 1, \dots, p$, sendo um fator de escala e $\mathbf{W}_0 = \mathbf{0}$.

Perturbação da Resposta

Considere uma perturbação aditiva para o vetor de respostas observadas $(\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$, isto é, adicionamos uma pequena perturbação às respostas de cada indivíduo, obtendo $\mathbf{Y}_{i\omega} = \mathbf{Y}_i + \boldsymbol{\omega}_i$ em que $\boldsymbol{\omega}_i$ representa um vetor de perturbação n_i -dimensional, $i = 1, \dots, M$. Nesse caso o vetor de não perturbação fica dado por $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^N$ com $N = \sum_{i=1}^M n_i$ e o logaritmo da verossimilhança de dados completos assume a forma $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) =$

$\sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$, com

$$L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{\kappa^{-1}(v_i)}{2} (\boldsymbol{\epsilon}_i + \boldsymbol{\omega}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\epsilon}_i + \boldsymbol{\omega}_i) \\ - \frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c.$$

É possível mostrar que,

$$\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} = \kappa^{-1}(v_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1}, \quad \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}_i^T} = \mathbf{0} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} = \kappa^{-1}(v_i) \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \boldsymbol{\Sigma}_i^{-1}.$$

para $r = 1, \dots, k$ e $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, $i = 1, \dots, M$.

Na seção a seguir abordamos a relação entre o esquema de perturbação aditivo no vetor de respostas e uma versão da matriz de alavancas generalizadas para o modelo linear misto hierárquico quando ambos \mathbf{b}_i e v_i são não observáveis.

Perturbação da Estrutura de Escala dentro dos Indivíduos

Para os esquemas de perturbação a serem discutidos a seguir vamos supor que a matriz de escala $\boldsymbol{\Sigma}_i$ pode ser decomposta como

$$\boldsymbol{\Sigma}_i = \mathbf{D}_i \mathbf{C}_i \mathbf{D}_i, \quad i = 1, \dots, M, \quad (2.19)$$

em que \mathbf{D}_i é matriz diagonal com elementos diagonais positivos e \mathbf{C}_i é matriz de correlação. Seja $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T)^T$ em que $\boldsymbol{\alpha}_1$ e $\boldsymbol{\alpha}_2$ representam vetores de parâmetros k_1 e k_2 dimensionais ($k = k_1 + k_2$) associados a \mathbf{D}_i e \mathbf{C}_i , respectivamente, isto é, $\mathbf{D}_i = \mathbf{D}_i(\boldsymbol{\alpha}_1)$ e $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\alpha}_2)$. Notamos que a decomposição dada por (2.19) traz maior flexibilidade para a modelagem da estrutura de escala. Alguns exemplos de funções de escala úteis para a modelagem de \mathbf{D}_i assim como estruturas de correlação são discutidas em Pinheiro e Bates (2000). Aqui usamos a decomposição de escala-correlação dada em (2.19) para introduzir dois esquemas de perturbação na estrutura de escala *dentro* dos indivíduos ($i = 1, \dots, M$).

Para esses dois esquemas de perturbação intra-indivíduos é conveniente introduzir a seguinte notação. Seja \mathbf{A} matriz $m \times n$ particionada como

$\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ em que \mathbf{a}_j é a j -ésima coluna de \mathbf{A} , então $\text{vec}(\mathbf{A})$ corresponde ao operador que ‘transforma’ \mathbf{A} no vetor mn -dimensional

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

Algumas propriedades e relações entre o produto Kroneker, o operador de vetorização e o traço de matrizes, denotados neste trabalho como \otimes , $\text{vec}(\cdot)$ e $\text{tr}(\cdot)$, respectivamente, são dadas, por exemplo, em Magnus e Neudecker (1988).

Para $\mathbf{A} \in \mathbb{R}^{m \times m}$ define-se $\text{vech}(\mathbf{A}) \in \mathbb{R}^{m(m+1)/2}$ como o vetor obtido a partir da vetorização dos elementos acima da diagonal de \mathbf{A} . $\text{vech}(\cdot)$ é conhecido como o operador de vetorização simétrico pois para $\mathbf{A} = \mathbf{A}^T$ temos que $\text{vech}(\mathbf{A})$ contém somente os elementos distintos de \mathbf{A} .

Considere \mathbf{A} matriz simétrica $m \times m$. Então, a relação entre $\text{vech}(\mathbf{A})$ e $\text{vec}(\mathbf{A})$ é dada por

$$\text{vec}(\mathbf{A}) = \mathbf{S}_m \text{vech}(\mathbf{A}),$$

em que $\mathbf{S}_m \in \mathbb{R}^{m^2 \times m(m+1)/2}$ é chamada matriz de duplicação de ordem m .

Em geral é possível vetorizar matrizes que seguem algum formato particular (Nel, 1980). Essa operação, tipicamente denotada por $\text{vecp}(\cdot)$, corresponde à vetorização da matriz em questão uma vez que têm sido excluídos todos seus elementos constantes assim como aqueles elementos repetidos.

Note por exemplo que, para $\mathbf{A} = \mathbf{A}^T$ temos $\text{vecp}(\mathbf{A}) = \text{vech}(\mathbf{A})$. A seguir particularizamos esse operador para matrizes diagonais. Seja $\mathbf{a} \in \mathbb{R}^m$, então denotamos por $\text{diag}(\mathbf{a})$ a matriz diagonal cujos elementos diagonais são dados pelos componentes do vetor \mathbf{a} . Vetorizando os elementos diferentes de $\mathbf{A} = \text{diag}(\mathbf{a})$, obtemos $\text{vecp}(\mathbf{A}) = \mathbf{a}$, ou seja, nesse caso $\text{vecp}(\cdot)$ denota o operador que vetoriza os elementos da diagonal de \mathbf{A} .

Temos que existe uma matriz de duplicação \mathbf{S}_m^* de dimensão $m^2 \times m$ (Nel, 1980) que permite relacionar $\text{vecp}(\mathbf{A})$ e $\text{vec}(\mathbf{A})$ para matrizes diagonais como

$$\text{vec}(\mathbf{A}) = \mathbf{S}_m^* \text{vecp}(\mathbf{A}) = \mathbf{S}_m^* \mathbf{a}.$$

Resultados úteis relativos às matrizes de duplicação são apresentados em Nel (1980) e Magnus e Neudecker (1988).

Perturbação da Matriz de Correlação dentro dos Indivíduos

Uma maneira de estudar como as observações dentro dos indivíduos influenciam a modelagem da sua estrutura de correlação é através da seguinte decomposição escala-correlação:

$$\Sigma_{i\omega} = \mathbf{D}_i \mathbf{C}_{i\omega} \mathbf{D}_i, \quad \text{com } \mathbf{C}_{i\omega} = \mathbf{C}_i + \mathbf{W}_i,$$

em que \mathbf{W}_i é matriz de perturbação simétrica $n_i \times n_i$ tal que $\mathbf{C}_{i\omega}$ é matriz de correlação. Considere $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^T, \dots, \boldsymbol{\omega}_M^T)^T$ em que $\boldsymbol{\omega}_i = \text{vech}(\mathbf{W}_i)$ para $i = 1, \dots, M$, então o vetor de não perturbação para esse esquema é $\boldsymbol{\omega}_0 = \mathbf{0}$ e o logaritmo da verossimilhança de dados completos assume a forma $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$, em que

$$\begin{aligned} L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} \log |\Sigma_i + \mathbf{D}_i \mathbf{W}_i \mathbf{D}_i| - \frac{\kappa^{-1}(v_i)}{2} \boldsymbol{\epsilon}_i^T (\Sigma_i + \mathbf{D}_i \mathbf{W}_i \mathbf{D}_i)^{-1} \boldsymbol{\epsilon}_i \\ &\quad - \frac{1}{2} \log |\Psi| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c. \end{aligned}$$

Mediante diferenciação obtemos que

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} &= -\kappa^{-1}(v_i) (\boldsymbol{\epsilon}_i^T \Sigma_i^{-1} \mathbf{D}_i \otimes \mathbf{X}_i^T \Sigma_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i}, \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_{1r} \partial \boldsymbol{\omega}_i^T} &= -\frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \Sigma_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \mathbf{D}_i^{-1} \dot{\mathbf{D}}_i(r) \mathbf{D}_i^{-1} \mathbf{C}_i^{-1}) \mathbf{S}_{n_i} \\ &\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{C}_i^{-1} \mathbf{D}_i^{-1} \dot{\mathbf{D}}_i(r) \mathbf{D}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \Sigma_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i}, \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_{2s} \partial \boldsymbol{\omega}_i^T} &= \frac{1}{2} \text{vec}^T (\mathbf{C}_i^{-1} \dot{\mathbf{C}}_i(s) \mathbf{C}_i^{-1}) \mathbf{S}_{n_i} \\ &\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \Sigma_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \mathbf{D}_i^{-1} \mathbf{C}_i^{-1} \dot{\mathbf{C}}_i(s) \mathbf{C}_i^{-1}) \mathbf{S}_{n_i} \\ &\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{C}_i^{-1} \dot{\mathbf{C}}_i(s) \mathbf{C}_i^{-1} \mathbf{D}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \Sigma_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}_i^T} &= \mathbf{0}, \end{aligned}$$

para $r = 1, \dots, k_1$, $s = 1, \dots, k_2$, em que $\text{vec}^T(\cdot)$ denota a transposta do vetor $\text{vec}(\cdot)$, $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$ para $i = 1, \dots, M$, $\dot{\mathbf{D}}_i(r) = \partial \mathbf{D}_i / \partial \alpha_{1r}$ e $\dot{\mathbf{C}}_i(s) = \partial \mathbf{C}_i / \partial \alpha_{2s}$ são avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Perturbação das Funções de Escala dentro dos Indivíduos

Considere perturbar as funções de escala mediante substituir \mathbf{D}_i na decomposição escala-correlação por $\mathbf{D}_{i\omega} = \mathbf{D}_i \mathbf{W}_i^{-1}$, em que \mathbf{W}_i é uma matriz diagonal com elementos positivos, isto é, $\mathbf{W}_i = \text{diag}(\boldsymbol{\omega}_i)$ com $\boldsymbol{\omega}_i =$

$(\omega_{i1}, \dots, \omega_{in_i})^T$ tal que $\omega_{ij} > 0$ para $i = 1, \dots, M$ e $j = 1, \dots, n_i$. Desse modo temos que

$$\boldsymbol{\Sigma}_{i\omega} = \mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{W}_i^{-1}.$$

Aqui $\boldsymbol{\omega}_0 = \mathbf{1}_N$ com $N = \sum_{i=1}^M n_i$ e o logaritmo da verossimilhança de dados completos fica dado por $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ em que

$$\begin{aligned} L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log |\mathbf{W}_i| - \frac{\kappa^{-1}(v_i)}{2} \boldsymbol{\epsilon}_i^T \mathbf{W}_i \boldsymbol{\Sigma}_i^{-1} \mathbf{W}_i \boldsymbol{\epsilon}_i \\ &\quad - \frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i + \log h(v_i; \boldsymbol{\nu}) + c. \end{aligned}$$

Usando diferenciação de matrizes temos que

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} &= -\kappa^{-1}(v_i) (\boldsymbol{\epsilon}_i^T \otimes \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1}) \mathbf{S}_{n_i}^*, \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} &= -\frac{\kappa^{-1}(v_i)}{2} \text{vec}^T(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T + \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \boldsymbol{\Sigma}_i^{-1}) \mathbf{S}_{n_i}^* \text{ e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}_i^T} &= \mathbf{0}, \end{aligned}$$

para $r = 1, \dots, k$ e $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$ para $i = 1, \dots, M$.

Esse esquema de perturbação permite estudar a sensibilidade das funções de escala ante observações aberrantes e pode ser útil para modelar a estrutura de escala mediante propor funções apropriadas para o modelo considerado. Por exemplo, através desse esquema de perturbação podemos sugerir estruturas de escala diferentes da estrutura simples, $\boldsymbol{\Sigma}_i = \phi \mathbf{I}_{n_i}$, $i = 1, \dots, M$. Note também que esse esquema pode ser estendido considerando $\boldsymbol{\Sigma}_{i\omega} = \mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{W}_i^{-1}$, em que \mathbf{W}_i é matriz simétrica definida positiva.

2.6.2 Influência Local quando os v_i são não observáveis

A seguir desenvolvemos a influência local no modelo linear misto hierárquico em que os efeitos aleatórios são integrados, isto é, considerando a formulação hierárquica dada por (1.12). Nesse caso o vetor de dados completos é $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{v}^T)^T$ em que $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ denota as respostas observadas para os M indivíduos e $\mathbf{v} = (v_1, \dots, v_M)^T$ correspondem às variáveis de mistura as quais são consideradas não observáveis. Temos que o logaritmo

da função de verossimilhança para os dados completos \mathbf{Y}_c é $L(\boldsymbol{\theta}|\mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}|\mathbf{Y}_c)$, com

$$L_i(\boldsymbol{\theta}|\mathbf{Y}_c) = -\frac{1}{2} \log |\mathbf{V}_i| - \frac{\kappa^{-1}(v_i)}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) + \log h(v_i; \boldsymbol{\nu}) + c, \quad (2.20)$$

em que $h(v_i; \boldsymbol{\nu})$ é a função de densidade da variável de mistura v_i , c representa uma constante e $\mathbf{V}_i = \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$. A Q -função associada para esse modelo assume a forma $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, com $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ apresentada em (1.21).

Matriz Hessiana, \ddot{Q}

Consideramos $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ como o vetor de parâmetros de interesse. A matriz hessiana $\ddot{Q}(\boldsymbol{\theta})$ (veja Apêndice A.3.2) avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ é dada por $\ddot{Q}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M \ddot{Q}_i(\hat{\boldsymbol{\theta}})$, em que $\ddot{Q}_i(\hat{\boldsymbol{\theta}})$ pode ser expressa como

$$\ddot{Q}_i(\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \ddot{Q}_{11,i} & \ddot{Q}_{12,i} & \ddot{Q}_{13,i} \\ \ddot{Q}_{12,i}^T & \ddot{Q}_{22,i} & \ddot{Q}_{23,i} \\ \ddot{Q}_{13,i}^T & \ddot{Q}_{23,i}^T & \ddot{Q}_{33,i} \end{pmatrix}, \quad (2.21)$$

em que

$$\begin{aligned} \ddot{Q}_{11,i} &= \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\hat{\kappa}_i \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i, \\ \ddot{Q}_{12,i} &= \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{e} \quad \ddot{Q}_{13,i} = \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \end{aligned}$$

com

$$\begin{aligned} \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \alpha_r} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= -\hat{\kappa}_i \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \quad \text{e} \\ \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \lambda_t} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= -\hat{\kappa}_i \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}), \end{aligned}$$

para $r = 1, \dots, k$; $t = 1, \dots, d$. Temos que

$$\ddot{Q}_{22,i} = \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{e} \quad \ddot{Q}_{23,i} = \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\lambda}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

com elementos dados por

$$\begin{aligned} \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \alpha_r \partial \alpha_s} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \widehat{\mathbf{V}}_i^{-1} \{ \dot{\boldsymbol{\Sigma}}_i(r) \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) - \ddot{\boldsymbol{\Sigma}}_i(r, s) \} \\ &\quad - \frac{1}{2} \widehat{\kappa}_i \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \{ \dot{\boldsymbol{\Sigma}}_i(r) \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(s) + \dot{\boldsymbol{\Sigma}}_i(s) \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) - \ddot{\boldsymbol{\Sigma}}_i(r, s) \} \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i \quad \text{e} \\ \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \alpha_r \partial \lambda_t} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \\ &\quad - \frac{1}{2} \widehat{\kappa}_i \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \{ \dot{\boldsymbol{\Sigma}}_i(r) \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T + \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \} \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i, \end{aligned}$$

para $r, s = 1, \dots, k$; $t = 1, \dots, d$ e

$$\ddot{Q}_{33,i} = \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

cujos elementos (t, u) fica dado por

$$\begin{aligned} \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \lambda_t \partial \lambda_u} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \{ \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(u) - \ddot{\boldsymbol{\Psi}}(t, u) \} \mathbf{Z}_i^T \\ &\quad - \frac{1}{2} \widehat{\kappa}_i \widehat{\mathbf{r}}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \{ \dot{\boldsymbol{\Psi}}(t) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(u) + \dot{\boldsymbol{\Psi}}(u) \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(t) \\ &\quad - \ddot{\boldsymbol{\Psi}}(t, u) \} \mathbf{Z}_i^T \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{r}}_i, \end{aligned}$$

para $t, u = 1, \dots, d$, em que $\widehat{\mathbf{V}}_i = \mathbf{Z}_i \widehat{\boldsymbol{\Psi}} \mathbf{Z}_i^T + \widehat{\boldsymbol{\Sigma}}_i$, $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}$ para $i = 1, \dots, M$, com $\widehat{\kappa}_i$ definido mediante (1.21) e como anteriormente, temos que $\dot{\boldsymbol{\Sigma}}_i(r) = \partial \boldsymbol{\Sigma}_i / \partial \alpha_r$, $\ddot{\boldsymbol{\Sigma}}_i(r, s) = \partial^2 \boldsymbol{\Sigma}_i / \partial \alpha_r \partial \alpha_s$, $\dot{\boldsymbol{\Psi}}(t) = \partial \boldsymbol{\Psi} / \partial \lambda_t$ e $\ddot{\boldsymbol{\Psi}}(t, u) = \partial^2 \boldsymbol{\Psi} / \partial \lambda_t \partial \lambda_u$ avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Esquemas de Perturbação

Apresentamos a matriz $\boldsymbol{\Delta}$ considerando os mesmos esquemas de perturbação abordados para o modelo linear misto hierárquico em que \mathbf{b}_i e v_i são considerados não observáveis. Para cada esquema de perturbação $\boldsymbol{\Delta}$ pode ser escrita como

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 \\ \boldsymbol{\Delta}_3 \end{pmatrix},$$

em que $\boldsymbol{\Delta}_1 = \partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T \in \mathbb{R}^{p \times q}$, $\boldsymbol{\Delta}_2 = \partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\alpha} \partial \boldsymbol{\omega}^T \in \mathbb{R}^{k \times q}$, $\boldsymbol{\Delta}_3 = \partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}^T \in \mathbb{R}^{d \times q}$ e q representa a dimensão do vetor de perturbação $\boldsymbol{\omega}$. Os diferenciais $d^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$ a partir dos quais é possível obter $\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c) / \partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T$ são dados no Apêndice A.3.2.

Ponderação de Casos

Considere a inclusão de pesos no logaritmo da função de verossimilhança de dados completos como

$$L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M \omega_i L_i(\boldsymbol{\theta} | \mathbf{Y}_c)$$

com $L_i(\boldsymbol{\theta} | \mathbf{Y}_c)$ dada por (2.20). Nesse caso o vetor de perturbação é $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$ tal que $0 \leq \omega_i \leq 1, \forall i$ e $\boldsymbol{\omega}_0 = \mathbf{1}_M$. Usando resultados de diferenciação de matrizes obtemos

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_i} &= \kappa^{-1}(v_i) \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}), \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \omega_i} &= -\frac{1}{2} \{ \text{tr} \mathbf{V}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) - \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \mathbf{V}_i^{-1} \mathbf{r}_i \} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \lambda_s \partial \omega_i} &= -\frac{1}{2} \{ \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T - \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \}, \end{aligned}$$

para $r = 1, \dots, k, s = 1, \dots, d$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}, i = 1, \dots, M$.

Mostramos a equivalência entre os diagnósticos obtidos por eliminação de casos e a influência local para esse esquema de perturbação no Apêndice B.

Perturbação da Matriz de Escala dos Efeitos Aleatórios

Seja $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$ vetor de perturbação $M \times 1$. Para este esquema de perturbação consideramos que a matriz de escala dos efeitos aleatórios fica dada por $\omega_i^{-1} \boldsymbol{\Psi}(\boldsymbol{\lambda})$, em que $\omega_i > 0, i = 1, \dots, M$. Temos que $\boldsymbol{\omega}_0 = \mathbf{1}_M$ e o logaritmo da verossimilhança de dados completos para o modelo perturbado é dado por $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$, com

$$L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = -\frac{1}{2} \log |\mathbf{V}_{i\omega}| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i + \log h(v_i; \boldsymbol{\nu}) + c,$$

em que $\mathbf{V}_{i\omega} = \omega_i^{-1} \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \boldsymbol{\Sigma}_i, \mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$. Mediante diferenciação obtemos

$$\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \omega_i} = \frac{\kappa^{-1}(v_i)}{2} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i,$$

$$\begin{aligned}\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \omega_i} &= -\frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \mathbf{V}_i^{-1} \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T \\ &\quad + \frac{\kappa^{-1}(v_i)}{2} \mathbf{r}_i^T \mathbf{V}_i^{-1} \{ \dot{\boldsymbol{\Sigma}}_i(r) \mathbf{V}_i^{-1} \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \} \mathbf{V}_i^{-1} \mathbf{r}_i \text{ e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \lambda_s \partial \omega_i} &= \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1} (\boldsymbol{\Sigma}_i - \kappa^{-1}(v_i) \mathbf{r}_i \mathbf{r}_i^T) \\ &\quad + \frac{\kappa^{-1}(v_i)}{2} \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \{ \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \boldsymbol{\Psi} + \boldsymbol{\Psi} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i,\end{aligned}$$

para $r = 1, \dots, k$ e $s = 1, \dots, d$.

Perturbação das Variáveis Explicativas

Perturbamos a matriz de planejamento \mathbf{X}_i , mediante uma perturbação aditiva numa particular variável explicativa contínua $\mathbf{x}_{it}(\boldsymbol{\omega}) = \mathbf{x}_{it} + \boldsymbol{\omega}_i s_i$ em que $\mathbf{x}_{it} \in \mathbb{R}^{n_i}$ denota a t -ésima coluna da matriz \mathbf{X}_i , $\boldsymbol{\omega}_i \in \mathbb{R}^{n_i}$ é vetor de perturbação n_i -dimensional e s_i é um fator de escala. Nesse caso $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^N$ com $N = \sum_{i=1}^M n_i$. Desse modo $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$, em que

$$\begin{aligned}L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} \log |\mathbf{V}_i| - \frac{\kappa^{-1}(v_i)}{2} (\mathbf{r}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i)^T \mathbf{V}_i^{-1} (\mathbf{r}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i) \\ &\quad + \log h(v_i; \boldsymbol{\nu}) + c,\end{aligned}$$

com $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, para $i = 1, \dots, M$ e \mathbf{c}_t vetor $p \times 1$ com 1 na t -ésima posição e zeros nas restantes. Obtemos que,

$$\begin{aligned}\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} &= -\kappa^{-1}(v_i) s_i (\boldsymbol{\beta}_t \mathbf{X}_i^T - \mathbf{c}_t \mathbf{r}_i^T) \mathbf{V}_i^{-1}, \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} &= -\kappa^{-1}(v_i) s_i \boldsymbol{\beta}_t \mathbf{r}_i^T \mathbf{V}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \mathbf{V}_i^{-1} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \lambda_s \partial \boldsymbol{\omega}_i^T} &= -\kappa^{-1}(v_i) s_i \boldsymbol{\beta}_t \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1},\end{aligned}$$

para $r = 1, \dots, k$, $s = 1, \dots, d$, em que $\boldsymbol{\beta}_t$ representa o t -ésimo elemento de $\boldsymbol{\beta}$.

Para esse esquema de perturbação a matriz de planejamento perturbada para o i -ésimo indivíduo fica dada por $\mathbf{X}_{i\omega} = \mathbf{X}_i + s_i \boldsymbol{\omega}_i \mathbf{c}_t^T$, para $i = 1, \dots, M$. Desse modo podemos estender esse esquema para perturbar todas as colunas da matriz \mathbf{X}_i , mediante $\mathbf{X}_{i\omega} = \mathbf{X}_i + \mathbf{W}_i \mathbf{S}_i$ com \mathbf{W}_i denotando uma matriz de perturbação $n_i \times p$, $\mathbf{S}_i = \text{diag}(s_{i1}, \dots, s_{ip})$ com s_{ij} para $j = 1, \dots, p$ sendo um fator de escala e a matriz de não perturbação é $\mathbf{W}_0 = \mathbf{0}$.

Perturbação da Resposta

Seja $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^T, \dots, \boldsymbol{\omega}_M^T)^T$ vetor de perturbação N -dimensional, em que $N = \sum_{i=1}^M n_i$ e $\boldsymbol{\omega}_0 = \mathbf{1}_M$. Perturbamos as respostas observadas $(\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$, como $\mathbf{Y}_{i\omega} = \mathbf{Y}_i + \boldsymbol{\omega}_i$, $i = 1, \dots, M$. Para esse esquema de perturbação o logaritmo da verossimilhança de dados completos fica dado por $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$, em que

$$L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = -\frac{1}{2} \log |\mathbf{V}_i| - \frac{\kappa^{-1}(v_i)}{2} (\mathbf{r}_i + \boldsymbol{\omega}_i)^T \mathbf{V}_i^{-1} (\mathbf{r}_i + \boldsymbol{\omega}_i) + \log h(v_i; \boldsymbol{\nu}) + c.$$

Diferenciando com relação a $\boldsymbol{\omega}_i$ e $\boldsymbol{\theta}$, obtemos

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} &= \kappa^{-1}(v_i) \mathbf{X}_i^T \mathbf{V}_i^{-1}, \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} &= \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \dot{\Sigma}_i(r) \mathbf{V}_i^{-1} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \lambda_s \partial \boldsymbol{\omega}_i^T} &= \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\Psi}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1}, \end{aligned}$$

para $r = 1, \dots, k$, $s = 1, \dots, d$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, $i = 1, \dots, M$.

Na próxima seção, investigamos a relação entre a matriz de alavancas generalizadas para dados não observáveis no modelo linear misto hierárquico em que os efeitos aleatórios são integrados e a influência local para o esquema de perturbação aditiva na resposta observada.

Perturbação da Matriz de Correlação dentro dos Indivíduos

Baseados na decomposição de escala-correlação dada em (2.19), podemos perturbar a matriz de correlação dentro dos indivíduos como

$$\boldsymbol{\Sigma}_{i\omega} = \mathbf{D}_i \mathbf{C}_{i\omega} \mathbf{D}_i, \quad \text{com} \quad \mathbf{C}_{i\omega} = \mathbf{C}_i + \mathbf{W}_i,$$

em que $\mathbf{C}_{i\omega}$ é matriz de correlação com \mathbf{W}_i sendo matriz de perturbação simétrica $n_i \times n_i$. Seja $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^T, \dots, \boldsymbol{\omega}_M^T)^T$ em que $\boldsymbol{\omega}_i = \text{vech}(\mathbf{W}_i) \in \mathbb{R}^{n_i(n_i+1)/2}$. Nesse caso o vetor de não perturbação é dado por $\boldsymbol{\omega}_0 = \mathbf{0}$ e o logaritmo da verossimilhança de dados completos para o modelo perturbado fica dado por $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$, com

$$L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = -\frac{1}{2} \log |\mathbf{V}_i + \mathbf{D}_i \mathbf{W}_i \mathbf{D}_i| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{r}_i^T (\mathbf{V}_i + \mathbf{D}_i \mathbf{W}_i \mathbf{D}_i)^{-1} \mathbf{r}_i + \log h(v_i; \boldsymbol{\nu}) + c.$$

Utilizando resultados de diferenciação de matrizes, obtemos

$$\begin{aligned}
\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} &= -\kappa^{-1}(v_i) (\mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \otimes \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i}, \\
\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_{1r} \partial \boldsymbol{\omega}_i^T} &= -\frac{1}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \dot{\mathbf{D}}_i(r) + \dot{\mathbf{D}}_i(r) \mathbf{V}_i^{-1} \mathbf{D}_i + \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{F}_i \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \{ \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{F}_i + \mathbf{F}_i \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \} \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \\
&\quad + \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\dot{\mathbf{D}}_i(r) \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i + \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \dot{\mathbf{D}}_i(r)) \mathbf{S}_{n_i}, \\
\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_{2s} \partial \boldsymbol{\omega}_i^T} &= \frac{1}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{D}_i \dot{\mathbf{C}}_i(s) \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \dot{\mathbf{C}}_i(s) \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{D}_i \dot{\mathbf{C}}_i(s) \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \quad \text{e} \\
\frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \lambda_t \partial \boldsymbol{\omega}_i^T} &= \frac{1}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\Psi}(t) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\Psi}(t) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\Psi}(t) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i) \mathbf{S}_{n_i},
\end{aligned}$$

para $r = 1, \dots, k_1$, $s = 1, \dots, k_2$, $t = 1, \dots, d$, em que $\mathbf{F}_i = \dot{\mathbf{D}}_i(r) \mathbf{C}_i \mathbf{D}_i + \mathbf{D}_i \mathbf{C}_i \dot{\mathbf{D}}_i(r)$, $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$ e \mathbf{S}_{n_i} denota a matriz de duplicação de ordem n_i , aqui $\dot{\mathbf{D}}_i(r) = \partial \mathbf{D}_i / \partial \alpha_{1r}$, $\dot{\mathbf{C}}_i(s) = \partial \mathbf{C}_i / \partial \alpha_{2s}$ são avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Perturbação das Funções de Escala dentro dos Indivíduos

Perturbamos as funções de escala na decomposição de escala-correlação dada em (2.19) como $\mathbf{D}_{i\omega} = \mathbf{D}_i \mathbf{W}_i^{-1}$, em que $\mathbf{W}_i = \text{diag}(\boldsymbol{\omega}_i)$ é matriz diagonal positiva definida. Nesse caso o vetor de não perturbação é $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^T, \dots, \boldsymbol{\omega}_M^T)^T$ e $\boldsymbol{\omega}_0 = \mathbf{1}_N$ com $N = \sum_{i=1}^M n_i$. O logaritmo da verossimilhança de dados completos para o modelo perturbado é dado por $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ em que

$$\begin{aligned}
L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} \log |\mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \boldsymbol{\Sigma}_{i\omega}| - \frac{\kappa^{-1}(v_i)}{2} \mathbf{r}_i^T (\mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \boldsymbol{\Sigma}_{i\omega})^{-1} \mathbf{r}_i \\
&\quad + \log h(v_i; \boldsymbol{\nu}) + c,
\end{aligned}$$

em que $\Sigma_{i\omega} = \mathbf{W}_i^{-1} \Sigma_i \mathbf{W}_i^{-1}$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$. Diferenciando com relação a \mathbf{W}_i e $\boldsymbol{\theta}$ obtemos que

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} &= \kappa^{-1}(v_i) (\mathbf{r}_i^T \mathbf{V}_i^{-1} \Sigma_i \otimes \mathbf{X}_i^T \mathbf{V}_i^{-1}) \mathbf{S}_{n_i}^*, \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T} &= \text{vec}(\mathbf{V}_i^{-1} \dot{\Sigma}_i(r) - \Sigma_i \mathbf{V}_i^{-1} \dot{\Sigma}_i(r) \mathbf{V}_i^{-1}) \mathbf{S}_{n_i}^* \\ &\quad + \kappa^{-1}(v_i) \text{vec}^T(\mathbf{V}_i^{-1} \dot{\Sigma}_i(r) \mathbf{G}_i \Sigma_i + \mathbf{G}_i \dot{\Sigma}_i(r) \mathbf{V}_i^{-1} \Sigma_i - \mathbf{G}_i \dot{\Sigma}_i(r)) \mathbf{S}_{n_i}^* \text{ e} \\ \frac{\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{Y}_c)}{\partial \lambda_s \partial \boldsymbol{\omega}_i^T} &= -\text{vec}^T(\mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\Psi}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \Sigma_i) \mathbf{S}_{n_i}^* \\ &\quad + \kappa^{-1}(v_i) \text{vec}^T(\mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\Psi}(s) \mathbf{Z}_i^T \mathbf{G}_i \Sigma_i + \mathbf{G}_i \mathbf{Z}_i \dot{\Psi}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \Sigma_i) \mathbf{S}_{n_i}^*, \end{aligned}$$

para $r = 1, \dots, k$; $s = 1, \dots, d$, $\mathbf{G}_i = \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1}$ e $\mathbf{S}_{n_i}^*$ sendo uma matriz de duplicação tal que $\text{vec}(\mathbf{W}_i) = \mathbf{S}_{n_i}^* \boldsymbol{\omega}_i$, $i = 1, \dots, M$.

2.7 Método de Alavanca Generalizado no Modelo Linear Misto Hierárquico

Nesta seção derivamos a matriz de alavancas generalizadas para o modelo linear misto hierárquico usando a proposta para modelos com dados incompletos dada em (2.8). Posteriormente, analisamos a relação entre o método de alavanca e a influência local considerando perturbações aditivas nas respostas observadas.

O vetor de posição para o i -ésimo indivíduo é dado por $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i$ e $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$ para o modelo linear misto em sua formulação hierárquica quando \mathbf{b}_i e v_i são não observáveis e quando os efeitos aleatórios \mathbf{b}_i são integrados, respectivamente. Para ambos os casos temos que

$$\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^T} = \mathbf{X}_i, \quad \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\alpha}^T} = \mathbf{0} \quad \text{e} \quad \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\lambda}^T} = \mathbf{0}, \quad i = 1, \dots, M.$$

Seja $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_M^T)^T \in \mathbb{R}^N$ com $N = \sum_{i=1}^M n_i$, então temos que a matriz Jacobiana $\mathbf{D} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\theta}^T$ é dada por $\mathbf{D} = (\mathbf{X}, \mathbf{0}, \mathbf{0})$, em que \mathbf{X} tem forma particionada

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_M \end{pmatrix}. \quad (2.22)$$

De (2.8) temos que a matriz de alavancas generalizadas para o modelo linear misto hierárquico é dada por

$$GL_Q(\hat{\boldsymbol{\theta}}) = D\{-\ddot{Q}(\hat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta},$$

em que a matriz hessiana $\ddot{Q}(\hat{\boldsymbol{\theta}})$ é dada por (2.18) e (2.21), quando \mathbf{b}_i e v_i são considerados não observáveis e quando os efeitos aleatórios são integrados, respectivamente, e $\boldsymbol{\Delta} = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\mathbf{Y}^T$ pode ser particionada como

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 \\ \boldsymbol{\Delta}_3 \end{pmatrix},$$

em que $\boldsymbol{\Delta}_1 = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\beta}\partial\mathbf{Y}^T \in \mathbb{R}^{p \times N}$, $\boldsymbol{\Delta}_2 = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\alpha}\partial\mathbf{Y}^T \in \mathbb{R}^{k \times N}$ e $\boldsymbol{\Delta}_3 = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\lambda}\partial\mathbf{Y}^T \in \mathbb{R}^{d \times N}$ são avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

A seguir apresentamos a matriz $\partial^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)/\partial\boldsymbol{\theta}\partial\mathbf{Y}^T$ para o modelo linear misto na sua formulação hierárquica, primeiramente no caso em que ambos \mathbf{b}_i e v_i são não observáveis e posteriormente quando somente as variáveis de mistura v_i são consideradas não observáveis.

Para o modelo linear misto hierárquico dado por (2.17) obtemos que

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)}{\partial\boldsymbol{\beta}\partial\mathbf{Y}_i^T} &= \kappa^{-1}(v_i)\mathbf{X}_i^T\boldsymbol{\Sigma}_i^{-1}, & \frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)}{\partial\boldsymbol{\lambda}\partial\mathbf{Y}_i^T} &= \mathbf{0} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)}{\partial\alpha_r\partial\mathbf{Y}_i^T} &= \kappa^{-1}(v_i)\boldsymbol{\epsilon}_i^T\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\boldsymbol{\Sigma}_i^{-1}, \end{aligned}$$

para $r = 1, \dots, k$ e $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i$, $i = 1, \dots, M$.

Seja $\ddot{Q}(\hat{\boldsymbol{\eta}}) = \sum_{i=1}^M \ddot{Q}_i(\hat{\boldsymbol{\eta}})$ em que $\ddot{Q}_i(\hat{\boldsymbol{\eta}}) = \partial^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ é dada em (2.18) e considere

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{12}^T & \mathbf{T}_{22} \end{pmatrix} = -\ddot{Q}(\hat{\boldsymbol{\eta}}).$$

Supondo $\mathbf{U} = \mathbf{T}_{11} - \mathbf{T}_{12}\mathbf{T}_{22}^{-1}\mathbf{T}_{12}^T$ matriz não-singular, então (veja Magnus e Neudecker, 1988)

$$\mathbf{T}^{-1} = \begin{pmatrix} \mathbf{U}^{-1} & -\mathbf{U}^{-1}\mathbf{T}_{12}\mathbf{T}_{22}^{-1} \\ -\mathbf{T}_{22}^{-1}\mathbf{T}_{12}^T\mathbf{U}^{-1} & \mathbf{T}_{22}^{-1} + \mathbf{T}_{22}^{-1}\mathbf{T}_{12}^T\mathbf{U}^{-1}\mathbf{T}_{12}\mathbf{T}_{22}^{-1} \end{pmatrix},$$

e portanto, obtemos que a matriz de alavancas generalizadas fica dada por

$$\begin{aligned} \mathbf{GL}_Q(\hat{\boldsymbol{\theta}}) &= (\mathbf{X}, \mathbf{0}, \mathbf{0}) \begin{pmatrix} \mathbf{T}^{-1} & \mathbf{0} \\ \mathbf{0} & \{-\ddot{\mathbf{Q}}(\hat{\boldsymbol{\lambda}})\}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 \\ \mathbf{0} \end{pmatrix} \\ &= \mathbf{GL}_{Q,\beta}(\hat{\boldsymbol{\theta}}) + \mathbf{GL}_{Q,\alpha}(\hat{\boldsymbol{\theta}}), \end{aligned}$$

em que

$$\mathbf{GL}_{Q,\beta}(\hat{\boldsymbol{\theta}}) = \mathbf{X}\mathbf{U}^{-1}\boldsymbol{\Delta}_1 \quad \text{e} \quad \mathbf{GL}_{Q,\alpha}(\hat{\boldsymbol{\theta}}) = -\mathbf{X}\mathbf{U}^{-1}\mathbf{T}_{12}\mathbf{T}_{22}^{-1}\boldsymbol{\Delta}_2.$$

Por outro lado, para o modelo linear misto hierárquico dado por (2.20) é possível mostrar que

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \mathbf{Y}_i^T} &= \kappa^{-1}(v_i) \mathbf{X}_i^T \mathbf{V}_i^{-1}, \\ \frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)}{\partial \alpha_r \partial \mathbf{Y}_i^T} &= \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(r) \mathbf{V}_i^{-1} \quad \text{e} \\ \frac{\partial^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)}{\partial \lambda_s \partial \mathbf{Y}_i^T} &= \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \dot{\boldsymbol{\Psi}}(s) \mathbf{Z}_i^T \mathbf{V}_i^{-1}, \end{aligned}$$

para $r = 1, \dots, k$, $s = 1, \dots, d$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$.

Considere $\mathbf{R} = \{-\ddot{\mathbf{Q}}(\hat{\boldsymbol{\theta}})\}^{-1}$ em que $\ddot{\mathbf{Q}}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M \ddot{\mathbf{Q}}_i(\hat{\boldsymbol{\theta}})$ com $\ddot{\mathbf{Q}}_i(\hat{\boldsymbol{\theta}})$ dada em (2.21). Particionando \mathbf{R} como

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{R}_{12}^T & \mathbf{R}_{22} & \mathbf{R}_{23} \\ \mathbf{R}_{13}^T & \mathbf{R}_{23}^T & \mathbf{R}_{33} \end{pmatrix},$$

obtemos que a matriz de alavancas generalizadas assume a forma

$$\begin{aligned} \mathbf{GL}_Q(\hat{\boldsymbol{\theta}}) &= (\mathbf{X}, \mathbf{0}, \mathbf{0}) \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{R}_{12}^T & \mathbf{R}_{22} & \mathbf{R}_{23} \\ \mathbf{R}_{13}^T & \mathbf{R}_{23}^T & \mathbf{R}_{33} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 \\ \boldsymbol{\Delta}_3 \end{pmatrix} \\ &= \mathbf{GL}_{Q,\beta}(\hat{\boldsymbol{\theta}}) + \mathbf{GL}_{Q,\alpha}(\hat{\boldsymbol{\theta}}) + \mathbf{GL}_{Q,\lambda}(\hat{\boldsymbol{\theta}}), \end{aligned}$$

em que

$$\begin{aligned} \mathbf{GL}_{Q,\beta}(\hat{\boldsymbol{\theta}}) &= \mathbf{X}\mathbf{R}_{11}\boldsymbol{\Delta}_1, \quad \mathbf{GL}_{Q,\alpha}(\hat{\boldsymbol{\theta}}) = \mathbf{X}\mathbf{R}_{12}\boldsymbol{\Delta}_2 \quad \text{e} \\ \mathbf{GL}_{Q,\lambda}(\hat{\boldsymbol{\theta}}) &= \mathbf{X}\mathbf{R}_{13}\boldsymbol{\Delta}_3, \end{aligned}$$

Para cada um dos enfoques considerados, isto é, tanto quando consideramos \mathbf{b}_i e v_i não observáveis assim como para o modelo em que as variáveis de mistura v_i são não observáveis, podemos definir a matrizes de alavancas generalizadas para o i -ésimo indivíduo como

$$\mathbf{GL}_{Q,i}(\hat{\boldsymbol{\theta}}) = \mathbf{D}_i \left\{ - \sum_{i=1}^M \ddot{\mathbf{Q}}_i(\hat{\boldsymbol{\theta}}) \right\}^{-1} \boldsymbol{\Delta}_i, \quad i = 1, \dots, M,$$

em que $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\theta}^T = (\mathbf{X}_i, \mathbf{0}, \mathbf{0})$ e $\boldsymbol{\Delta}_i = \partial^2 Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta} \partial \mathbf{Y}_i^T \in \mathbb{R}^{p^* \times n_i}$, para $i = 1, \dots, M$ com $p^* = p + k + d$.

Notamos alguma similaridade entre as matrizes de alavancas generalizadas para os modelos lineares mistos em suas formulações marginal e hierárquica. Em efeito, tanto para o modelo marginal assim como para o modelo hierárquico em que os efeitos aleatórios são integrados a matriz de alavancas generalizadas $\mathbf{GL}_Q(\hat{\boldsymbol{\theta}})$ depende das quantidades $\mathbf{GL}_{Q,\beta}(\hat{\boldsymbol{\theta}})$, $\mathbf{GL}_{Q,\alpha}(\hat{\boldsymbol{\theta}})$ e $\mathbf{GL}_{Q,\lambda}(\hat{\boldsymbol{\theta}})$.

Por outro lado, a matriz de alavancas generalizadas no modelo hierárquico em que ambos \mathbf{b}_i e v_i são considerados não observáveis, temos que $\mathbf{GL}_Q(\hat{\boldsymbol{\theta}})$ é decomposto somente em $\mathbf{GL}_{Q,\beta}(\hat{\boldsymbol{\theta}})$, $\mathbf{GL}_{Q,\alpha}(\hat{\boldsymbol{\theta}})$, ainda assim, a partir da equação (1.17) e notando que $\hat{\mathbf{b}}_i = \mathbf{b}_i(\hat{\boldsymbol{\theta}})$ observamos que informação relativa a $\hat{\boldsymbol{\lambda}}$ é incluída em $\mathbf{GL}_{Q,\alpha}(\hat{\boldsymbol{\theta}})$ uma vez que essa quantidade depende do resíduo condicional $\mathbf{e}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i$.

2.7.1 Conexão entre o Método de Alavanca Generalizado para Dados Incompletos e o Procedimento de Influência Local

Para examinar a relação entre a matriz de alavancas generalizadas e a influência local sob perturbações aditivas no vetor de respostas observadas consideramos que $\boldsymbol{\alpha}$ e $\boldsymbol{\lambda}$ são parâmetros fixos. Desse modo, a matriz de alavancas generalizadas para o modelo linear misto hierárquico em que \mathbf{b}_i e v_i são não observáveis fica dada por

$$\begin{aligned} \mathbf{GL}_Q(\hat{\boldsymbol{\theta}}) &= \mathbf{X} \{ -\ddot{\mathbf{Q}}(\hat{\boldsymbol{\beta}}) \}^{-1} \mathbf{X}^T \hat{\boldsymbol{\Sigma}}^* \\ &= \mathbf{X} \left(\sum_{i=1}^M \hat{\kappa}_i \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}^T \hat{\boldsymbol{\Sigma}}^*, \end{aligned} \quad (2.23)$$

com \mathbf{X} sendo definida em (2.22) e $\hat{\boldsymbol{\Sigma}}^* = \text{blc diag}(\hat{\kappa}_1 \hat{\boldsymbol{\Sigma}}_1, \dots, \hat{\kappa}_M \hat{\boldsymbol{\Sigma}}_M)$. Temos que a curvatura normal para este caso é dada por

$$\mathbf{B} = -\boldsymbol{\Delta}_1^T \{ \ddot{\mathbf{Q}}(\hat{\boldsymbol{\beta}}) \}^{-1} \boldsymbol{\Delta}_1,$$

com $\Delta_1 = \mathbf{X}^T \widehat{\Sigma}^*$. Comparando \mathbf{B} com (2.23) temos que a relação entre a curvatura normal e a matriz de alavancas generalizadas fica como $\mathbf{B} = \widehat{\Sigma}^* \mathbf{GL}_Q(\widehat{\theta})$.

Para o modelo linear misto hierárquico em que somente as variáveis de mistura v_i são não observáveis temos que a matriz de alavancas generalizadas é dada por

$$\begin{aligned} \mathbf{GL}_Q(\widehat{\theta}) &= \mathbf{X} \{-\ddot{\mathbf{Q}}(\widehat{\beta})\}^{-1} \mathbf{X}^T \widehat{\mathbf{V}}^* \\ &= \mathbf{X} \left(\sum_{i=1}^M \widehat{\kappa}_i \mathbf{X}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}^T \widehat{\mathbf{V}}^*, \end{aligned} \quad (2.24)$$

em que $\widehat{\mathbf{V}}^* = \text{blc diag}(\widehat{\kappa}_1 \widehat{\mathbf{V}}_1, \dots, \widehat{\kappa}_M \widehat{\mathbf{V}}_M)$ com $\widehat{\mathbf{V}}_i = \mathbf{Z}_i \widehat{\Psi} \mathbf{Z}_i^T + \widehat{\Sigma}_i$, $i = 1, \dots, M$. A curvatura normal associada ao modelo linear misto hierárquico definido por (2.20) assume a forma

$$\mathbf{B} = -\Delta_1^T \{\ddot{\mathbf{Q}}(\widehat{\beta})\}^{-1} \Delta_1,$$

em que $\Delta_1 = \mathbf{X}^T \widehat{\mathbf{V}}^*$. Para $\mathbf{GL}_Q(\widehat{\theta})$ definida por (2.24) temos que a relação entre a curvatura normal e a matriz de alavancas generalizadas fica dada como $\mathbf{B} = \widehat{\mathbf{V}}^* \mathbf{GL}_Q(\widehat{\theta})$.

A conexão explorada acima sugere que a proposta do método de alavanca para dados incompletos baseado na Q -função proveniente da esperança condicional do logaritmo da função de verossimilhança de dados completos, pode ser útil para propósitos de diagnóstico em modelos complexos.

2.8 Considerações Computacionais

Uma crítica habitual ao enfoque de influência local proposto por Cook (1986) é que o cálculo de $C_{h_{\max}}$ e \mathbf{h}_{\max} envolve a decomposição espectral da matriz $\mathbf{B} = -\Delta^T \ddot{\mathbf{L}}^{-1} \Delta$ ou $\mathbf{B} = -\Delta^T \ddot{\mathbf{Q}}^{-1} \Delta$, quando consideramos a função Q -afastamento, que dependendo da dimensão do vetor de perturbação $\omega \in \Omega \subset \mathbb{R}^q$ pode ser uma operação de alto custo computacional (veja por exemplo, Lesaffre e Verbeke, 1998, Cadigan e Farrell, 2002 e Pan e Fang, 2002). Verbeke e Molenberghs (2000) consideram um procedimento para evitar esse problema mediante notar que o posto da matriz \mathbf{B} é $p^* = p + k + d$ o qual freqüentemente é pequeno com relação a q . O método baseia-se em considerar um fator raiz-quadrada de $(-\ddot{\mathbf{L}})^{-1} = \mathbf{G}^T \mathbf{G}$ e obter a decomposição

espectral de $\mathbf{B}^* = \mathbf{G}\mathbf{\Delta}\mathbf{\Delta}^T\mathbf{G}^T$, e como \mathbf{B} e \mathbf{B}^* contêm os mesmos autovalores não zeros, \mathbf{h}_{\max} pode ser obtido a partir de \mathbf{B}^* . Contudo, formar a matriz de produtos cruzados \mathbf{B}^* não é um procedimento recomendado do ponto de vista computacional (veja Golub e van Loan, 1996).

Nesta seção propomos calcular \mathbf{h}_{\max} e $C_{\mathbf{h}_{\max}}$ usando a decomposição valor singular (Golub e van Loan, 1996) de $\mathbf{M} = \mathbf{\Delta}^T\mathbf{G}^T$ em que \mathbf{G} é um fator raiz-quadrada de $(-\ddot{\mathbf{L}})^{-1}$ ou $(-\ddot{\mathbf{Q}})^{-1}$ no caso de utilizar como medida de influência a função Q -afastamento. Notamos que a informação dada pela decomposição espectral de \mathbf{B} , isto é,

$$\mathbf{B} = \sum_{j=1}^q \delta_j \mathbf{v}_j \mathbf{v}_j^T,$$

em que $\{(\delta_j, \mathbf{v}_j^T)^T\}_{j=1}^q$ correspondem aos autovalores e autovetores de \mathbf{B} tais que $\delta_1 \geq \dots \geq \delta_{p^*} > \delta_{p^*+1} = \dots = \delta_q = 0$, pode ser recuperada considerando a decomposição valor singular (SVD) de $\mathbf{M} = \mathbf{\Delta}^T\mathbf{G}^T$ dada por $\mathbf{M} = \mathbf{V}\mathbf{D}\mathbf{U}^T$, em que $\mathbf{V} \in \mathbb{R}^{q \times p^*}$ é matriz ortogonal por colunas, $\mathbf{D} = \text{diag}(\delta_1^{1/2}, \dots, \delta_{p^*}^{1/2})$ e \mathbf{U} é matriz ortogonal $p^* \times p^*$. Usando a decomposição SVD temos que a curvatura normal assume a forma $\mathbf{B} = \mathbf{M}\mathbf{M}^T = \mathbf{V}\mathbf{D}^2\mathbf{V}^T$ em que a matriz \mathbf{D}^2 contém os autovalores não zeros de \mathbf{B} e seus os autovetores associados são dados pelas colunas de \mathbf{V} .

Algumas das vantagens da decomposição SVD para a obtenção da direção \mathbf{h}_{\max} são: não é requerido formar explicitamente a matriz $\mathbf{B} = -\mathbf{\Delta}^T\ddot{\mathbf{L}}^{-1}\mathbf{\Delta}$ (ou $\mathbf{B} = -\mathbf{\Delta}^T\ddot{\mathbf{Q}}^{-1}\mathbf{\Delta}$), os requerimentos de armazenamento são mínimos uma vez que é possível armazenar \mathbf{V} no espaço requerido para o armazenamento de $\mathbf{\Delta}$, também notamos que usando essa decomposição obtemos toda a informação requerida para a curvatura \mathbf{B} , pois

$$\mathbf{B} = \sum_{j=1}^{p^*} \delta_j \mathbf{v}_j \mathbf{v}_j^T = \mathbf{V}\mathbf{D}^2\mathbf{V}^T,$$

e finalmente observamos que os algoritmos disponíveis para o cálculo da decomposição SVD são superiores em termos de eficiência e estabilidade do que suas contrapartes para obter a decomposição espectral (veja Belsley, 1991 e Golub e van Loan, 1996).

2.8.1 Implementação

Implementamos os procedimentos para a análise de influência local e o método de alavanca generalizado em modelos lineares mistos hierárquicos considerando a classe de distribuições de mistura de escala normal na biblioteca para S-PLUS, [roblme](#).

Para o cálculo da curvatura normal \mathbf{B} assim como a direção \mathbf{h}_{\max} utilizamos as bibliotecas de programas BLAS níveis 2 e 3 (Dongarra, Du Croz, Hammarling e Hanson, 1988 e Dongarra, Du Croz, Hammarling e Duff, 1990) e LINPACK (Dongarra, Bunch, Moler e Stewart, 1979). A biblioteca [roblme](#) é disponível em <http://www.ime.usp.br/~osorio/roblme/>.

Capítulo 3

Aplicações

Para ilustrar as metodologias desenvolvidas nos capítulos anteriores realizamos estudos de diagnóstico de influência em três conjuntos de dados. Consideramos o estudo de ortodontia apresentado por Potthoff e Roy (1964), o estudo reprodutivo em roedores reportado por Dempster et al. (1984) e finalmente o conjunto de dados de capacidade vital pulmonar introduzido por Barnett (1969). Na seção final são apresentadas algumas considerações.

3.1 Dados Dentais

Considere o conjunto de dados dentais introduzidos por Potthoff e Roy (1964) em que medições dentais foram realizadas em 11 meninas e 16 meninos nas idades 8, 10, 12 e 14 anos. A variável resposta foi a distância, em milímetros, da pituitaria à fissura pterigomaxilar. A Figura 3.1 apresenta os perfis individuais dos grupos de meninas e meninos. Vários autores têm analisado este conjunto de dados. Por exemplo, Pendergast e Broffitt (1985) consideram procedimentos de estimação robusta em modelos de curvas de crescimento enquanto Pinheiro, Liu e Wu (2001) usaram um modelo linear misto considerando a distribuição t de Student numa formulação hierárquica. Estudos de sensibilidade são abordados por Pan, Fang e von Rosen (1997) e Pan e Bai (2003) usando o método de influência local em modelos de curvas de crescimento e Pan (2002) considera procedimentos de eliminação de casos. Mais recentemente, Savalli, Paula e Cysneiros (2006) aplicam um teste tipo-score para os componentes de variância. Em todos esses trabalhos observações aberrantes assim como influentes foram detectadas sob erros normais, indicando que métodos robustos podem ser usados.

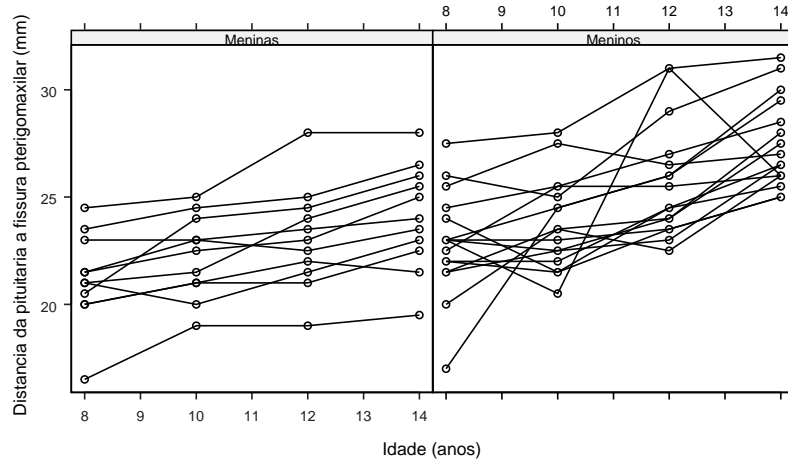


Figura 3.1: Perfis individuais para o conjunto de dados de ortodontia.

Desse modo, sugerimos analisar este conjunto de dados usando o modelo de intercepto e inclinação aleatórios:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 27,$$

em que \mathbf{Y}_i é um vetor aleatório 4-dimensional para as respostas do i -ésimo indivíduo, \mathbf{X}_i corresponde à matriz de planejamento 4×4 dada por,

$$\mathbf{X}_i = \begin{pmatrix} 1 & 8 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 12 & 0 & 0 \\ 1 & 14 & 0 & 0 \end{pmatrix}$$

para o grupo das meninas, e

$$\mathbf{X}_i = \begin{pmatrix} 0 & 0 & 1 & 8 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 1 & 12 \\ 0 & 0 & 1 & 14 \end{pmatrix}$$

para o grupo dos meninos, $\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2)^T$ é o vetor de parâmetros associados aos efeitos fixos, em que α_1 e β_1 representam o intercepto e a inclinação para o grupo das meninas enquanto α_2 e β_2 correspondem ao intercepto e inclinação para o grupo dos meninos, respectivamente, e \mathbf{Z}_i é a matriz de planejamento 4×2 para os efeitos aleatórios \mathbf{b}_i dada por

$$\mathbf{Z}_i^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix}, \quad i = 1, \dots, 27.$$

Assumimos que a distribuição conjunta de $(\mathbf{Y}_i^T, \mathbf{b}_i^T)^T$ é da forma

$$\begin{pmatrix} \mathbf{Y}_i \\ \mathbf{b}_i \end{pmatrix} \sim EC_6 \left(\begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \phi \mathbf{I}_4 & \mathbf{Z}_i \boldsymbol{\Psi} \\ \boldsymbol{\Psi} \mathbf{Z}_i^T & \boldsymbol{\Psi} \end{pmatrix}; g \right),$$

em que $\boldsymbol{\Psi}$ é considerada uma matriz simétrica não-estruturada 2×2 com elementos ψ_{11}, ψ_{12} e ψ_{22} . As inferências serão baseadas no modelo marginal dado por $\mathbf{Y}_i \sim EC_4(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i(\boldsymbol{\tau}); g)$, para $i = 1, \dots, 27$, em que $\mathbf{V}_i(\boldsymbol{\tau}) = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \phi \mathbf{I}_4$ com $\boldsymbol{\tau} = (\psi_{11}, \psi_{12}, \psi_{22}, \phi)^T$. Note que o modelo está na formulação marginal apresentada em (1.2).

Análises dos Modelos Ajustados

Para fins comparativos consideramos nas nossas análises as distribuições t de Student, exponencial potência e normal. Seguindo Lange, Little e Taylor (1989) usamos o critério de informação de Schwarz para a escolha dos graus de liberdade ν para a distribuição t de Student e do parâmetro de forma λ no caso da distribuição exponencial potência; desse modo, obtivemos $\nu = 5$ e $\lambda = \frac{2}{3}$, respectivamente. Portanto, para ambos os modelos distribuições com caudas pesadas serão assumidas para os erros aleatórios. As estimativas de máxima verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\tau}^T)^T$ para os três modelos ajustados são apresentadas na Tabela 3.1.

Tabela 3.1: Estimativas de máxima verossimilhança para os três modelos ajustados ao conjunto de dados dentais.

Parâmetro	Normal		t de Student		Exponencial potência	
	Estimativa	E.P.	Estimativa	E.P.	Estimativa	E.P.
α_1	17,373	(1,182)	17,610	(0,992)	17,568	(1,095)
β_1	0,480	(0,100)	0,459	(0,084)	0,462	(0,093)
α_2	16,341	(0,980)	16,948	(0,823)	16,699	(0,908)
β_2	0,784	(0,083)	0,716	(0,070)	0,744	(0,077)
ψ_{11}	4,557	(4,672)	3,270	(2,950)	1,185	(1,100)
ψ_{12}	-0,198	(0,379)	-0,133	(0,233)	-0,053	(0,088)
ψ_{22}	0,024	(0,034)	0,020	(0,022)	0,007	(0,008)
ϕ	1,716	(0,330)	0,887	(0,223)	0,358	(0,079)

Podemos notar pela Tabela 3.1 que as estimativas do intercepto e da inclinação são similares entre os três modelos ajustados, contudo os erros padrão para os modelos t de Student e exponencial potência são menores do

que os obtidos para o modelo normal, indicando que os dois modelos com caudas mais pesadas parecem produzir estimativas de máxima verossimilhança mais precisas. As inferências para os componentes da escala são similares para os três modelos ajustados, porém as estimativas não são comparáveis pois estão em diferentes escalas.

Com o intuito de detectar observações aberrantes em modelos elípticos multivariados a distância de Mahalanobis $U_i = (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})$, $i = 1, \dots, M$, tem sido considerada (Little, 1988, Lange, Little e Taylor, 1989 e Copt e Victoria-Feser, 2006). Para o caso normal, temos que U_i segue uma distribuição qui-quadrado com n_i graus de liberdade. Deste modo, podemos usar os percentis $\chi_{n_i}^2(\xi)$, em que $0 < \xi < 1$ como uma medida para identificação de pontos aberrantes. As distâncias de Mahalanobis modificadas, $F_i = U_i/n_i \sim F_{n_i, \nu}$ e $T_i = U_i^\lambda \sim Gama(\frac{n_i}{2\lambda}, \frac{1}{2})$ podem também ser consideradas para as distribuições t de Student com ν graus de liberdade e exponencial potência com parâmetro de forma λ , respectivamente. A Figura 3.2 apresenta os gráficos de índices de tais distâncias para os três modelos ajustados. As linhas segmentadas correspondem ao percentil $\xi = 0,975$.

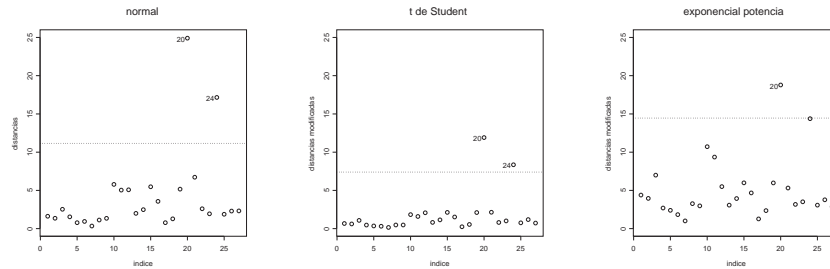


Figura 3.2: Gráficos de índices das distâncias de Mahalanobis para os três modelos ajustados.

Pode-se apreciar a partir dessas figuras que as observações 20 e 24 correspondentes às medições associadas aos meninos 9 e 13, aparecem como possíveis observações aberrantes. As ponderações atribuídas pelo procedimento de estimação para esses dois meninos são as menores para ambos os modelos ajustados em que distribuições com caudas pesadas são utilizadas (veja Figura 3.3), confirmando deste modo os aspectos de robustez dos estimadores de máxima verossimilhança contra observações aberrantes. Com fins comparativos, a função de pesos para o caso normal ($q_i(\boldsymbol{\theta}) = 1, \forall i$), é indicada como uma linha segmentada nos gráficos da Figura 3.3.

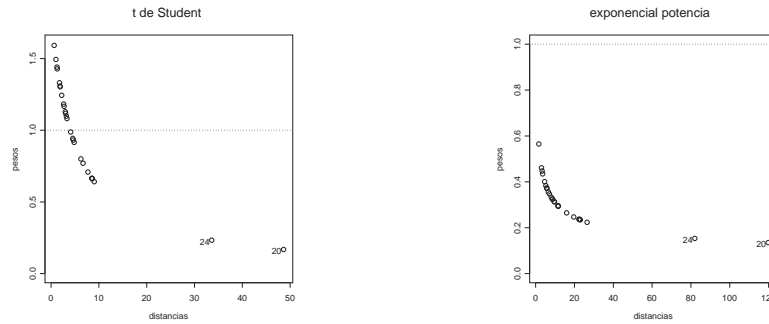


Figura 3.3: Pesos estimados para os modelos t de Student e exponencial potência.

A Figura 3.4 apresenta os gráficos de distâncias transformadas (Lange et al., 1989, veja também Galea et al., 2005) para os três modelos ajustados, que parecem indicar que tanto o modelo t quanto o exponencial potência proporcionam um melhor ajuste do que o modelo sob erro normal.

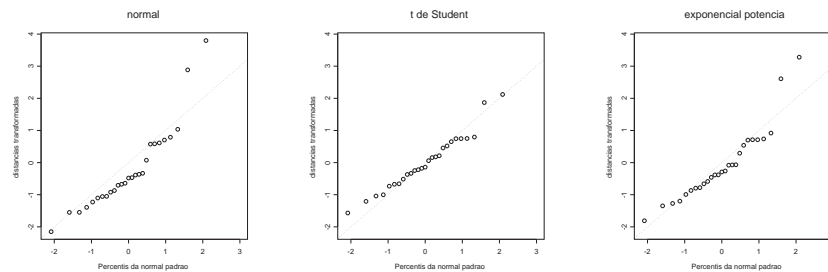


Figura 3.4: Gráficos de distâncias transformadas para os três modelos ajustados.

Para identificar observações influentes nos modelos ajustados no conjunto de dados de ortodontia, apresentamos os diagramas de índices da curvatura conformal B_i sob três esquemas de perturbação discutidos na Seção 2.4.

Ponderação de Casos: baseados neste esquema de perturbação, os gráficos de índices de $B_i(\beta)$ e $B_i(\tau)$ para os três modelos ajustados são apresentados nas Figuras 3.5-3.6, respectivamente.

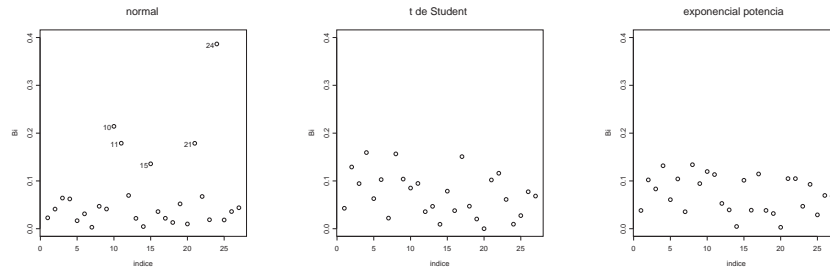


Figura 3.5: Gráficos de índices de B_i para $\hat{\beta}$ sob ponderação de casos.

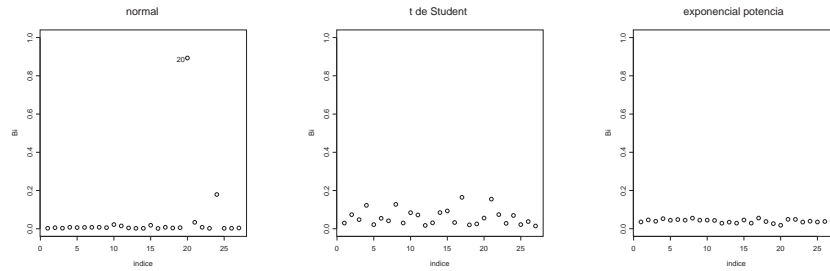


Figura 3.6: Gráficos de índices de B_i para $\hat{\tau}$ sob ponderação de casos.

Notamos que sob erros normais o indivíduo 24 se destaca como o mais influente em $\hat{\beta}$ assim como uma influência menor dos indivíduos 10, 11, 15 e 21. Não são apreciadas observações que influenciem de forma relevante os resultados sob erros t de Student e exponencial potência. Esses resultados concordam com os reportados por Pan, Fang e von Rosen (1997), Pan (2002) e Pan e Bai (2003), e também com os comentários dados em Pinheiro, Liu e Wu (2001) e Savalli, Paula e Cysneiros (2006). Por outro lado, observando a Figura 3.6 pode-se notar a grande influência da observação 20 em $\hat{\tau}$ sob erro normal. Os gráficos de índices de $B_i(\theta)$ são similares aos apresentados na Figura 3.6 e portanto tais considerações também são válidas para $\hat{\theta}$.

Perturbação da Matriz de Escala: as Figuras 3.7 e 3.8 apresentam os gráficos de índices de $B_i(\beta)$ e $B_i(\tau)$, para os modelos normal, t de Student e exponencial potência. Os gráficos de índices de $B_i(\theta)$ não são apresentados

por serem muito similares àqueles apresentados na Figura 3.8.

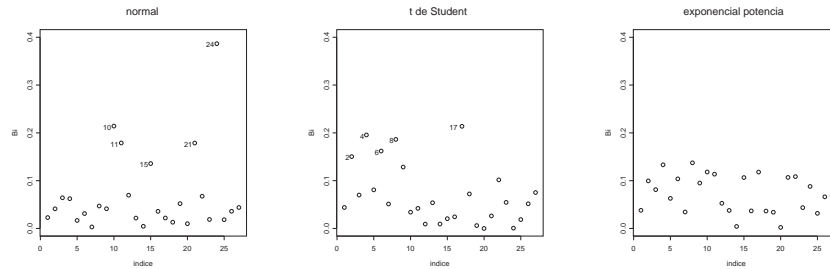


Figura 3.7: Gráficos de índices de B_i para $\hat{\beta}$ sob perturbação da matriz de escala.

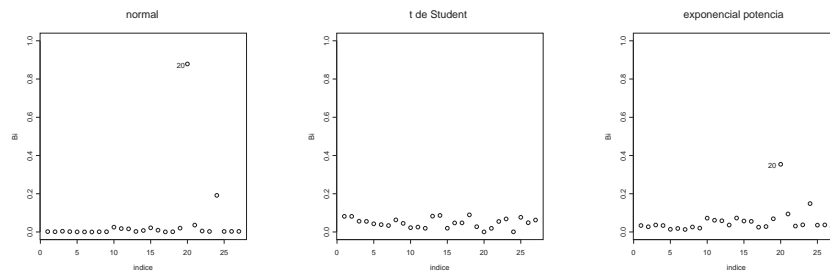


Figura 3.8: Gráficos de índices de B_i para $\hat{\tau}$ sob perturbação da matriz de escala.

As Figuras 3.7 e 3.8 são similares às Figuras 3.5 e 3.6, exceto que sob erro t de Student cinco observações têm influência moderada em $\hat{\beta}$, enquanto que sob erro exponencial potência uma observação é apontada com alguma influência em $\hat{\tau}$. Tais resultados concordam com os reportados por Pan, Fang e von Rosen (1997), Pan (2002) e Pan e Bai (2003) sob erros normais.

Perturbação da Resposta: para o esquema de perturbação da resposta apresentamos o diagrama de índices de $B_i(\theta)$ na Figura 3.9. Notamos que a partir desse esquema de perturbação podemos examinar a influência das medições *dentro* da cada um dos indivíduos. A partir da Figura 3.9 apreciamos alguma influência no modelo sob erro normal, especificamente quando as respostas dos indivíduos 20 e 24 são perturbadas.

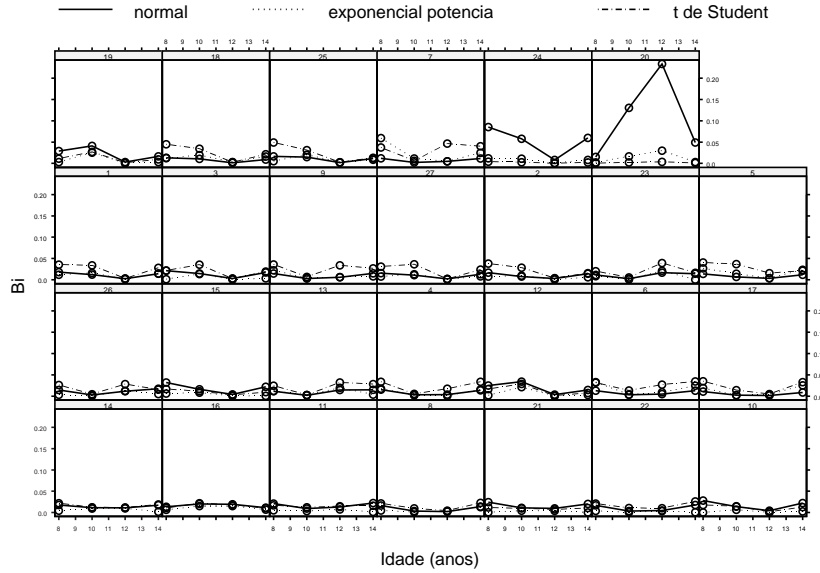


Figura 3.9: Gráfico de índices de B_i para $\hat{\theta}$ sob perturbação da resposta.

3.2 Estudo Reprodutivo em Roedores

Estes dados foram reportados por Dempster et al. (1984) e correspondem a um estudo em roedores com o objetivo de determinar o efeito de um composto toxicológico experimental no desempenho reprodutivo. A variável resposta foi o peso ao nascer, em gramas, de suas crias. Dez fêmeas reprodutoras foram destinadas aleatoriamente a 3 tratamentos (controle, dose baixa e alta de uma certa substância experimental). Tem-se observações somente para 7 ninhadas no grupo que recebeu a dose alta e um número diferente de crias é disponível para cada uma das ninhadas. Existem duas fontes principais de variação aleatória: entre e dentro da ninhada. Dempster et al. (1984) adotaram um modelo com efeitos mistos sob normalidade e incluíram efeitos fixos de tratamento, tamanho da ninhada e sexo da cria. Desse modo consideram o modelo

$$\begin{aligned}
 Y_{ijkl} &= \beta_j + b_i + \gamma L_{ij} + \delta S_k + \epsilon_{ijkl}, \quad i = 1, \dots, 27, \\
 j &= 1, 2, 3, \quad k = 0, 1 \quad \text{e} \quad l = 1, \dots, l_i,
 \end{aligned}
 \tag{3.1}$$

em que Y_{ijkl} representa o peso ao nascer da l -ésima cria na i -ésima ninhada ($i = 1, \dots, 27$) para o j -ésimo tratamento ($j = 1, 2, 3$) com sexo k ($k = 0, 1$); β_j é o efeito fixo do tratamento j ; b_i denota o efeito aleatório da ninhada no tratamento j ; L_{ij} é o tamanho da ninhada i associada ao j -ésimo tratamento; S_k representa o efeito fixo do sexo da cria, e ϵ_{ijkl} é o erro aleatório dentro da ninhada. A Figura 3.10 apresenta o peso das crias organizado por ninhada.

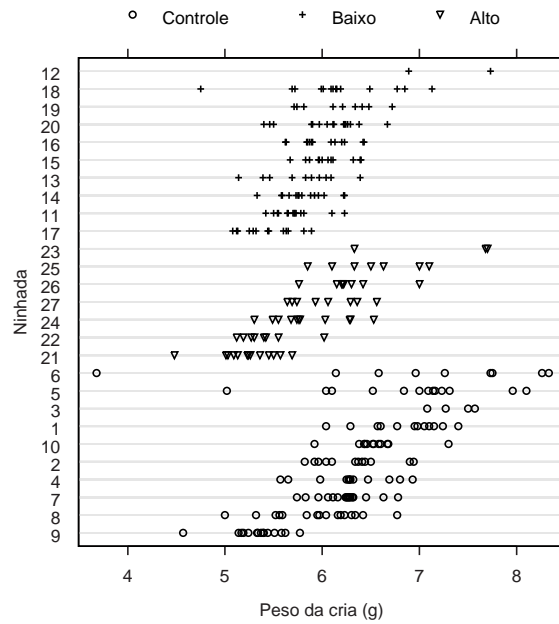


Figura 3.10: Peso das crias para cada ninhada.

Rosa, Padovani e Gianola (2003) observam que existem algumas crias cujo peso não é bem ajustado pelo modelo com efeitos mistos sob normalidade e consideram a estimação no modelo linear misto usando misturas de escala normal sob uma perspectiva bayesiana.

O modelo em (3.1) pode ser escrito como:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_ib_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 27,$$

em que \mathbf{Y}_i denota o vetor n_i -dimensional dos pesos observados para as crias na i -ésima ninhada, \mathbf{X}_i matriz de planejamento associada ao modelo (3.1) com $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \gamma, \delta)^T$, $\mathbf{Z}_i = \mathbf{1}_{n_i}$ é a matriz de planejamento para os efeitos aleatórios e $\boldsymbol{\epsilon}_i$ representa um erro aleatório.

Assumimos o modelo:

$$\begin{aligned} \mathbf{Y}_i | b_i, v_i &\stackrel{ind}{\sim} N_{n_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i b_i, \kappa(v_i) \phi \mathbf{I}_{n_i}), \quad b_i | v_i \stackrel{ind}{\sim} N(0, \kappa(v_i) \psi) \quad e \\ v_i &\stackrel{ind}{\sim} H(v_i; \boldsymbol{\nu}), \quad i = 1, \dots, 27, \end{aligned}$$

em que $H(v_i; \boldsymbol{\nu})$ é a função de distribuição para as variáveis de mistura e $\kappa(\cdot)$ é uma função positiva. Logo, apreciamos que o modelo está na formulação hierárquica dada em (1.4).

Análises dos Modelos Ajustados

Neste conjunto de dados, consideramos que a variável de mistura segue uma distribuição gama e uma estável positiva, isto é, assumimos que a resposta marginal segue uma distribuição t de Student e exponencial potência (veja Seção 1.3.3). Também apresentamos os resultados para o modelo normal. Foram escolhidos $\nu = 4$ e $\lambda = \frac{2}{3}$ para os graus de liberdade na distribuição t de Student e para o parâmetro de forma da exponencial potência, respectivamente. As estimativas de máxima verossimilhança para os três modelos ajustados são apresentadas na Tabela 3.2.

Tabela 3.2: Estimativas de máxima verossimilhança para os três modelos ajustados aos dados de peso de roedores.

Parâmetro	Normal		t de Student		Exponencial potência	
	Estimativa	E.P.	Estimativa	E.P.	Estimativa	E.P.
β_1	7,853	(0,231)	7,757	(0,231)	7,862	(0,191)
β_2	7,430	(0,227)	7,312	(0,227)	7,435	(0,185)
β_3	6,969	(0,188)	6,821	(0,188)	6,974	(0,157)
γ	-0,119	(0,016)	-0,110	(0,016)	-0,119	(0,013)
δ	0,297	(0,048)	0,254	(0,037)	0,268	(0,044)
ϕ	0,168	(0,016)	0,090	(0,018)	0,022	(0,002)
ψ	0,059	(0,003)	0,059	(0,002)	0,010	(0,001)

Pode-se apreciar que as estimativas para os três modelos considerados são muito similares, contudo para os parâmetros associados à escala $\boldsymbol{\tau} = (\phi, \psi)^T$ tais estimativas não são comparáveis.

Usando o gráfico de distâncias de Mahalanobis (Figura 3.11) detectamos as ninhadas 5 e 6 como possíveis observações aberrantes. Note que tais observações recebem ponderações pequenas no processo de estimação tanto sob o modelo t de Student quanto exponencial potência (veja Figura 3.12).

Este fato indica, portanto, que os estimadores de máxima verossimilhança obtidos a partir de modelos na classe das misturas de escala normal apresentam qualidades robustas.

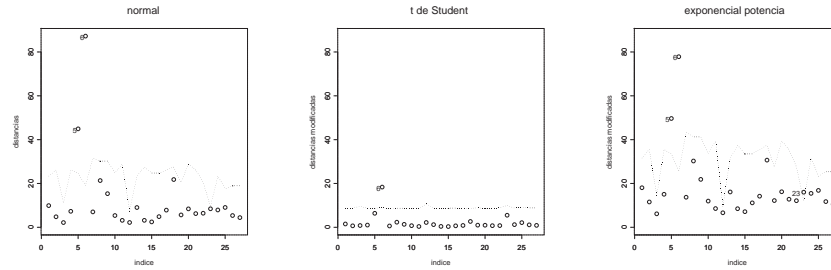


Figura 3.11: Gráficos de índices das distâncias de Mahalanobis para os três modelos ajustados.

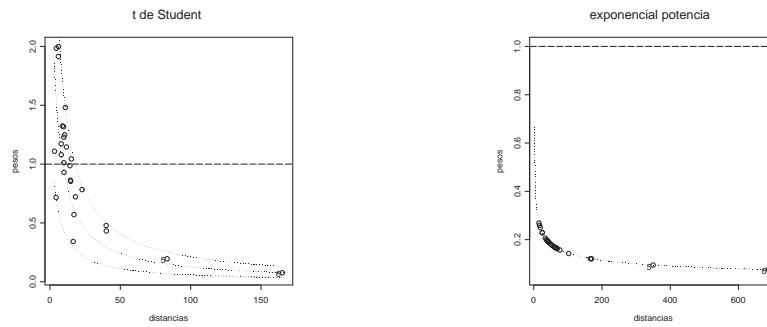


Figura 3.12: Pesos estimados para os modelos t de Student e exponencial potência.

Os gráficos de distâncias transformadas para os três modelos ajustados dados na Figura 3.13 indicam que o modelo t proporciona um melhor ajuste do que os modelos normal ou exponencial potência.

Considere os gráficos de resíduos condicionais $e_i^* = (\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \mathbf{Z}_i\hat{b}_i)/\hat{\phi}^{1/2}$ e das \mathbf{b} -distâncias $U_{b_i} = \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i$ estimadas (sob normalidade) apresentados na Figura 3.14. Pode-se apreciar que existem crias nas ninhadas 5, 6 e 18 que podem ser catalogadas como ϵ -outliers (compare com Figura 3.11). Note que sob normalidade $E(U_{b_i}) = q$, deste modo, é esperado que para o modelo normal, U_{b_i} esteja próximo de q . Isto indica, portanto, que a ninhada 9 pode ser considerada um \mathbf{b} -outlier.

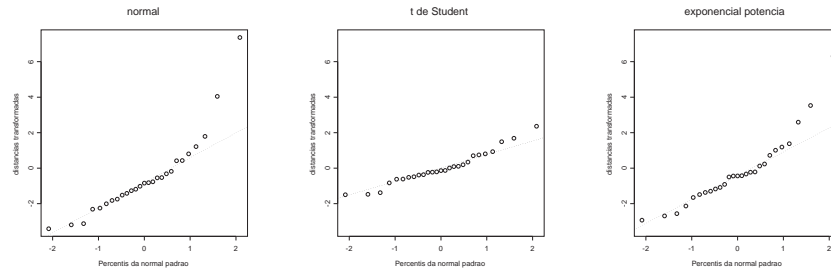


Figura 3.13: Gráficos de distâncias transformadas para os três modelos ajustados.

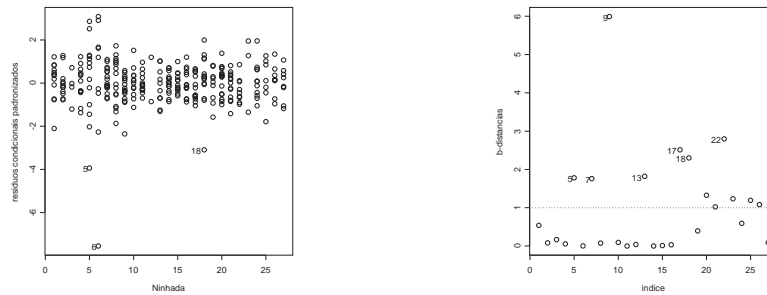


Figura 3.14: Gráficos de resíduos condicionais e b -distâncias para o modelo normal.

A seguir identificamos observações influentes usando o procedimento de influência local. Consideramos os esquemas de perturbação de ponderação de casos, da matriz de escala dos efeitos aleatórios e da resposta.

Ponderação de Casos: apresentamos nas Figuras 3.15-3.16 os gráficos de índices de $B_i(\beta)$ e $B_i(\tau)$ para os três modelos ajustados.

Observamos a partir da Figura 3.15 que a ninhada 6 destaca-se como a que exerce a maior influência em $\hat{\beta}$ no modelo sob erros normais; para os modelos sob erros t de Student e exponencial potência, a influência da ninhada 6 é reduzida. Contudo, nesses modelos nota-se alguma influência das ninhadas 7 e 9.

Com relação à Figura 3.16, podemos observar que para todos os modelos ajustados, as ninhadas 7, 8 e 9 têm influência sobre as estimativas dos parâmetros associados à escala $\hat{\tau}$. Isto parece indicar alguma vulnerabilidade

das estimativas de máxima verossimilhança para os três modelos considerados com relação a \mathbf{b} -outliers (veja Figura 3.14).

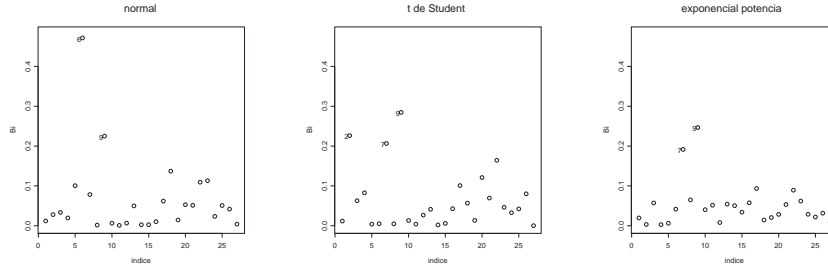


Figura 3.15: Gráficos de índices de B_i para $\hat{\beta}$ sob ponderação de casos.

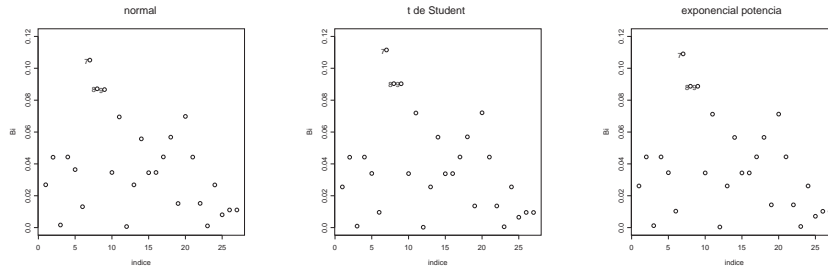


Figura 3.16: Gráficos de índices de B_i para $\hat{\tau}$ sob ponderação de casos.

Perturbação da Matriz de Escala dos Efeitos Aleatórios: para este esquema de perturbação podemos notar que a curvatura normal assume a forma (veja Seção 2.6.1):

$$\begin{aligned} C_{f_Q, h}(\theta) &= -2\mathbf{h}^T \Delta_{\omega_0}^T \ddot{Q}^{-1}(\hat{\theta}) \Delta_{\omega_0} \mathbf{h} \\ &= -2\mathbf{h}^T \Delta_{3, \omega_0}^T \ddot{Q}^{-1}(\hat{\lambda}) \Delta_{3, \omega_0} \mathbf{h} \\ &= C_{f_Q, h}(\lambda), \end{aligned}$$

em que $\ddot{Q}^{-1}(\hat{\lambda})$ é dada na equação 2.18 e Δ_{3, ω_0} é a submatriz de Δ associada ao vetor de parâmetros λ . Deste modo, este esquema de perturbação coloca especial ênfase na influência sobre os parâmetros associados aos efeitos aleatórios.

Baseados nos gráficos de índices de $B_i(\psi)$ para os três modelos ajustados dados na Figura 3.17 podemos notar que a influência exercida pelas ob-

servações é bastante similar para os três modelos considerados e revelam pouca habilidade por parte dos modelos com caudas mais pesadas do que a normal para diminuir a influência das ninhadas 7, 8 e 9. Isto pode ser devido à existência de observações consideradas como \mathbf{b} -outliers (veja Figura 3.14).

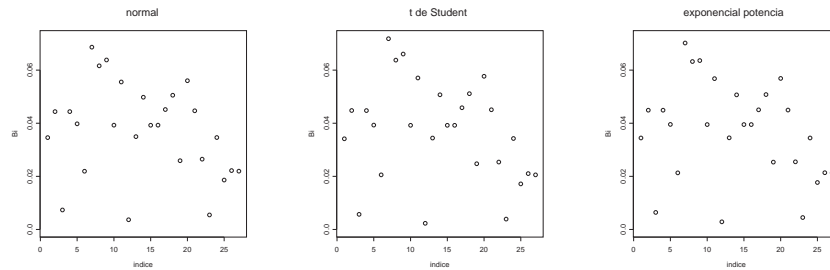


Figura 3.17: Gráficos de índices de B_i para $\hat{\psi}$ sob ponderação da matriz de escala dos efeitos aleatórios.

Perturbação da Resposta: considerando o diagrama de índices de $B_i(\beta)$ dado na Figura 3.18 notamos que para os três modelos ajustados, as ninhadas 3 e 12 exercem forte influência nas medições do peso das crias possivelmente devido ao fato de serem ninhadas de tamanho pequeno, com uma maior vulnerabilidade para o modelo sob erro exponencial potência.

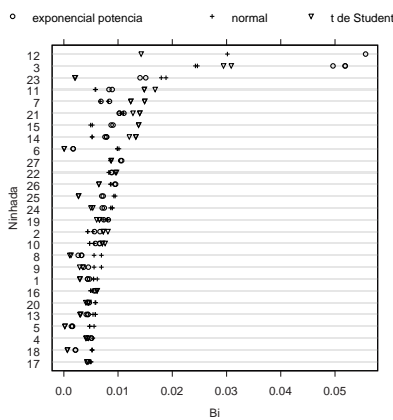


Figura 3.18: Gráfico de índices de B_i para $\hat{\beta}$ sob perturbação da resposta.

3.3 Capacidade Vital Pulmonar

A seguir analisamos o conjunto de dados apresentados por Barnett (1969) em que é avaliada a qualidade de dois instrumentos, um de tipo padrão e outro novo, para medir a capacidade vital pulmonar num grupo de 72 pacientes. O instrumento padrão requer um operador experiente enquanto o instrumento novo é mais fácil de operar. Foram consideradas quatro combinações instrumento-operador (Theobald e Mallinson, 1978):

Instrumento 1: instrumento padrão e operador experiente,

Instrumento 2: instrumento padrão e operador não experiente,

Instrumento 3: instrumento novo e operador experiente e

Instrumento 4: instrumento novo e operador não experiente.

Consideramos o seguinte modelo:

$$\mathbf{Y}_i = \boldsymbol{\mu} + \mathbf{1}z_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 72, \quad (3.2)$$

em que $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{i4})^T$ com Y_{ij} sendo a medição da capacidade vital fornecida pelo instrumento j para o paciente i ; $i = 1, \dots, 72$ e $j = 1, \dots, 4$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_4)^T$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{i4})^T$ é vetor aleatório tetra-dimensional e z_i denota uma variável aleatória com medida de posição 0 e escala ϕ_x .

Vamos supor a seguinte formulação hierárquica:

$$\begin{aligned} \mathbf{Y}_i | z_i &\stackrel{ind}{\sim} SMN_4(\boldsymbol{\mu} + \mathbf{1}z_i, D(\boldsymbol{\phi}); H) \quad e \\ z_i &\stackrel{ind}{\sim} SMN(0, \kappa(v_i)\phi_x; H), \quad i = 1, \dots, 72, \end{aligned}$$

em que $D(\boldsymbol{\phi}) = \text{diag}(\phi_1, \dots, \phi_4)$ e H denota a função de distribuição da variável de mistura e $\kappa(\cdot)$ é uma função positiva.

O modelo dado em (3.2) corresponde a um caso particular do modelo linear com efeitos mistos. Contudo, freqüentemente pode ser interessante considerar tal modelo como um caso particular do modelo de Barnett (Barnett, 1969). Nesse contexto, o modelo (3.2) é conhecido como modelo de Grubbs (Grubbs, 1948, 1973, 1983). Detalhes relativos ao procedimento de estimação por máxima verossimilhança assim como diagnóstico de influência no modelo de Grubbs em (3.2) são dados no Apêndice C.

Análises dos Modelos Ajustados

Consideramos que as variáveis de mistura seguem uma distribuição gama, estável positiva com massa pontual em v_i , ou equivalentemente, que a resposta marginal tem distribuição t de Student, exponencial potência e normal, respectivamente. Foram escolhidos $\nu = 8$ e $\lambda = 0,75$ para os graus de liberdade na distribuição t e para o parâmetro de forma da distribuição exponencial potência, respectivamente. Todas as nossas análises foram obtidas usando a biblioteca para S-PLUS [robssc](http://www.ime.usp.br/~osorio/robssc/) disponível em <http://www.ime.usp.br/~osorio/robssc/>. A seguir apresentamos as estimativas dos parâmetros para o conjunto de dados de capacidade vital pulmonar.

Tabela 3.3: Estimativas de máxima verossimilhança para os três modelos ajustados ao conjunto de dados de capacidade pulmonar.

Parâmetro	Normal		t de Student		Exponencial potência	
	Estimativa	E.P.	Estimativa	E.P.	Estimativa	E.P.
μ_1	22,461	(0,971)	21,846	(0,952)	22,013	(0,965)
μ_2	21,757	(0,945)	21,196	(0,930)	21,355	(0,940)
μ_3	21,486	(0,967)	20,881	(0,950)	21,080	(0,961)
μ_4	21,022	(0,969)	20,562	(0,953)	20,724	(0,964)
ϕ_1	4,998	(1,021)	3,592	(0,872)	1,918	(0,459)
ϕ_2	1,413	(0,666)	0,999	(0,775)	0,538	(0,975)
ϕ_3	4,383	(0,924)	3,392	(0,836)	1,703	(0,422)
ϕ_4	4,633	(0,963)	3,713	(0,895)	1,832	(0,444)

Obtivemos também $\hat{\phi}_x$ como 62,907; 52,351 e 25,725 para os modelos normal, t de Student e exponencial potência, respectivamente. Podemos notar que as inferências para os três modelos ajustados são similares e que os modelos t de Student e exponencial potência apresentam uma tendência a produzir estimativas mais precisas. Usando o gráfico de distâncias de Mahalanobis modificadas (veja Figura 3.19) detectamos os pacientes 4, 25 e 67 como possíveis observações aberrantes. Pode-se notar porém, que para o caso do modelo com erro t , nenhuma observação é catalogada como aberrante.

O diagrama dos pesos estimados contra as distâncias de Mahalanobis dado na Figura 3.20 revela claramente que observações aberrantes recebem ponderações pequenas no processo de estimação por máxima verossimilhança quando distribuições com caudas pesadas são utilizadas.

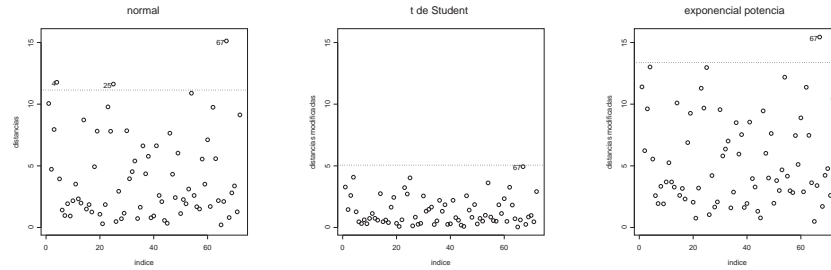


Figura 3.19: Gráficos de índices das distâncias de Mahalanobis para os três modelos ajustados.

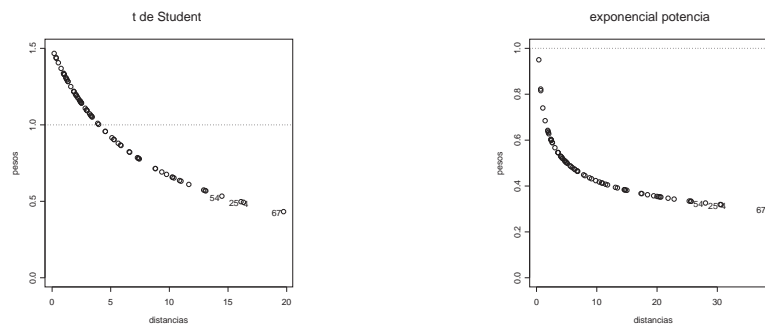


Figura 3.20: Pesos estimados para os modelos t de Student e exponencial potência.

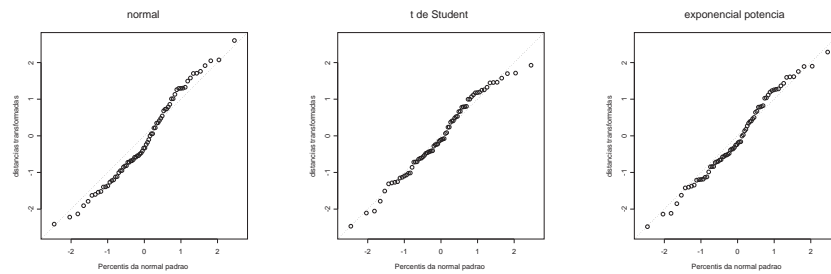


Figura 3.21: Gráficos de distâncias transformadas para os três modelos ajustados.

A Figura 3.21 mostra o gráfico das distâncias transformadas para os três modelos ajustados. Como podemos observar tanto o modelo com erro t de Student quanto o exponencial potência apresentam um melhor ajuste do que o modelo sob erro normal.

A seguir identificamos observações influentes para o conjunto de dados de Barnett usando a curvatura conformal B_i . Focamos nossa atenção em $\theta_1 = (\mu^T, \phi^T)^T$ e consideramos os esquemas de perturbação de ponderação de casos e a perturbação simultânea das medições dos instrumentos. Diagnósticos de influência neste conjunto de dados têm sido considerados por Galea, Bolfarine e de Castro (2002), Galea, Bolfarine e Vilca (2005) e Lachos, Vilca e Galea (2006).

Ponderação de Casos: considerando as Figuras 3.22-3.23, notamos que sob erro normal são identificados os pacientes 4, 25 e 67 como influentes tanto em $\hat{\mu}$ quanto em $\hat{\phi}$. Tais resultados coincidem com os reportados por Lachos, Vilca e Galea (2006). Como esperado, a influência dessas observações é reduzida quando consideramos distribuições com caudas mais pesadas que a normal e, tal como é indicado por Galea, Bolfarine e Vilca (2005) para este conjunto de dados o modelo t com graus de liberdade pequenos acomoda melhor as observações influentes.

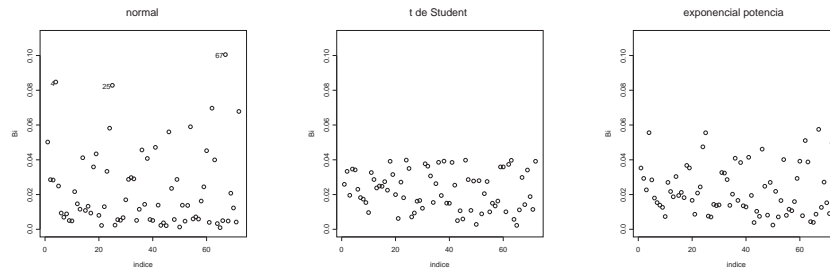


Figura 3.22: Gráficos de índices de B_i para $\hat{\mu}$ sob ponderação de casos.

Perturbação da Resposta: podemos destacar na Figura 3.24 os pacientes 4, 24, 25, 62, 63 e 67 indicando que este esquema de perturbação parece distinguir observações com alta variabilidade entre as medições fornecidas pelos diferentes instrumentos. Contudo, novamente é possível apreciar que para os modelos t de Student e exponencial potência, a influência dessas observações é reduzida.

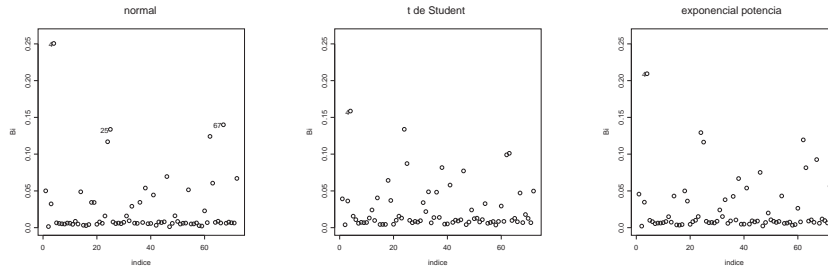


Figura 3.23: Gráficos de índices de B_i para $\hat{\phi}$ sob ponderação de casos.

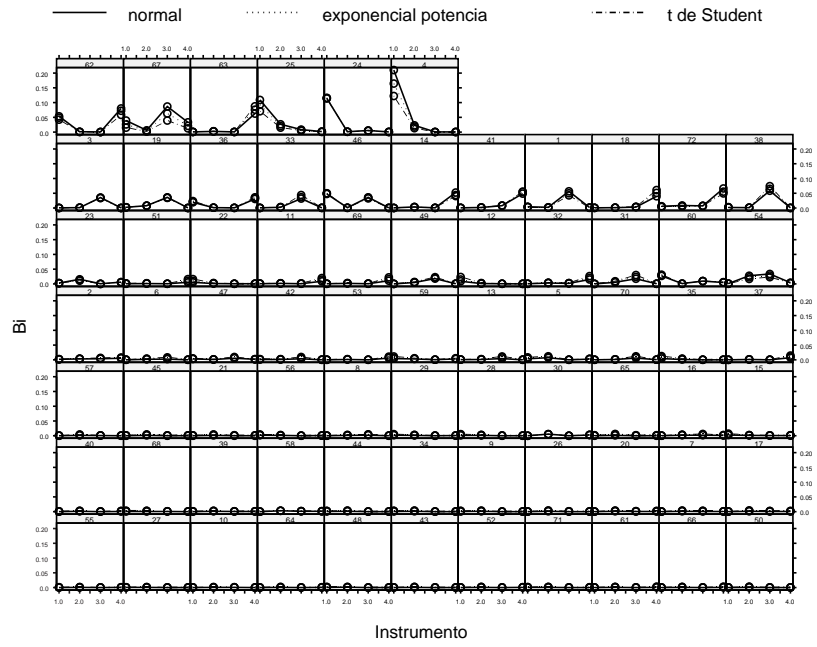


Figura 3.24: Gráfico de índices de B_i para $\hat{\theta}_1$ sob perturbação da resposta.

3.4 Comentários Finais

Através dos estudos de influência considerados podemos notar alguns aspectos de robustez dos modelos elípticos com caudas mais pesadas do que a distribuição normal. É importante salientar a capacidade de tais modelos para atribuir ponderações pequenas no processo de estimação para as observações aberrantes. Tais resultados concordam com as considerações que a este respeito apresentam Lange, Little e Taylor (1989) e Kowalski, Mendoza-Blanco, Tu e Gleser (1999) assim como com os comentários de Pينهiro, Liu e Wu (2001) e Savalli, Paula e Cysneiros (2006) para modelos lineares com efeitos mistos e Galea, Bolfarine e Vilca (2005) para modelos de calibração comparativa estrutural. Contudo, como é notado por alguns autores (veja por exemplo, Lange, Little e Taylor, 1989) modelos baseados em distribuições de contornos elípticos estão longe de serem uma panacéia, de fato eles ainda podem sofrer o efeito de observações aberrantes e/ou influentes.

Nas análises consideradas focamos nossa atenção em alguns membros típicos da classe de distribuições elípticas, porém várias outras alternativas podem ser úteis para modelagem robusta, tais como as distribuições slash e normal contaminada. Por outro lado, esses estudos de sensibilidade também revelam a necessidade de considerar procedimentos de diagnóstico para modelos desenvolvidos usando a classe de distribuições de contornos elípticos.

Capítulo 4

Considerações Finais

4.1 Conclusões

Neste trabalho investigamos diagnósticos de influência em modelos com efeitos mistos sob distribuições de contornos elípticos. Modelos baseados nesta classe de distribuições representam alternativas robustas para modelos desenvolvidos sob normalidade.

No Capítulo 1 foram considerados dois enfoques para a inclusão de distribuições de contornos elípticos no modelo linear misto e revisamos a estimação por máxima verossimilhança para cada um desses enfoques. Mostramos que a densidade marginal para a resposta obtida a partir do modelo elíptico linear misto em sua formulação hierárquica também pertence à classe das elípticas, simplificando desse modo a inferência estatística. Discutimos também procedimentos que permitem facilitar a implementação computacional do algoritmo EM para o modelo elíptico linear misto em sua formulação hierárquica.

No Capítulo 2 usamos o enfoque de influência proposto por Cook (1986) para estudar a influência local no modelo elíptico linear misto. Devido à complexidade da função de verossimilhança de dados observados para o modelo em sua formulação hierárquica optamos por desenvolver a influência neste caso baseados na função Q -afastamento (Zhu e Lee, 2001). Para cada um dos enfoques considerados para a inclusão de distribuições de contornos elípticos obtivemos expressões em forma fechada para a curvatura normal considerando vários esquemas de perturbação. Consideramos também, tanto para o modelo em sua formulação marginal quanto hierárquica a relação

entre o método de alavanca generalizado (Wei, Hu e Fung, 1998) e o procedimento de influência local. Com o intuito de obter a matriz de alavancas generalizadas no modelo elíptico linear misto sob o enfoque hierárquico estendemos a definição dada por Wei et al. (1998) para manipular modelos com dados incompletos. Salientamos, todavia, que a aplicação da definição proposta não está restrita somente a modelos lineares com efeitos mistos. Por outro lado, freqüentemente tem sido mencionado que a análise de influência local requer o cálculo de autovalores e autovetores de uma matriz de dimensão potencialmente grande. Usamos a estrutura do problema para propor uma estratégia equivalente, porém de custo computacional muito menor usando a decomposição de valor singular.

Ilustramos a metodologia desenvolvida mediante alguns exemplos, cujos resultados são apresentados no Capítulo 3. A partir deles pode-se apreciar a habilidade dos modelos baseados em distribuições com caudas mais pesadas que a normal para acomodar observações aberrantes. Notamos também que as observações diagnosticadas como influentes exercem uma influência menor sob esse tipo de distribuições indicando desse modo que correspondem a boas alternativas para modelagem robusta. É importante destacar que os diagnósticos desenvolvidos mediante o procedimento de influência local não somente permitem a identificação de observações influentes, mas também permitem compreender melhor o processo de modelagem.

O diagnóstico de influência no modelo de Grubbs (veja Apêndice C) indica que os resultados desenvolvidos como parte deste trabalho para o modelo linear com efeitos mistos podem ser aplicados em diversas situações.

4.2 Perspectivas de Trabalhos Futuros

Este trabalho centra sua atenção no modelo linear com efeitos mistos. Sentimos que baseados no enfoque hierárquico para a inclusão de distribuições na classe das elípticas é possível estender os resultados apresentados para modelagem robusta assim como diagnóstico de influência para modelos mistos não-lineares (veja Vonesh e Chinchilli, 1997, Pinheiro e Bates, 2000 e Lee e Xu, 2004).

Um dos resultados apresentados neste trabalho indica que a densidade marginal obtida a partir do modelo em sua formulação hierárquica pertence à classe de distribuições de misturas de escala normal. Deste modo, pode ser interessante considerar métodos para aproximar a integral (1.13) com o

intuito de estudar a estimação e influência nesse modelo. Estudos comparativos entre esse modelo e os apresentados neste trabalho também podem ser úteis.

Vários autores têm considerado a estimação dos parâmetros de forma em modelos desenvolvidos usando distribuições de contornos elípticos (veja por exemplo, Fernández e Steel, 1999 e Pinheiro, Liu e Wu, 2001). Um enfoque alternativo corresponde à estimação da função geradora de densidade g no caso do modelo em sua formulação marginal ou da função de distribuição H para a formulação hierárquica mediante métodos não-paramétricos (Hodgson, Linton e Vorkink, 2002). Acreditamos que estender os modelos considerados neste trabalho usando tal proposta pode trazer bastante flexibilidade na modelagem de modelos com efeitos mistos.

4.3 Desenvolvimento de Software

O sucesso de qualquer técnica estatística depende em grande parte da disponibilidade de software para realizar as análises. Em conjunto com este trabalho disponibilizamos duas bibliotecas para S-PLUS em que podem ser realizadas análises no modelo linear com efeitos mistos e no modelo de Grubbs sob distribuições de contornos elípticos. O estágio de desenvolvimento dessas bibliotecas ainda é muito preliminar; pretendemos estender as facilidades disponíveis nessas bibliotecas para manipular modelos mais gerais. Temos o objetivo de submeter este software para a coleção de código S do repositório [StatLib](#).

O código em C do software desenvolvido como parte deste trabalho utiliza código FORTRAN das bibliotecas de programas BLAS e LINPACK. Temos o propósito de reescrever as partes críticas das rotinas desenvolvidas com o intuito de utilizar código FORTRAN mais eficiente disponível na biblioteca LAPACK (Anderson et al., 1995) assim como utilizar a biblioteca PORT (Fox, Hall e Schryer, 1978) para otimização não restrita.

Por outro lado, notamos que tanto os modelos com efeitos mistos quanto as técnicas de diagnóstico discutidas requerem métodos versáteis para visualizar a informação em forma gráfica. Acreditamos que o formato de apresentação em painéis da biblioteca Trellis (Becker, Cleveland e Shyu, 1996) disponível no sistema S-PLUS pode ajudar neste respeito.

Finalmente, temos o objetivo de escrever uma versão do código desenvolvido para seu uso com R (R Development Core Team, 2005).

Apêndice A

Cálculo dos Diferenciais

Neste apêndice apresentamos os diferenciais a partir dos quais obtemos as medidas de diagnóstico para o modelo elíptico linear com efeitos mistos. Tais medidas são obtidas mediante a técnica de diferenciação matricial proposta por Neudecker e outros autores. Este procedimento é descrito em Nel (1980) e completamente documentado em Magnus e Neudecker (1988). Uma característica de interesse do método de diferenciação de Magnus e Neudecker é que permite o cálculo eficiente de derivadas matriciais.

A aplicação deste procedimento em diferentes tópicos de Estatística e Econometria é discutido em Magnus e Neudecker (1988) e na derivação da curvatura normal requerida para o método de influência local, por Liu (2000, 2002), Díaz-García, Galea e Leiva-Sánchez (2003) e Osorio, Paula e Galea (2006).

A.1 Derivação da Curvatura no Modelo Linear Misto Marginal

No modelo elíptico linear misto marginal o logaritmo da função de verossimilhança é dado por

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

em que $L_i(\boldsymbol{\theta})$ é dado em (1.5). A seguir obtemos os diferenciais $d_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$ e $d_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta})$ para esse modelo.

Matriz de Informação Observada

Usando resultados de diferenciação de matrizes obtemos que

$$d_{\theta} L(\boldsymbol{\theta}) = \sum_{i=1}^M d_{\theta} L_i(\boldsymbol{\theta}) \quad \text{e} \quad d_{\theta}^2 L(\boldsymbol{\theta}) = \sum_{i=1}^M d_{\theta}^2 L_i(\boldsymbol{\theta}).$$

Os diferenciais $d_{\theta} L_i(\boldsymbol{\theta})$ e $d_{\theta}^2 L_i(\boldsymbol{\theta})$ para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ são apresentados a seguir.

Tomando diferenciais de $L_i(\boldsymbol{\theta})$ com relação a $\boldsymbol{\beta}$ obtemos

$$d_{\boldsymbol{\beta}} L_i(\boldsymbol{\theta}) = -2W_g(u_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta} \quad \text{e} \quad (\text{A.1})$$

$$d_{\boldsymbol{\beta}}^2 L_i(\boldsymbol{\theta}) = 2(d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} \{W_g(u_i) \mathbf{V}_i + 2W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^T\} \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta}, \quad (\text{A.2})$$

em que $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$. A partir de (A.1) segue que

$$d_{\alpha\boldsymbol{\beta}}^2 L_i(\boldsymbol{\theta}) = 2\mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \{W_g(u_i) \mathbf{V}_i + W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^T\} \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta} \quad \text{e}$$

$$d_{\lambda\boldsymbol{\beta}}^2 L_i(\boldsymbol{\theta}) = 2\mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \{W_g(u_i) \mathbf{V}_i + W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^T\} \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta}.$$

Agora, diferenciando $L_i(\boldsymbol{\theta})$ com relação a $\boldsymbol{\alpha}$, temos

$$d_{\boldsymbol{\alpha}} L_i(\boldsymbol{\theta}) = -\frac{1}{2} \text{tr} \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i - W_g(u_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \mathbf{r}_i \quad \text{e} \quad (\text{A.3})$$

$$\begin{aligned} d_{\boldsymbol{\alpha}}^2 L_i(\boldsymbol{\theta}) &= \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i - \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} d^2 \boldsymbol{\Sigma}_i \\ &\quad + \mathbf{r}_i^T \mathbf{V}_i^{-1} \{W'_g(u_i) (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) - W_g(u_i) d^2 \boldsymbol{\Sigma}_i \\ &\quad + W_g(u_i) (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i + W_g(u_i) (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i\} \mathbf{V}_i^{-1} \mathbf{r}_i. \end{aligned} \quad (\text{A.4})$$

Usando (A.3) fica que

$$\begin{aligned} d_{\lambda\boldsymbol{\alpha}}^2 L_i(\boldsymbol{\theta}) &= -\frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i \\ &\quad + \mathbf{r}_i^T \mathbf{V}_i^{-1} \{W'_g(u_i) \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \\ &\quad + W_g(u_i) \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) + W_g(u_i) (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T\} \mathbf{V}_i^{-1} \mathbf{r}_i. \end{aligned}$$

Finalmente, o primeiro e o segundo diferencial de $L_i(\boldsymbol{\theta})$ com relação a $\boldsymbol{\lambda}$ são dados pelas expressões,

$$d_{\boldsymbol{\lambda}} L_i(\boldsymbol{\theta}) = -\frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T - W_g(u_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \quad \text{e}$$

$$\begin{aligned} d_{\boldsymbol{\lambda}}^2 L_i(\boldsymbol{\theta}) &= \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) \mathbf{Z}_i^T - \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i (d^2 \Psi) \mathbf{Z}_i^T \\ &\quad + \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \{W'_g(u_i) (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\Psi) - W_g(u_i) d^2 \Psi \\ &\quad + W_g(u_i) (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i d\Psi + W_g(u_i) (d\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i d\Psi\} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i. \end{aligned}$$

Avaliando esses diferenciais em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e usando os teoremas de identificação dados em Magnus e Neudecker (1988) obtemos a matriz de informação observada dada em (2.10).

Esquemas de Perturbação

Para cada um dos esquemas de perturbação obtemos o diferencial $d_{\widehat{\boldsymbol{\theta}}|\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})$ e posteriormente conseguimos a matriz $\boldsymbol{\Delta} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}$ por meio da avaliação de $d_{\widehat{\boldsymbol{\theta}}|\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})$ em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

Ponderação de Casos

Para este esquema de perturbação temos que

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= q_i(d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) d\omega_i, \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -\frac{1}{2}\{\text{tr } \mathbf{V}_i^{-1}(d\boldsymbol{\Sigma}_i) - q_i\mathbf{r}_i^T \mathbf{V}_i^{-1}(d\boldsymbol{\Sigma}_i)\mathbf{V}_i^{-1}\mathbf{r}_i\} d\omega_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -\frac{1}{2}\{\text{tr } \mathbf{V}_i^{-1}\mathbf{Z}_i(d\boldsymbol{\Psi})\mathbf{Z}_i^T - q_i\mathbf{r}_i^T \mathbf{V}_i^{-1}\mathbf{Z}_i(d\boldsymbol{\Psi})\mathbf{Z}_i^T \mathbf{V}_i^{-1}\mathbf{r}_i\} d\omega_i, \end{aligned}$$

com $q_i = q_i(\boldsymbol{\theta})$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}$ para $i = 1, \dots, M$.

Perturbação da Matriz de Escala

Tomando diferenciais de $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ com relação a $\boldsymbol{\theta}$ e ω_i , obtemos

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -2\{W_g(u_{i\omega}) + \omega_i u_i W'_g(u_{i\omega})\}(d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i d\omega_i, \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -\{W_g(u_{i\omega}) + \omega_i u_i W'_g(u_{i\omega})\}\mathbf{r}_i^T \mathbf{V}_i^{-1}(d\boldsymbol{\Sigma}_i)\mathbf{V}_i^{-1}\mathbf{r}_i d\omega_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -\{W_g(u_{i\omega}) + \omega_i u_i W'_g(u_{i\omega})\}\mathbf{r}_i^T \mathbf{V}_i^{-1}\mathbf{Z}_i(d\boldsymbol{\Psi})\mathbf{Z}_i^T \mathbf{V}_i^{-1}\mathbf{r}_i d\omega_i, \end{aligned}$$

em que $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}$ e $u_{i\omega} = \omega_i u_i$ com $u_i = \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i$, $i = 1, \dots, M$.

Perturbação das Variáveis Explicativas

Diferenciando $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ com relação a $\boldsymbol{\theta}$ e ω_i , fica

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -s_i 2W_g(u_{i\omega})(d\boldsymbol{\beta})^T (\mathbf{c}_t^T \boldsymbol{\beta} \mathbf{X}_{i\omega} - \mathbf{c}_t \mathbf{r}_{i\omega}^T) \mathbf{V}_i^{-1} d\omega_i \\ &\quad + s_i 4W'_g(u_{i\omega}) \mathbf{c}_t^T \boldsymbol{\beta} (d\boldsymbol{\beta})^T \mathbf{X}_{i\omega}^T \mathbf{V}_i^{-1} \mathbf{r}_{i\omega} \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} d\omega_i, \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= 2s_i \mathbf{c}_t^T \boldsymbol{\beta} \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1}(d\boldsymbol{\Sigma}_i)\mathbf{V}_i^{-1}\{W_g(u_{i\omega})\mathbf{V}_i + W'_g(u_{i\omega})\mathbf{r}_{i\omega}\mathbf{r}_{i\omega}^T\}\mathbf{V}_i^{-1} d\omega_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= 2s_i \mathbf{c}_t^T \boldsymbol{\beta} \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1}\mathbf{Z}_i(d\boldsymbol{\Psi})\mathbf{Z}_i^T \mathbf{V}_i^{-1}\{W_g(u_{i\omega})\mathbf{V}_i + W'_g(u_{i\omega})\mathbf{r}_{i\omega}\mathbf{r}_{i\omega}^T\}\mathbf{V}_i^{-1} d\omega_i, \end{aligned}$$

em que $\mathbf{X}_{i\omega} = \mathbf{X}_i + s_i \boldsymbol{\omega}_i \mathbf{c}_t^T$, com s_i um fator de escala, $\boldsymbol{\omega}_i$ vetor de perturbação $n_i \times 1$ e \mathbf{c}_t vetor p -dimensional com 1 na t -ésima posição e zeros nas restantes. Aqui $\mathbf{r}_{i\omega} = \mathbf{r}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i$ e $u_{i\omega} = \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} \mathbf{r}_{i\omega}$ com $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$.

Perturbação da Resposta

Tomando diferenciais de $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ com relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos

$$\begin{aligned} d_{\beta \boldsymbol{\omega}_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -2(d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} \{W_g(u_{i\omega}) \mathbf{V}_i + 2W'_g(u_{i\omega}) \mathbf{r}_{i\omega} \mathbf{r}_{i\omega}^T\} \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i, \\ d_{\alpha \boldsymbol{\omega}_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -2\mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \{W_g(u_{i\omega}) \mathbf{V}_i + W'_g(u_{i\omega}) \mathbf{r}_{i\omega} \mathbf{r}_{i\omega}^T\} \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i \quad \text{e} \\ d_{\lambda \boldsymbol{\omega}_i}^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -2\mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \{W_g(u_{i\omega}) \mathbf{V}_i + W'_g(u_{i\omega}) \mathbf{r}_{i\omega} \mathbf{r}_{i\omega}^T\} \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i, \end{aligned}$$

em que $u_{i\omega} = \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} \mathbf{r}_{i\omega}$ e $\mathbf{r}_{i\omega} = \mathbf{r}_i + \boldsymbol{\omega}_i$, com $\boldsymbol{\omega}_i$ vetor de perturbação $n_i \times 1$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, $i = 1, \dots, M$.

A.2 Método de Alavanca Generalizado no Modelo Linear Misto Marginal

Considere $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$, logo o vetor de posição $\boldsymbol{\mu}$ fica dado como $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_M^T)^T$. Desse modo

$$d_{\beta} \boldsymbol{\mu}_i = \mathbf{X}_i d\boldsymbol{\beta}, \quad d_{\alpha} \boldsymbol{\mu}_i = \mathbf{0} \quad \text{e} \quad d_{\lambda} \boldsymbol{\mu}_i = \mathbf{0}, \quad i = 1, \dots, M.$$

E tomando diferenciais de $L(\boldsymbol{\theta})$ com relação a $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ e \mathbf{Y}_i , obtemos

$$\begin{aligned} d_{\beta \mathbf{Y}_i}^2 L(\boldsymbol{\theta}) &= -2(d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} \{W_g(u_i) \mathbf{V}_i + 2W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^T\} \mathbf{V}_i^{-1} d\mathbf{Y}_i, \\ d_{\alpha \mathbf{Y}_i}^2 L(\boldsymbol{\theta}) &= -2\mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \{W_g(u_i) \mathbf{V}_i + W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^T\} \mathbf{V}_i^{-1} d\mathbf{Y}_i \quad \text{e} \\ d_{\lambda \mathbf{Y}_i}^2 L(\boldsymbol{\theta}) &= -2\mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \{W_g(u_i) \mathbf{V}_i + W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^T\} \mathbf{V}_i^{-1} d\mathbf{Y}_i, \end{aligned}$$

em que $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$.

Avaliando os diferenciais $d_{\beta \mathbf{Y}_i}^2 L(\boldsymbol{\theta})$, $d_{\alpha \mathbf{Y}_i}^2 L(\boldsymbol{\theta})$ e $d_{\lambda \mathbf{Y}_i}^2 L(\boldsymbol{\theta})$ em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ obtemos $\partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \mathbf{Y}^T |_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$.

A.3 Derivação da Curvatura no Modelo Linear Misto Hierárquico

Nesta seção derivamos a curvatura normal baseada na função Q -afastamento para o modelo linear com efeitos mistos usando a formulação hierárquica dada em (1.4), no caso em que \mathbf{b}_i e v_i são considerados não observáveis assim como no caso em que os efeitos aleatórios são integrados.

A.3.1 Derivação da Curvatura quando \mathbf{b}_i e v_i são não observáveis

Considerando o vetor de dados completos $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{b}^T, \mathbf{v}^T)^T$ em que $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ representa o vetor de respostas observadas e, quando tanto os efeitos aleatórios $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_M^T)^T$ quanto as variáveis de mistura $\mathbf{v} = (v_1, \dots, v_M)^T$ são considerados não observáveis, obtemos que a função Q para o modelo linear misto hierárquico dado em (1.4) assume a forma

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}),$$

em que $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) + Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$ com $Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\lambda}|\hat{\boldsymbol{\theta}})$, dadas em (1.15) e (1.16), respectivamente.

Matriz Hessiana, \ddot{Q}

Usando resultados de diferenciação de matrizes obtemos que

$$d_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M d_{\boldsymbol{\theta}} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) \quad \text{e} \quad d_{\boldsymbol{\theta}}^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M d_{\boldsymbol{\theta}}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}).$$

Os diferenciais $d_{\boldsymbol{\theta}} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ e $d_{\boldsymbol{\theta}}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ são apresentados a seguir.

Tomando diferenciais de $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\beta}$ obtemos

$$d_{\boldsymbol{\beta}} Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) = \hat{\kappa}_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \hat{\mathbf{b}}_i)^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta} \quad \text{e} \quad (\text{A.5})$$

$$d_{\boldsymbol{\beta}}^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) = -\hat{\kappa}_i (d\boldsymbol{\beta})^T \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta}. \quad (\text{A.6})$$

Usando (A.5), segue que

$$d_{\alpha\beta}^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\hat{\boldsymbol{\theta}}) = -\hat{\kappa}_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \hat{\mathbf{b}}_i)^T \boldsymbol{\Sigma}_i^{-1} (d\boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta}.$$

Agora, diferenciando $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\alpha}$, temos

$$\begin{aligned} d_{\boldsymbol{\alpha}} Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}) &= -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Sigma}_i - \mathbf{Z}_i \widehat{\boldsymbol{\Omega}}_i \mathbf{Z}_i^T) \\ &\quad + \frac{\widehat{\kappa}_i}{2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \widehat{\mathbf{b}}_i)^T \boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \widehat{\mathbf{b}}_i) \quad \text{e} \\ d_{\boldsymbol{\alpha}}^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}) &= \frac{1}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \{ (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} d \boldsymbol{\Sigma}_i - d^2 \boldsymbol{\Sigma}_i \} \\ &\quad - \frac{\widehat{\kappa}_i}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \{ (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} d \boldsymbol{\Sigma}_i + (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} d \boldsymbol{\Sigma}_i - d^2 \boldsymbol{\Sigma}_i \} \boldsymbol{\Sigma}_i^{-1} \mathbf{M}_i. \end{aligned}$$

Finalmente, o primeiro e o segundo diferencial de $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\lambda}$ são dados pelas expressões,

$$\begin{aligned} d_{\boldsymbol{\lambda}} Q_{2i}(\boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}) &= -\frac{1}{2} \text{tr} \boldsymbol{\Psi}^{-1} (d \boldsymbol{\Psi}) \boldsymbol{\Psi}^{-1} (\boldsymbol{\Psi} - \widehat{\mathbf{N}}_i) \quad \text{e} \\ d_{\boldsymbol{\lambda}}^2 Q_{2i}(\boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}) &= \frac{1}{2} \text{tr} \boldsymbol{\Psi}^{-1} \{ (d \boldsymbol{\Psi}) \boldsymbol{\Psi}^{-1} d \boldsymbol{\Psi} - d^2 \boldsymbol{\Psi} \} \\ &\quad - \frac{\widehat{\kappa}_i}{2} \text{tr} \boldsymbol{\Psi}^{-1} \{ (d \boldsymbol{\Psi}) \boldsymbol{\Psi}^{-1} d \boldsymbol{\Psi} + (d \boldsymbol{\Psi}) \boldsymbol{\Psi}^{-1} d \boldsymbol{\Psi} - d^2 \boldsymbol{\Psi} \} \boldsymbol{\Psi}^{-1} \widehat{\mathbf{N}}_i, \end{aligned}$$

em que $\mathbf{M}_i = \widehat{\kappa}_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \widehat{\mathbf{b}}_i) (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \widehat{\mathbf{b}}_i)^T + \mathbf{Z}_i \widehat{\boldsymbol{\Omega}}_i \mathbf{Z}_i^T$, $\widehat{\mathbf{N}}_i = \widehat{\kappa}_i \widehat{\mathbf{b}}_i \widehat{\mathbf{b}}_i^T + \widehat{\boldsymbol{\Omega}}_i$, para $i = 1, \dots, M$, com $\widehat{\kappa}_i$, $\widehat{\mathbf{b}}_i$ e $\widehat{\boldsymbol{\Omega}}_i$ dados em (1.17) e (1.18). Avaliando esses diferenciais em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e usando os teoremas de identificação dados em Magnus e Neudecker (1988) obtemos a matriz de hessiana, $\ddot{\mathbf{Q}}$ dada em (2.18).

Esquemas de Perturbação

Para cada um dos esquemas de perturbação calculamos o diferencial $d_{\boldsymbol{\theta}\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$ e obtemos a matriz $\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c) / \partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T |_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}$ avaliando $d_{\boldsymbol{\theta}\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$ em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Para os resultados obtidos nesta seção assumimos que é possível trocar as operações de integração e diferenciação.

Ponderação de Casos

Para este esquema de perturbação temos que

$$\begin{aligned} d_{\beta \boldsymbol{\omega}_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c) &= \kappa^{-1}(v_i) (d \boldsymbol{\beta})^T \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i) d \boldsymbol{\omega}_i, \\ d_{\alpha \boldsymbol{\omega}_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c) &= -\frac{1}{2} \{ \text{tr} \boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) - \kappa^{-1}(v_i) \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i \} d \boldsymbol{\omega}_i \quad \text{e} \\ d_{\lambda \boldsymbol{\omega}_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c) &= -\frac{1}{2} \{ \text{tr} \boldsymbol{\Psi}^{-1} d \boldsymbol{\Psi} - \kappa^{-1}(v_i) \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} (d \boldsymbol{\Psi}) \boldsymbol{\Psi}^{-1} \mathbf{b}_i \} d \boldsymbol{\omega}_i, \end{aligned}$$

em que $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$ para $i = 1, \dots, M$.

Perturbação da Matriz de Escala dos Efeitos Aleatórios

Tomando diferenciais de $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e ω_i , obtemos

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \mathbf{0}, & d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \mathbf{0} \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{\kappa^{-1}(v_i)}{2} \mathbf{b}_i^T \boldsymbol{\Psi}^{-1} (d \boldsymbol{\Psi}) \boldsymbol{\Psi}^{-1} \mathbf{b}_i d \omega_i, \end{aligned}$$

para $i = 1, \dots, M$.

Perturbação das Variáveis Explicativas

Diferenciando $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e ω_i , fica

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) s_i (d \boldsymbol{\beta})^T (\mathbf{c}_t^T \boldsymbol{\beta} \mathbf{X}_{i\omega}^T - \mathbf{c}_t \boldsymbol{\epsilon}_{i\omega}^T) \boldsymbol{\Sigma}_i^{-1} d \boldsymbol{\omega}_i \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\epsilon}_{i\omega}^T \boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} d \boldsymbol{\omega}_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \mathbf{0}, \end{aligned}$$

em que $\mathbf{X}_{i\omega} = \mathbf{X}_i + s_i \omega_i \mathbf{c}_t^T$, com s_i um fator de escala, ω_i vetor de perturbação $n_i \times 1$ e \mathbf{c}_t vetor p -dimensional com 1 na t -ésima posição e zeros nas restantes. Aqui $\boldsymbol{\epsilon}_{i\omega} = \boldsymbol{\epsilon}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \omega_i$ e $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$ para $i = 1, \dots, M$.

Perturbação da Resposta

Tomando diferenciais de $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e ω_i , obtemos

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) (d \boldsymbol{\beta})^T \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} d \boldsymbol{\omega}_i, & d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \mathbf{0} \quad \text{e} \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) \boldsymbol{\epsilon}_{i\omega}^T \boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} d \boldsymbol{\omega}_i, \end{aligned}$$

em que $\boldsymbol{\epsilon}_{i\omega} = \boldsymbol{\epsilon}_i + \boldsymbol{\omega}_i$, com ω_i vetor de perturbação $n_i \times 1$ e $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, $i = 1, \dots, M$.

Perturbação da Matriz de Correlação dentro dos Indivíduos

Mediante diferenciação podemos mostrar que

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) (d \boldsymbol{\beta})^T (\boldsymbol{\epsilon}^T \mathbf{D}_i^{-1} \mathbf{C}_{i\omega}^{-1} \otimes \mathbf{X}_i^T \mathbf{D}_i^{-1} \mathbf{C}_{i\omega}^{-1}) \mathbf{S}_{n_i} d \boldsymbol{\omega}_i \quad \text{e} \\ d_{\alpha_1\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{C}_{i\omega}^{-1} \mathbf{D}_i^{-1} (d \mathbf{D}_i) \mathbf{D}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \mathbf{D}_i^{-1} \mathbf{C}_{i\omega}^{-1}) \mathbf{S}_{n_i} d \boldsymbol{\omega}_i \\ &\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{C}_{i\omega}^{-1} \mathbf{D}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \mathbf{D}_i^{-1} (d \mathbf{D}_i) \mathbf{D}_i^{-1} \mathbf{C}_{i\omega}^{-1}) \mathbf{S}_{n_i} d \boldsymbol{\omega}_i, \end{aligned}$$

$$\begin{aligned}
d_{\alpha_2 \omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \frac{1}{2} \text{vec}^T (\mathbf{C}_{i\omega}^{-1} (d \mathbf{C}_i) \mathbf{C}_{i\omega}^{-1}) \mathbf{S}_{n_i} d \boldsymbol{\omega}_i \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{C}_{i\omega}^{-1} (d \mathbf{C}_i) \mathbf{C}_{i\omega}^{-1} \mathbf{D}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \mathbf{D}_i^{-1} \mathbf{C}_{i\omega}^{-1}) \mathbf{S}_{n_i} d \boldsymbol{\omega}_i \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{C}_{i\omega}^{-1} \mathbf{D}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \mathbf{D}_i^{-1} \mathbf{C}_{i\omega}^{-1} (d \mathbf{C}_i) \mathbf{C}_{i\omega}^{-1}) \mathbf{S}_{n_i} d \boldsymbol{\omega}_i \quad \text{e} \\
d_{\lambda \omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \mathbf{0},
\end{aligned}$$

em que $\boldsymbol{\Sigma}_{i\omega} = \mathbf{D}_i \mathbf{C}_{i\omega} \mathbf{D}_i$ com $\mathbf{C}_{i\omega} = \mathbf{C}_i + \mathbf{W}_i$ sendo uma matriz de correlação e \mathbf{W}_i matriz de perturbação simétrica $n_i \times n_i$. Aqui \mathbf{S}_{n_i} é matriz de duplicação que satisfaz $\text{vec}(\mathbf{W}_i) = \mathbf{S}_{n_i} \boldsymbol{\omega}_i$ em que $\boldsymbol{\omega}_i = \text{vech}(\mathbf{W}_i)$, $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, para $i = 1, \dots, M$.

Perturbação das Funções de Escala dentro dos Indivíduos

Diferenciando $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos

$$\begin{aligned}
d_{\beta \omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) (d \boldsymbol{\beta})^T (\boldsymbol{\epsilon}_i^T \otimes \mathbf{X}_i^T \mathbf{W}_i \boldsymbol{\Sigma}_i^{-1}) \mathbf{S}_{n_i}^* d \boldsymbol{\omega}_i, \\
d_{\alpha \omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i \mathbf{W}_i \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T) \mathbf{S}_{n_i}^* d \boldsymbol{\omega}_i \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \mathbf{W}_i \boldsymbol{\Sigma}_i^{-1} (d \boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1}) \mathbf{S}_{n_i}^* d \boldsymbol{\omega}_i \quad \text{e} \\
d_{\lambda \omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \mathbf{0},
\end{aligned}$$

em que $\mathbf{S}_{n_i}^*$ é uma matriz de duplicação tal que $\text{vec}(\mathbf{W}_i) = \mathbf{S}_{n_i}^* \boldsymbol{\omega}_i$ com $\mathbf{W}_i = \text{diag}(\boldsymbol{\omega}_i)$ matriz de perturbações $n_i \times n_i$, $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, $i = 1, \dots, M$.

A.3.2 Derivação da Curvatura quando os v_i são não observáveis

A seguir consideramos o vetor de dados completos $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{v}^T)^T$ em que $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ representa o vetor de respostas e $\mathbf{v} = (v_1, \dots, v_M)^T$ denota as variáveis de mistura, as quais são consideradas não observáveis, isto é, usamos a formulação hierárquica dada em (1.12) em que os efeitos aleatórios têm sido retirados via integração. Neste caso, a função Q adota a forma

$$Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}),$$

em que $Q_i(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})$ é dada na equação (1.21).

Matriz Hessiana, \ddot{Q}

Diferenciando a esperança condicional do logaritmo da função de verossimilhança para dados completos obtemos que

$$d_{\theta} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M d_{\theta} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) \quad \text{e} \quad d_{\theta}^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M d_{\theta}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}).$$

A seguir apresentamos os diferenciais $d_{\theta} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ e $d_{\theta}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$.

Tomando diferenciais de $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\beta}$ obtemos

$$d_{\beta} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \hat{\kappa}_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta} \quad \text{e} \quad (\text{A.7})$$

$$d_{\beta}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = -\hat{\kappa}_i (d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta}, \quad (\text{A.8})$$

e diferenciando (A.7) com relação a $\boldsymbol{\alpha}$ e $\boldsymbol{\lambda}$, segue que

$$\begin{aligned} d_{\alpha\beta}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= -\hat{\kappa}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta} \quad \text{e} \\ d_{\lambda\beta}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= -\hat{\kappa}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i d\boldsymbol{\beta}. \end{aligned}$$

Tomando diferenciais de $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\alpha}$, obtemos

$$\begin{aligned} d_{\alpha} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= -\frac{1}{2} \text{tr} \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i + \frac{\hat{\kappa}_i}{2} \mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \mathbf{r}_i \quad \text{e} \quad (\text{A.9}) \\ d_{\alpha}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \{ (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i - d^2 \boldsymbol{\Sigma}_i \} \\ &\quad - \frac{\hat{\kappa}_i}{2} \mathbf{r}_i^T \mathbf{V}_i^{-1} \{ (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i + (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i - d^2 \boldsymbol{\Sigma}_i \} \mathbf{V}_i^{-1} \mathbf{r}_i, \end{aligned}$$

a partir de (A.9) obtemos que

$$\begin{aligned} d_{\lambda\alpha}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i \\ &\quad - \frac{\hat{\kappa}_i}{2} \mathbf{r}_i^T \mathbf{V}_i^{-1} \{ \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} d\boldsymbol{\Sigma}_i + (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \} \mathbf{V}_i^{-1} \mathbf{r}_i. \end{aligned}$$

Finalmente, os diferenciais de $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\lambda}$ são dados por

$$\begin{aligned} d_{\lambda} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= -\frac{1}{2} \text{tr} \mathbf{V}_i^{-1} d\boldsymbol{\Psi} + \frac{\hat{\kappa}_i}{2} \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \quad \text{e} \\ d_{\lambda}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{1}{2} \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i \{ (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i d\boldsymbol{\Psi} - d^2 \boldsymbol{\Sigma}_i \} \mathbf{Z}_i^T \\ &\quad - \frac{\hat{\kappa}_i}{2} \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \{ (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i d\boldsymbol{\Psi} + (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i d\boldsymbol{\Psi} \\ &\quad - d^2 \boldsymbol{\Psi} \} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i, \end{aligned}$$

em que $\mathbf{V}_i = \mathbf{Z}_i \Psi \mathbf{Z}_i^T + \Sigma_i$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \beta$, para $i = 1, \dots, M$. Avaliando as expressões dos diferenciais dados acima em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e utilizando os teoremas de identificação dados em Magnus e Neudecker (1988) obtemos a hessiana, dada na equação (2.21).

Esquemas de Perturbação

Para cada um dos esquemas de perturbação calculamos o diferencial $d_{\boldsymbol{\theta}\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ e obtemos a matriz $\partial^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) / \partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}$ avaliando $d_{\boldsymbol{\theta}\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. É assumido que podemos trocar as operações de integração e diferenciação.

Ponderação de Casos

Mediante diferenciação obtemos que

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) (\mathbf{d}\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) d\omega_i, \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} \{ \text{tr} \mathbf{V}_i^{-1} (\mathbf{d}\Sigma_i) - \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} (\mathbf{d}\Sigma_i) \mathbf{V}_i^{-1} \mathbf{r}_i \} d\omega_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} \{ \text{tr} \mathbf{V}_i^{-1} \mathbf{Z}_i (\mathbf{d}\Psi) \mathbf{Z}_i^T - \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (\mathbf{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \} d\omega_i, \end{aligned}$$

em que $\mathbf{V}_i = \mathbf{Z}_i \Psi \mathbf{Z}_i^T + \Sigma_i$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$.

Perturbação da Matriz de Escala dos Efeitos Aleatórios

Tomando diferenciais de $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e ω_i , obtemos

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \frac{\omega_i^{-2} \kappa^{-1}(v_i)}{2} (\mathbf{d}\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i \Psi \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i d\omega_i \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{\omega_i^{-2}}{2} \text{tr} \mathbf{V}_{i\omega}^{-1} (\mathbf{d}\Sigma_i) \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i \Psi \mathbf{Z}_i^T d\omega_i \\ &\quad + \frac{\omega_i^{-2} \kappa^{-1}(v_i)}{2} \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \{ (\mathbf{d}\Sigma_i) \mathbf{V}_{i\omega} \mathbf{Z}_i \Psi \mathbf{Z}_i^T + \mathbf{Z}_i \Psi \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{d}\Sigma_i \} \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i d\omega_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{\omega_i^{-2}}{2} \text{tr} \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i (\mathbf{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} (\Sigma_i - \kappa^{-1}(v_i) \mathbf{r}_i \mathbf{r}_i^T) d\omega_i \\ &\quad + \frac{\omega_i^{-3} \kappa^{-1}(v_i)}{2} \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i \{ (\mathbf{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_{i\omega} \mathbf{Z}_i \Psi + \Psi \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i \mathbf{d}\Psi \} \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i d\omega_i, \end{aligned}$$

em que $\mathbf{V}_{i\omega} = \omega_i^{-1} \mathbf{Z}_i \Psi \mathbf{Z}_i^T + \Sigma_i$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$.

Perturbação das Variáveis Explicativas

Diferenciando $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) s_i (d\boldsymbol{\beta})^T (\mathbf{c}_t^T \boldsymbol{\beta} \mathbf{X}_i^T - \mathbf{c}_t \mathbf{r}_{i\omega}^T) \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) s_i \mathbf{c}_t^T \boldsymbol{\beta} \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) s_i \mathbf{c}_t^T \boldsymbol{\beta} \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i, \end{aligned}$$

em que $\mathbf{r}_{i\omega} = \mathbf{r}_i - s_i \mathbf{c}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i$, $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, com s_i um fator de escala, $\boldsymbol{\omega}_i$ vetor de perturbação $n_i \times 1$, $i = 1, \dots, M$ e \mathbf{c}_t vetor p -dimensional com 1 na posição t -ésima e zeros nas restantes.

Perturbação da Resposta

Tomando diferenciais de $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) (d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i, \\ d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i \quad \text{e} \\ d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) \mathbf{r}_{i\omega}^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} d\boldsymbol{\omega}_i, \end{aligned}$$

em que $\mathbf{r}_{i\omega} = \mathbf{r}_i + \boldsymbol{\omega}_i$, com $\boldsymbol{\omega}_i$ vetor de perturbação $n_i \times 1$ e $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, para $i = 1, \dots, M$.

Perturbação da Matriz de Correlação dentro dos Indivíduos

Diferenciando $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos que

$$\begin{aligned} d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\kappa^{-1}(v_i) (d\boldsymbol{\beta})^T (\mathbf{r}^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i \otimes \mathbf{X}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} d\boldsymbol{\omega}_i, \\ d_{\alpha_1\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} d\mathbf{D}_i + (d\mathbf{D}_i) \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i + \mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{F}_i \mathbf{V}_{i\omega}^{-1}) \mathbf{S}_{n_i} d\boldsymbol{\omega}_i \\ &\quad + \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \{\mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{F}_i + \mathbf{F}_i \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T\} \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} d\boldsymbol{\omega}_i, \\ &\quad + \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T ((d\mathbf{D}_i) \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i + \mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} d\mathbf{D}_i) \mathbf{S}_{n_i} d\boldsymbol{\omega}_i, \\ d_{\alpha_2\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \frac{1}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i (d\mathbf{C}_i) \mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} d\boldsymbol{\omega}_i \\ &\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i (d\mathbf{C}_i) \mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} d\boldsymbol{\omega}_i \\ &\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i (d\mathbf{C}_i) \mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} d\boldsymbol{\omega}_i \quad \text{e} \end{aligned}$$

$$\begin{aligned}
d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \frac{1}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i (\text{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \text{d}\boldsymbol{\omega}_i \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i (\text{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \text{d}\boldsymbol{\omega}_i \\
&\quad - \frac{\kappa^{-1}(v_i)}{2} \text{vec}^T (\mathbf{D}_i \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i (\text{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{D}_i) \mathbf{S}_{n_i} \text{d}\boldsymbol{\omega}_i,
\end{aligned}$$

em que $\mathbf{V}_{i\omega} = \mathbf{Z}_i \Psi \mathbf{Z}_i^T + \boldsymbol{\Sigma}_{i\omega}$ com $\boldsymbol{\Sigma}_{i\omega} = \mathbf{D}_i \mathbf{C}_{i\omega} \mathbf{D}_i$ tal que $\mathbf{C}_{i\omega} = \mathbf{C}_i + \mathbf{W}_i$ é matriz de correlação e \mathbf{W}_i é matriz de perturbação simétrica $n_i \times n_i$, \mathbf{S}_{n_i} matriz de duplicação que satisfaz $\text{vec}(\mathbf{W}_i) = \mathbf{S}_{n_i} \boldsymbol{\omega}_i$ com $\boldsymbol{\omega}_i = \text{vech}(\mathbf{W}_i)$, $\mathbf{F}_{i\omega} = (\text{d}\mathbf{D}_i) \mathbf{C}_{i\omega} \mathbf{D}_i + \mathbf{D}_i \mathbf{C}_{i\omega} \text{d}\mathbf{D}_i$, $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, para $i = 1, \dots, M$.

Perturbação das Funções de Escala dentro dos Indivíduos

Mediante diferenciação obtemos que

$$\begin{aligned}
d_{\beta\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) (\text{d}\boldsymbol{\beta})^T (\mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1} \boldsymbol{\Sigma}_{i\omega} \otimes \mathbf{X}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{W}_i^{-1}) \mathbf{S}_{n_i}^* \text{d}\boldsymbol{\omega}_i, \\
d_{\alpha\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \text{vec}^T (\mathbf{W}_i^{-1} \mathbf{V}_{i\omega}^{-1} \mathbf{W}_i^{-1} (\text{d}\boldsymbol{\Sigma}_i) \mathbf{W}_i^{-1} \\
&\quad - \boldsymbol{\Sigma}_{i\omega} \mathbf{V}_{i\omega}^{-1} \mathbf{W}_i^{-1} (\text{d}\boldsymbol{\Sigma}_i) \mathbf{W}_i \mathbf{V}_{i\omega}^{-1} \mathbf{W}_i^{-1}) \mathbf{S}_{n_i}^* \text{d}\boldsymbol{\omega}_i \\
&\quad + \kappa^{-1}(v_i) \text{vec}^T (\mathbf{W}_i^{-1} \mathbf{V}_{i\omega}^{-1} \mathbf{W}_i^{-1} (\text{d}\boldsymbol{\Sigma}_i) \mathbf{W}_i^{-1} \mathbf{G}_{i\omega} \boldsymbol{\Sigma}_{i\omega} \\
&\quad + \mathbf{W}_i^{-1} \mathbf{G}_{i\omega} \mathbf{W}_i^{-1} (\text{d}\boldsymbol{\Sigma}_i) \mathbf{W}_i^{-1} \mathbf{V}_{i\omega}^{-1} \boldsymbol{\Sigma}_{i\omega} - \mathbf{W}_i^{-1} \mathbf{G}_{i\omega} \mathbf{W}_i^{-1} \text{d}\boldsymbol{\Sigma}_i) \mathbf{S}_{n_i}^* \text{d}\boldsymbol{\omega}_i \quad \text{e} \\
d_{\lambda\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= - \text{vec}^T (\mathbf{W}_i^T \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i (\text{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \boldsymbol{\Sigma}_{i\omega}) \mathbf{S}_{n_i}^* \text{d}\boldsymbol{\omega}_i \\
&\quad + \kappa^{-1}(v_i) \text{vec}^T (\mathbf{W}_i^{-1} \mathbf{V}_{i\omega}^{-1} \mathbf{Z}_i (\text{d}\Psi) \mathbf{Z}_i^T \mathbf{G}_{i\omega} \boldsymbol{\Sigma}_{i\omega} \\
&\quad + \mathbf{W}_i^{-1} \mathbf{G}_{i\omega} \mathbf{Z}_i (\text{d}\Psi) \mathbf{Z}_i^T \mathbf{V}_{i\omega}^{-1} \boldsymbol{\Sigma}_{i\omega}) \mathbf{S}_{n_i}^* \text{d}\boldsymbol{\omega}_i,
\end{aligned}$$

em que $\mathbf{V}_{i\omega} = \mathbf{Z}_i \Psi \mathbf{Z}_i^T + \boldsymbol{\Sigma}_{i\omega}$ com $\boldsymbol{\Sigma}_{i\omega} = \mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{W}_i^{-1}$, $\mathbf{W}_i = \text{diag}(\boldsymbol{\omega}_i)$ matriz de perturbações $n_i \times n_i$, $\mathbf{G}_{i\omega} = \mathbf{V}_{i\omega}^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_{i\omega}^{-1}$, $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, aqui $\mathbf{S}_{n_i}^*$ é uma matriz de duplicação tal que $\text{vec}(\mathbf{W}_i) = \mathbf{S}_{n_i}^* \boldsymbol{\omega}_i$, para $i = 1, \dots, M$.

A.4 Método de Alavanca Generalizado no Modelo Linear Misto Hierárquico

Obtemos os diferenciais $\text{d}_{\boldsymbol{\theta}} \boldsymbol{\mu}_i$ e $\text{d}_{\boldsymbol{\theta} \mathbf{Y}_i}^2 L(\boldsymbol{\theta} | \mathbf{Y}_c)$ para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ requeridos para o cálculo das matrizes $\partial \boldsymbol{\mu} / \partial \boldsymbol{\theta}^T$ e $\partial^2 Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta} \partial \mathbf{Y}^T$. Temos que $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i$ e $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$ para o modelo linear misto hierárquico em que \mathbf{b}_i e v_i são considerados não observáveis e quando os efeitos aleatórios são integrados, respectivamente.

Para qualquer um dos casos anteriores, temos que

$$d_{\beta} \boldsymbol{\mu}_i = \mathbf{X}_i d\boldsymbol{\beta}, \quad d_{\alpha} \boldsymbol{\mu}_i = \mathbf{0} \quad \text{e} \quad d_{\lambda} \boldsymbol{\mu}_i = \mathbf{0}, \quad i = 1, \dots, M.$$

A seguir apresentamos os diferenciais $d_{\beta Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)$, $d_{\alpha Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)$ e $d_{\lambda Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)$ para cada uma das formulações consideradas. Avaliando tais diferenciais em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ é possível obter $\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \mathbf{Y}^T|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

Método de Alavanca Generalizado quando \mathbf{b}_i e v_i são não observáveis

E tomando diferenciais de $L(\boldsymbol{\theta}|\mathbf{Y}_c)$ com relação a $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ e \mathbf{Y}_i , obtemos

$$\begin{aligned} d_{\beta Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c) &= \kappa^{-1}(v_i)(d\boldsymbol{\beta})^T \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} d\mathbf{Y}_i, \\ d_{\alpha Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c) &= \kappa^{-1}(v_i) \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} (d\boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} d\mathbf{Y}_i \quad \text{e} \\ d_{\lambda Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c) &= \mathbf{0}, \end{aligned}$$

em que $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$ para $i = 1, \dots, M$.

Método de Alavanca Generalizado quando os v_i são não observáveis

Diferenciando $L(\boldsymbol{\theta}|\mathbf{Y}_c)$ com relação a $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T$ e \mathbf{Y}_i , temos que

$$\begin{aligned} d_{\beta Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c) &= \kappa^{-1}(v_i)(d\boldsymbol{\beta})^T \mathbf{X}_i^T \mathbf{V}_i^{-1} d\mathbf{Y}_i, \\ d_{\alpha Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c) &= \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} (d\boldsymbol{\Sigma}_i) \mathbf{V}_i^{-1} d\mathbf{Y}_i \quad \text{e} \\ d_{\lambda Y_i}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c) &= \kappa^{-1}(v_i) \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i (d\boldsymbol{\Psi}) \mathbf{Z}_i^T \mathbf{V}_i^{-1} d\mathbf{Y}_i, \end{aligned}$$

em que $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ para $i = 1, \dots, M$.

Apêndice B

Equivalência entre Medidas de Eliminação de Casos e o Método de Influência Local

Os procedimentos de diagnóstico por eliminação de casos, também conhecidos como influência global determinam o efeito de observações influentes no resultado do ajuste do modelo mediante a completa retirada do i -ésimo indivíduo. Neste apêndice focamos nosso interesse na comparação da diferença entre $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{(i)}$ em que $\hat{\boldsymbol{\theta}}_{(i)}$ denota a estimativa de máxima verossimilhança para $\boldsymbol{\theta}$ quando efetuamos a retirada do i -ésimo sujeito $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{Z}_i)$ do modelo linear com efeitos mistos dado em (1.1), isto é, $\hat{\boldsymbol{\theta}}_{(i)}$ maximiza o logaritmo da verossimilhança,

$$L_{(i)}(\boldsymbol{\theta}) = \sum_{j \neq i} L_j(\boldsymbol{\theta}).$$

Métodos de diagnóstico obtidos mediante eliminação de observações têm sido amplamente discutidos para modelos de regressão linear sob normalidade; veja, por exemplo, Belsley, Kuh e Welsh (1980), Cook e Weisberg (1982) e Chatterjee e Hadi (1988), dentre outros. Tais métodos foram estendidos para dados com estrutura longitudinal por De Gruttola, Ware e Louis (1987), para modelos lineares mistos sob normalidade e modelos lineares mistos generalizados por Christensen, Pearson e Johnson (1992) e Tan, Ouwens e Berger (2001), e Xiang, Tse e Lee (2002) e Lee e Xu (2004), respectivamente.

Como uma medida de comparação entre $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{(i)}$ consideramos a distância

$$D_i = (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)})^T \{-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})\}^{-1} (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)}), \quad i = 1, \dots, M, \quad (\text{B.1})$$

e observamos que o cálculo de $\widehat{\boldsymbol{\theta}}_{(i)}$ pode ficar muito trabalhoso para modelos elípticos lineares mistos. Para contornar esse problema, Pregibon (1981) e Cook e Weisberg (1982) sugerem utilizar a seguinte aproximação de um passo $\widehat{\boldsymbol{\theta}}_{(i)}^1$ para $\widehat{\boldsymbol{\theta}}_{(i)}$ como

$$\widehat{\boldsymbol{\theta}}_{(i)}^1 = \widehat{\boldsymbol{\theta}}_{(i)} + \{-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})\}^{-1} \mathbf{U}_{(i)}(\widehat{\boldsymbol{\theta}}),$$

em que $\mathbf{U}_{(i)}(\widehat{\boldsymbol{\theta}}) = \partial L_{(i)}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T |_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$. Como $\mathbf{U}(\widehat{\boldsymbol{\theta}}) = \partial L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T |_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathbf{0}$ e $\mathbf{U}_{(i)}(\widehat{\boldsymbol{\theta}}) = -\mathbf{U}_i(\widehat{\boldsymbol{\theta}}) = \partial L_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T |_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, concluímos que D_i pode ser aproximado por

$$D_i^1 = \mathbf{U}_i^T(\widehat{\boldsymbol{\theta}}) \{-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})\}^{-1} \mathbf{U}_i(\widehat{\boldsymbol{\theta}}), \quad i = 1, \dots, M. \quad (\text{B.2})$$

Agora, para o esquema de perturbação de ponderação de casos temos que

$$\left. \frac{\partial^2 L(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{U}_i(\widehat{\boldsymbol{\theta}}).$$

Desse modo, a matriz $\boldsymbol{\Delta}$ na curvatura normal $\mathbf{B} = -\boldsymbol{\Delta}^T \ddot{\mathbf{L}}^{-1}(\widehat{\boldsymbol{\theta}}) \boldsymbol{\Delta}$ assume a forma

$$\boldsymbol{\Delta} = (\mathbf{U}_1(\widehat{\boldsymbol{\theta}}), \dots, \mathbf{U}_M(\widehat{\boldsymbol{\theta}})).$$

Logo, a curvatura normal $C_h(\widehat{\boldsymbol{\theta}})$ na direção \mathbf{h}_i , em que \mathbf{h}_i é um vetor $M \times 1$ com 1 na i -ésima posição e zeros nas restantes fica dada por

$$C_{\mathbf{h}_i}(\widehat{\boldsymbol{\theta}}) = \mathbf{U}_i^T(\widehat{\boldsymbol{\theta}}) \{-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})\}^{-1} \mathbf{U}_i(\widehat{\boldsymbol{\theta}}) = 2 D_i^1,$$

para $i = 1, \dots, M$.

Esse resultado também é válido considerando a distância

$$QD_i = (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)})^T \{-\ddot{\mathbf{Q}}(\widehat{\boldsymbol{\theta}})\}^{-1} (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)}), \quad i = 1, \dots, M, \quad (\text{B.3})$$

que pode ser aproximada como

$$QD_i^1 = \dot{\mathbf{Q}}_i^T(\widehat{\boldsymbol{\theta}}) \{-\ddot{\mathbf{Q}}(\widehat{\boldsymbol{\theta}})\}^{-1} \dot{\mathbf{Q}}_i(\widehat{\boldsymbol{\theta}}), \quad i = 1, \dots, M, \quad (\text{B.4})$$

em que, para esse caso $\widehat{\boldsymbol{\theta}}_{(i)}$ maximiza a função $Q_{(i)}(\boldsymbol{\theta} | \widehat{\boldsymbol{\theta}}) = \text{E}\{L(\boldsymbol{\theta} | \mathbf{Y}_{(i)c}) | \mathbf{Y}_{(i)}, \boldsymbol{\theta}\}$ e $\mathbf{Y}_{(i)c}$, $\mathbf{Y}_{(i)}$ representam os dados completos e observados para o modelo

linear misto hierárquico em que $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{Z}_i)$ têm sido eliminados, $\ddot{Q}(\hat{\boldsymbol{\theta}}) = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ e $\dot{Q}_i(\hat{\boldsymbol{\theta}}) = \partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}^T|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. É possível mostrar que

$$C_{f_Q, h_i}(\hat{\boldsymbol{\theta}}) = \dot{Q}_i^T(\hat{\boldsymbol{\theta}})\{-\ddot{Q}(\hat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_i(\hat{\boldsymbol{\theta}}) = 2QD_i^1,$$

para $i = 1, \dots, M$.

Esses resultados indicam que para modelos elípticos lineares com efeitos mistos para qualquer das suas formulações marginal ou hierárquica, não é necessário o cálculo de medidas de diagnóstico por eliminação de casos, uma vez que esta informação é obtida através do método de influência local.

Apêndice C

Diagnóstico de Influência no Modelo de Grubbs

Neste apêndice estudamos a influência local no modelo de Grubbs (Grubbs, 1948) considerando distribuições de mistura de escala normal. O principal objetivo do modelo proposto por Grubbs (1948, 1973, 1983) corresponde a examinar a qualidade de vários instrumentos de medição usados para estimar a mesma quantidade desconhecida x em um grupo comum de indivíduos ou unidades experimentais. O problema da comparação de instrumentos de medição cujas características diferem em custo, velocidade e outros fatores tais como eficiência tem recebido uma crescente atenção em áreas como engenharia, medicina e agricultura, dentre outras (veja por exemplo, Barnett, 1969; Grubbs, 1948, 1973, 1983 e Fuller, 1987).

O modelo de Grubbs pode ser entendido como um caso particular do modelo linear com efeitos mistos, contudo esse modelo também pode ser considerado um caso particular do modelo de Barnett (Barnett, 1969), também conhecido como modelo de calibração comparativa estrutural. Salientamos que tal modelo não é um caso particular do modelo linear misto usual.

Estudos de influência local no modelo de calibração comparativa estrutural sob normalidade foram considerados por Galea, Bolfarine e de Castro (2002), usando um enfoque para introduzir distribuições de contornos elípticos por Galea, Bolfarine e Vilca (2005) e mais recentemente Lachos, Vilca e Galea (2006) estudam diagnóstico de influência no modelo de Grubbs considerando erros normais. Neste apêndice adotamos um enfoque diferente ao de Galea, Bolfarine e Vilca (2005) para a inclusão de distribuições na

classe das elípticas, tais como a família de distribuições de mistura de escala normal, no modelo de Grubbs. Finalmente, abordamos a influência local nesta classe de modelos sob vários esquemas de perturbação e obtemos também a matriz de alavancas generalizadas.

Grubbs (1948) propôs um modelo para comparar as medições de p instrumentos num grupo de M indivíduos, dado por

$$Y_{ij} = \alpha_j + x_i + \epsilon_{ij}, \quad i = 1, \dots, M; j = 1, \dots, p, \quad (\text{C.1})$$

em que Y_{ij} representa a medição usando o j -ésimo instrumento na i -ésima unidade experimental. Os erros de medição ϵ_{ij} são assumidos independentes dos valores aleatórios x_1, \dots, x_M . Frequentemente é assumido que os x_j e ϵ_{ij} são independentes seguindo uma distribuição normal $N(\mu_x, \phi_x)$ e $N(0, \phi_i)$, respectivamente.

Para assegurar a identificabilidade do modelo em (C.1) pode-se assumir que $\alpha_1 = 0$ (veja, por exemplo, Shyr e Gleser, 1986 e Lachos, Vilca e Galea, 2006). Neste trabalho evitamos manipular restrições no espaço paramétrico, considerando a transformação $z_i = x_i - \mu_x$, $i = 1, \dots, M$ (Theobald e Mallinson, 1978). Desse modo o modelo de Grubbs em notação matricial fica dado por

$$\mathbf{Y}_i = \boldsymbol{\mu} + \mathbf{1}_p z_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M, \quad (\text{C.2})$$

em que $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T$ é vetor p -dimensional, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^T$ e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{ip})^T$ denotam vetores aleatórios $p \times 1$ e z_i representa uma variável aleatória com posição 0 e escala ϕ_x .

Similarmente aos modelos desenvolvidos na Seção 1.3 introduzimos distribuições de contornos elípticos no modelo (C.2) mediante considerar a seguinte estrutura hierárquica:

$$\begin{aligned} \mathbf{Y}_i | z_i &\stackrel{ind}{\sim} SMN_p(\boldsymbol{\mu} + \mathbf{1}_p z_i, D(\boldsymbol{\phi}); H) \\ z_i &\stackrel{ind}{\sim} SMN(0, \phi_x; H), \quad i = 1, \dots, M, \end{aligned}$$

isto é,

$$\begin{aligned} \mathbf{Y}_i | z_i, v_i &\stackrel{ind}{\sim} N_p(\boldsymbol{\mu} + \mathbf{1}_p z_i, \kappa(v_i) D(\boldsymbol{\phi})), \quad z_i | v_i \stackrel{ind}{\sim} N(0, \kappa(v_i) \phi_x), \\ v_i &\stackrel{ind}{\sim} H(v_i; \boldsymbol{\nu}), \quad i = 1, \dots, M, \end{aligned} \quad (\text{C.3})$$

em que $D(\boldsymbol{\phi})$ denota uma matriz diagonal cujos elementos diagonais são dados pelos elementos do vetor p -dimensional $\boldsymbol{\phi}$, ou seja, $D(\boldsymbol{\phi}) = \text{diag}(\phi_1, \dots, \phi_p)$.

Utilizamos um algoritmo tipo-EM para a estimação de máxima verossimilhança no modelo definido em (C.2), considerando o vetor de dados completos $\mathbf{Y}_c = (\mathbf{Y}^T, \mathbf{z}^T, \mathbf{v}^T)^T$, em que $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$ corresponde ao vetor de respostas para os M indivíduos, $\mathbf{z} = (z_1, \dots, z_M)^T$ e $\mathbf{v} = (v_1, \dots, v_M)^T$. Supondo que $(\mathbf{z}^T, \mathbf{v}^T)^T$ são não observáveis, o logaritmo da verossimilhança baseada no vetor de dados completos fica $L(\boldsymbol{\theta}|\mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}|\mathbf{Y}_c)$, em que

$$L_i(\boldsymbol{\theta}|\mathbf{Y}_c) = -\frac{1}{2} \log |D(\boldsymbol{\phi})| - \frac{\kappa^{-1}(v_i)}{2} (\mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p z_i)^T D^{-1}(\boldsymbol{\phi}) (\mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p z_i) - \frac{1}{2} \log \phi_x - \frac{\kappa^{-1}(v_i)}{2\phi_x} z_i^2 + \log h(v_i; \boldsymbol{\nu}) + c, \quad (\text{C.4})$$

com $h(v_i; \boldsymbol{\nu})$ sendo a função de densidade da variável de mistura v_i com parâmetro $\boldsymbol{\nu}$ e c denota uma constante.

A seguir apresentamos a $(r+1)$ -ésima etapa de um algoritmo ECM multiciclo para a estimação de $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \phi_x, \boldsymbol{\phi}^T)^T$ no modelo (C.2).

Passo E: usando estimativas da r -ésima etapa, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(r)}$ calcular $\hat{z}_i, \hat{\tau}$ como

$$\hat{z}_i = \text{E}(z_i|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}) = \frac{\hat{\phi}_x}{k} \mathbf{1}_p^T D^{-1}(\hat{\boldsymbol{\phi}}) (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}) \quad \text{e}$$

$$\hat{\tau} = \kappa^{-1}(v_i) \text{Var}(z_i|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}) = \frac{\hat{\phi}_x}{k},$$

para $i = 1, \dots, M$, em que $k = 1 + \phi_x \mathbf{1}_p^T D^{-1}(\boldsymbol{\phi}) \mathbf{1}_p$. Logo, obter $\hat{u}_i = u_i(\hat{\boldsymbol{\theta}})$ como

$$\begin{aligned} \hat{u}_i &= (\mathbf{Y}_i - \hat{\boldsymbol{\mu}})^T (D(\hat{\boldsymbol{\phi}}) + \hat{\phi}_x \mathbf{1}_p \mathbf{1}_p^T)^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}) \\ &= \mathbf{e}_i^T D^{-1}(\hat{\boldsymbol{\phi}}) \mathbf{e}_i + \hat{z}_i^2 / \hat{\phi}_x, \end{aligned}$$

em que $\mathbf{e}_i = \mathbf{Y}_i - \hat{\boldsymbol{\mu}} - \mathbf{1}_p \hat{z}_i$, para $i = 1, \dots, M$, e avaliar $\hat{\kappa}_i = \text{E}(\kappa^{-1}(v_i)|\mathbf{Y}_i, \hat{\boldsymbol{\theta}})$ usando as expressões desenvolvidas na Seção 1.3.3 para alguma mistura de escala normal particular.

Passo CM multiciclo: maximização da esperança do logaritmo da função de verossimilhança de dados completos, $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, com $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = Q_{1i}(\boldsymbol{\mu}, \boldsymbol{\phi}|\hat{\boldsymbol{\theta}}) + Q_{2i}(\phi_x|\hat{\boldsymbol{\theta}})$, em que

$$\begin{aligned} Q_{1i}(\boldsymbol{\mu}, \boldsymbol{\phi}|\hat{\boldsymbol{\theta}}) &= -\frac{1}{2} \log |D(\boldsymbol{\phi})| - \frac{\hat{\tau}}{2} \mathbf{1}_p^T D^{-1}(\boldsymbol{\phi}) \mathbf{1}_p \\ &\quad - \frac{\hat{\kappa}_i}{2} (\mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p \hat{z}_i)^T D^{-1}(\boldsymbol{\phi}) (\mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p \hat{z}_i) \quad \text{e} \\ Q_{2i}(\phi_x|\hat{\boldsymbol{\theta}}) &= -\frac{1}{2} \log \phi_x - \frac{1}{2\phi_x} (\hat{\kappa}_i \hat{z}_i^2 + \hat{\tau}), \end{aligned} \quad (\text{C.5})$$

usando uma seqüência de passos CM.

Passo CM-1: atualizar $\boldsymbol{\mu}^{(r+1)}$ como

$$\boldsymbol{\mu}^{(r+1)} = \frac{1}{\sum_{j=1}^M \hat{\kappa}_j} \sum_{i=1}^M \hat{\kappa}_i (\mathbf{Y}_i - \mathbf{1}_p \hat{z}_i).$$

Passo E: executar o passo E com $\hat{\boldsymbol{\theta}} = (\boldsymbol{\mu}^{(r+1)T}, \phi_x^{(r)}, \boldsymbol{\phi}^{(r)T})^T$.

Passo CM-2: fixar $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}^{(r+1)}$, e atualizar $\boldsymbol{\phi}^{(r+1)}$ como

$$\boldsymbol{\phi}^{(r+1)} = \hat{\tau} \mathbf{1}_p + \frac{1}{M} \sum_{i=1}^M \hat{\kappa}_i D(\mathbf{e}_i) \mathbf{e}_i,$$

com $\mathbf{e}_i = \mathbf{Y}_i - \hat{\boldsymbol{\mu}} - \mathbf{1}_p \hat{z}_i$, para $i = 1, \dots, M$.

Passo E: executar o passo E com $\hat{\boldsymbol{\theta}} = (\boldsymbol{\mu}^{(r+1)T}, \phi_x^{(r)}, \boldsymbol{\phi}^{(r+1)T})^T$.

Passo CM-3: atualizar $\phi_x^{(r+1)}$ como

$$\phi_x^{(r+1)} = \hat{\tau} + \frac{1}{M} \sum_{i=1}^M \hat{\kappa}_i \hat{z}_i^2.$$

Então, fazer $\boldsymbol{\theta}^{(r)} = (\boldsymbol{\mu}^{(r+1)T}, \phi_x^{(r+1)}, \boldsymbol{\phi}^{(r+1)T})^T$ e voltar à etapa E. Este algoritmo ECM multiciclo itera entre os passos E e CM até que a sequência $\boldsymbol{\theta}^{(r)}$ atinja a convergência.

Salientamos que o algoritmo ECM desenvolvido neste apêndice compartilha as mesmas propriedades de simplicidade e estabilidade dos algoritmos para estimação por máxima verossimilhança em modelos de calibração comparativa propostos por Bolfarine e Galea (1995, 1996).

Influência Local

Nesta seção derivamos a curvatura normal no modelo de Grubbs quando consideramos distribuições de mistura de escala normal. Analogamente ao desenvolvimento apresentado na Seção 2.6, obtemos a curvatura normal usando a função Q -afastamento como medida de influência. A seguir apresentamos os diferenciais $d_{\boldsymbol{\theta}}^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ e $d_{\boldsymbol{\theta}\omega}^2 L(\boldsymbol{\theta}|\mathbf{Y}_c)$ para vários esquemas de perturbação. As matrizes $\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ e $\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T$ podem ser obtidas a partir desses diferenciais mediante os teoremas de identificação dados em Magnus e Neudecker (1988).

Matriz Hessiana, \ddot{Q}

Seja $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \phi_x, \boldsymbol{\phi}^T)^T$ o vetor de parâmetros de interesse. Temos que o segundo diferencial de $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\theta}$ fica dado por:

$$d_{\boldsymbol{\theta}}^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^M d_{\boldsymbol{\theta}}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}),$$

em que $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ é dado em (C.5). A seguir apresentamos os diferenciais $d_{\boldsymbol{\theta}}^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ para $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \phi_x, \boldsymbol{\phi}^T)^T$.

Tomando diferenciais de $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\mu}$ e então com relação a $\boldsymbol{\phi}$, obtemos

$$\begin{aligned} d_{\boldsymbol{\mu}}^2 Q_{1i}(\boldsymbol{\mu}, \boldsymbol{\phi}|\hat{\boldsymbol{\theta}}) &= -\hat{\kappa}_i (d\boldsymbol{\mu})^T D^{-1}(\boldsymbol{\phi}) d\boldsymbol{\mu}, & \text{e} \\ d_{\boldsymbol{\mu}\boldsymbol{\phi}}^2 Q_{1i}(\boldsymbol{\mu}, \boldsymbol{\phi}|\hat{\boldsymbol{\theta}}) &= -\hat{\kappa}_i (d\boldsymbol{\mu})^T D^{-2}(\boldsymbol{\phi}) D(\mathbf{e}_i^*) d\boldsymbol{\phi}, \end{aligned}$$

em que $\mathbf{e}_i^* = \mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p \hat{z}_i$, $i = 1, \dots, M$ e $D^{-m}(\mathbf{a}) = \text{diag}(a_1^{-m}, \dots, a_p^{-m})$, para \mathbf{a} vetor p -dimensional e $m > 0$.

Diferenciando $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\phi}$, obtemos

$$\begin{aligned} d_{\boldsymbol{\phi}}^2 Q_{1i}(\boldsymbol{\mu}, \boldsymbol{\phi}|\hat{\boldsymbol{\theta}}) &= \frac{1}{2} (d\boldsymbol{\phi})^T D^{-2}(\boldsymbol{\phi}) d\boldsymbol{\phi} - \hat{\tau} (d\boldsymbol{\phi})^T D^{-3}(\boldsymbol{\phi}) d\boldsymbol{\phi} \\ &\quad - \hat{\kappa}_i (d\boldsymbol{\phi})^T D(\mathbf{e}_i^*) D^{-3}(\boldsymbol{\phi}) D(\mathbf{e}_i^*) d\boldsymbol{\phi}. \end{aligned}$$

Finalmente, tomando diferenciais de $Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com relação a ϕ_x , obtemos

$$d_{\phi_x}^2 Q_{2i}(\phi_x|\hat{\boldsymbol{\theta}}) = \frac{1}{2\phi_x^2} d^2 \phi_x - \frac{1}{\phi_x^3} (\hat{\kappa}_i \hat{z}_i^2 + \hat{\tau}) d^2 \phi_x.$$

Avaliando estes diferenciais em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ podemos obter a matriz hessiana \ddot{Q} .

Esquemas de Perturbação

Para cada um dos esquemas de perturbação obtemos o diferencial $d_{\boldsymbol{\theta}\boldsymbol{\omega}}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$. Avaliando esses diferenciais em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ podemos calcular $\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T$ para o esquema de perturbação particular. A seguir é assumido que é possível trocar as operações de integração e diferenciação.

Ponderação de Casos

Considere a inclusão de pesos para cada unidade experimental no logaritmo da verossimilhança para dados completos, como

$$L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M \omega_i L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$$

com $L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ dado em (C.4), aqui $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$ em que $0 \leq \omega_i \leq 1$ para $i = 1, \dots, M$ e $\boldsymbol{\omega}_0 = \mathbf{1}_M$. Para esse esquema de perturbação temos que

$$\begin{aligned} d_{\mu\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) (d\boldsymbol{\mu})^T D^{-1}(\boldsymbol{\phi}) (\mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p z_i) d\omega_i, \\ d_{\phi\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} (d\boldsymbol{\phi})^T \{D^{-1}(\boldsymbol{\phi}) \mathbf{1}_p - \kappa^{-1}(v_i) D(\boldsymbol{\epsilon}_i) D^{-2}(\boldsymbol{\phi}) \boldsymbol{\epsilon}_i\} d\omega_i \quad \text{e} \\ d_{\phi_x\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= -\frac{1}{2} \left\{ \frac{1}{\phi_x} + \frac{\kappa^{-1}(v_i)}{\phi_x^2} z_i^2 \right\} d\phi_x d\omega_i, \end{aligned}$$

em que $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p z_i$, para $i = 1, \dots, M$.

Perturbação das Observações

Consideramos perturbações aditivas nas medições obtidas através dos instrumentos usados no estudo. Seja $\mathbf{Y}_{i\omega}$ o vetor de observações perturbadas para cada unidade experimental. Abordamos os seguintes casos de interesse:

Perturbação simultânea das medições fornecidas pelos p instrumentos:

isto é, substituímos \mathbf{Y}_i por $\mathbf{Y}_{i\omega} = \mathbf{Y}_i + \boldsymbol{\omega}_i$ em que $\boldsymbol{\omega}_i \in \mathbb{R}^p$, $i = 1, \dots, M$, em cujo caso temos que $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^{Mp}$. O logaritmo da verossimilhança de dados completos fica dado por $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ em que

$$\begin{aligned} L_i(\boldsymbol{\theta} | \mathbf{Y}_c) &= -\frac{1}{2} \log |D(\boldsymbol{\phi})| - \frac{\kappa^{-1}(v_i)}{2} (\boldsymbol{\epsilon}_i + \boldsymbol{\omega}_i)^T D^{-1}(\boldsymbol{\phi}) (\boldsymbol{\epsilon}_i + \boldsymbol{\omega}_i) \\ &\quad - \frac{1}{2} \log \phi_x - \frac{\kappa^{-1}(v_i)}{2\phi_x} z_i^2 + \log h(v_i; \boldsymbol{\nu}) + c \end{aligned}$$

com $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p z_i$, para $i = 1, \dots, M$.

Tomando diferenciais de $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos

$$\begin{aligned} d_{\mu\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) (d\boldsymbol{\mu})^T D^{-1}(\boldsymbol{\phi}) d\boldsymbol{\omega}_i, \\ d_{\phi\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \kappa^{-1}(v_i) (d\boldsymbol{\phi})^T D(\boldsymbol{\epsilon}_i + \boldsymbol{\omega}_i) D^{-2}(\boldsymbol{\phi}) d\boldsymbol{\omega}_i \quad \text{e} \\ d_{\phi_x\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) &= \mathbf{0}. \end{aligned}$$

Perturbação das medições obtidas desde um dos instrumentos: suponha que temos interesse em perturbar as medições fornecidas pelo t -ésimo instrumento, $t = 1, \dots, p$. Neste caso $\mathbf{Y}_{i\omega} = \mathbf{Y}_i + \omega_i \mathbf{c}_t$ em que \mathbf{c}_t denota um vetor p -dimensional com 1 na t -ésima posição, e zeros nas restantes. O vetor de perturbação fica dado por $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)^T$ e $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^M$. O logaritmo da verossimilhança de dados completos sob este esquema de perturbação assume a forma $L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \sum_{i=1}^M L_i(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c)$ em que

$$L_i(\boldsymbol{\theta} | \mathbf{Y}_c) = -\frac{1}{2} \log |D(\boldsymbol{\phi})| - \frac{\kappa^{-1}(v_i)}{2} (\boldsymbol{\epsilon}_i + \omega_i \mathbf{c}_t)^T D^{-1}(\boldsymbol{\phi}) (\boldsymbol{\epsilon}_i + \omega_i \mathbf{c}_t) \\ - \frac{1}{2} \log \phi_x - \frac{\kappa^{-1}(v_i)}{2\phi_x} z_i^2 + \log h(v_i; \boldsymbol{\nu}) + c$$

com $\boldsymbol{\epsilon}_i = \mathbf{Y}_i - \boldsymbol{\mu} - \mathbf{1}_p z_i$, para $i = 1, \dots, M$.

Mediante diferenciação obtemos que

$$d_{\boldsymbol{\mu}\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \kappa^{-1}(v_i) (d\boldsymbol{\mu})^T D^{-1}(\boldsymbol{\phi}) \mathbf{c}_t d\omega_i, \\ d_{\phi\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = \kappa^{-1}(v_i) (d\boldsymbol{\phi})^T D(\boldsymbol{\epsilon}_i + \omega_i \mathbf{c}_t) D^{-2}(\boldsymbol{\phi}) \mathbf{c}_t d\omega_i \quad \text{e} \\ d_{\phi_x\omega_i}^2 L(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{Y}_c) = 0.$$

Método de Alavanca Generalizado no Modelo de Grubbs

Derivamos os diferenciais requeridos para o cálculo da matriz de alavancas generalizadas para o modelo de Grubbs usando a metodologia proposta na Seção 2.3.

Considere $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \mathbf{1}_p z_i$, para $i = 1, \dots, M$. A seguir obtemos os diferenciais $d_{\boldsymbol{\theta}} \boldsymbol{\mu}_i$ e $d_{\boldsymbol{\theta}Y_i}^2 L(\boldsymbol{\theta} | \mathbf{Y}_c)$ a partir dos quais é possível obter as matrizes $\partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^T$ e $\partial^2 Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}) / \partial\boldsymbol{\theta} \partial \mathbf{Y}^T$. Usando o método de diferenciação de matrizes obtemos

$$d_{\boldsymbol{\mu}} \boldsymbol{\mu}_i = d\boldsymbol{\mu}, \quad d_{\phi} \boldsymbol{\mu}_i = \mathbf{0} \quad \text{e} \quad d_{\phi_x} \boldsymbol{\mu}_i = \mathbf{0}, \quad i = 1, \dots, M.$$

Tomando diferenciais de $L(\boldsymbol{\theta} | \mathbf{Y}_c)$ com relação a $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \phi_x^T, \boldsymbol{\phi}^T)^T$ e \mathbf{Y}_i , obtemos que

$$d_{\boldsymbol{\mu}Y_i}^2 L(\boldsymbol{\theta} | \mathbf{Y}_c) = \kappa^{-1}(v_i) (d\boldsymbol{\mu})^T D^{-1}(\boldsymbol{\phi}) d\mathbf{Y}_i, \\ d_{\phi Y_i}^2 L(\boldsymbol{\theta} | \mathbf{Y}_c) = \kappa^{-1}(v_i) (d\boldsymbol{\phi})^T D(\boldsymbol{\epsilon}_i) D^{-2}(\boldsymbol{\phi}) d\mathbf{Y}_i \quad \text{e} \\ d_{\phi_x Y_i}^2 L(\boldsymbol{\theta} | \mathbf{Y}_c) = \mathbf{0}.$$

É possível apreciar que assim como para os resultados desenvolvidos nas Seções 2.5 e 2.7 existe uma relação entre o método de alavanca e a influência local considerando perturbações aditivas nas observações.

Referências Bibliográficas

- [1] Anderson, E.; Z. Bai; C. Bischof; J. Demmel; J. Dongarra; J. DuCroz, A. Greenbaum; S. Hammarling; A. McKenney; S. Ostrouchov e D. Sorensen, (1995). *LAPACK Users Guide, Release 2.0, 2nd ed.*, SIAM Publications, Philadelphia.
- [2] Andrade, F. C., (2004). *Pontos de Alavanca em Regressão*. Dissertação de Mestrado, Departamento de Estatística, Universidade de São Paulo.
- [3] Andrews, D. F. e C. L. Mallows, (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36**, 99–102.
- [4] Arellano, R., (1994). *Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão*. Tese de Doutorado, Departamento de Estatística, Universidade de São Paulo.
- [5] Armstrong, R. D. e E. L. Frome, (1976). A comparison of two algorithms for absolute deviation curve fitting. *Journal of the American Statistical Association* **71**, 328–330.
- [6] Arslan, O., (2001). Family of multivariate generalized t distributions. *Journal of Multivariate Analysis* **89**, 329–337.
- [7] Atkinson, A. C., (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika* **68**, 13–20.
- [8] Banerjee, M. e E. W. Frees, (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association* **92**, 999–1005.
- [9] Barnett, V. D., (1969). Simultaneous pairwise linear structural relationships. *Biometrics* **25**, 129–142.

- [10] Bates, D. M. e J. C. Pinheiro, (1998). Computational methods for multilevel models. *Technical Memorandum BL0112140-980226-01TM*, Bell Labs, Lucent Technologies, Murray Hill, NJ.
- [11] Bates, D. M. e S. DebRoy, (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis* **91**, 1–17.
- [12] Becker, R. A.; W. S. Cleveland e M. Shyu, (1996). The visual design and control of trellis display. *Journal of Computational and Graphical Statistics* **5**, 123–155.
- [13] Beckman, R. J.; C. J. Nachtsheim e R. D. Cook, (1987). Diagnostics for mixed-model analysis of variance. *Technometrics* **29**, 413–426.
- [14] Belsley, D. A.; E. Kuh e R. E. Welsh, (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.
- [15] Belsley, D. A., (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. Wiley, New York.
- [16] Bolfarine, H. e M. Galea, (1995). Structural comparative calibration using the EM algorithm. *Journal of Applied Statistics* **2**, 277–292.
- [17] Bolfarine, H. e M. Galea, (1996). One structural comparative calibration under a t -model. *Computational Statistics* **11**, 63–85.
- [18] Cadigan, N. G. e P. J. Farrel, (2002). Generalized local influence with applications to fish stock cohort analysis. *Applied Statistics* **51**, 469–483.
- [19] Chatterjee, S. e A. S. Hadi, (1988). *Sensitivity Analysis in Linear Regression*. Wiley, New York.
- [20] Chen, M. e D. K. Dey, (1998). Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions. *Sankhyā, Series A* **60**, 322–343.
- [21] Choy, S. T. B. e A. F. M. Smith, (1997). Hierarchical models with scale mixtures of normal distributions. *Test* **6**, 205–221.
- [22] Choy, S. T. B. e S. G. Walker, (2003). The extended exponential power distribution and Bayesian robustness. *Statistics & Probability Letters* **65**, 227–232.

- [23] Christensen, R.; L. M. Pearson e W. Johnson, (1992). Case-deletion diagnostics for mixed models. *Technometrics* **34**, 38–45.
- [24] Cook, R. D., (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B* **48**, 133–169.
- [25] Cook, R. D., (1997). Local Influence. Em S. Kotz; C. B. Read e D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences, Update, Vol. 1*, p. 380–385. Wiley.
- [26] Cook, R. D. e S. Weisberg (1982). *Residuals and Influence in Regression*. Chapman & Hall, London.
- [27] Copt, S. e M. Victoria-Feser, (2006). High breakdown inference in the mixed linear model. *Journal of the American Statistical Association* **101**, 292–300.
- [28] Cysneiros, F. J. A. e G. A. Paula, (2004). One-sided test in linear models with multivariate t -distribution. *Communications in Statistics - Simulation & Computation* **33**, 747–771.
- [29] Cysneiros, F. J. A. e G. A. Paula, (2005). Restricted methods in symmetrical linear regression models. *Computational Statistics & Data Analysis* **49**, 689–708.
- [30] Davidian, M e R. J. Carroll, (1987). Variance function estimation. *Journal of the American Statistical Association* **82**, 1079–1091.
- [31] De Gruttola, V.; H. J. Ware e T. A. Louis, (1987). Influence analysis of generalized least squares estimators. *Journal of the American Statistical Association* **82**, 911–917.
- [32] Dempster, A. P.; N. M. Laird e D. B. Rubin, (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- [33] Dempster, A. P.; N. M. Laird e D. B. Rubin, (1980). Iteratively reweighted least squares for linear regression when errors are Normal/Independent distributed. Em P. R. Krishnaiah (Ed.), *Multivariate Analysis V*, p. 35–57. North-Holland.
- [34] Dempster, A. P.; M. R. Selwyn; C. M. Patel e A. J. Roth, (1984). Statistical and computational aspects of mixed model analysis. *Applied Statistics* **33**, 203–214.

- [35] Dennis, J. E. e R. B. Schnabel, (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM Publications, Philadelphia.
- [36] Díaz-García, J. A.; M. Galea e V. Leiva-Sánchez, (2003). Influence diagnostics for elliptical multivariate linear regression models. *Communications in Statistics - Theory & Methods* **32**, 625–641.
- [37] Dongarra, J. J.; J. R. Bunch; C. B. Moler e G. W. Stewart, (1979). *Linpac Users Guide*. SIAM Publications, Philadelphia.
- [38] Dongarra, J. J.; J. DuCroz; S. Hammarling e R. Hanson, (1988). An extended set of Fortran basic linear algebra subprograms. *ACM Transactions on Mathematical Software* **14**, 1–17.
- [39] Dongarra, J. J.; J. DuCroz; S. Hammarling e I. Duff, (1990). A set of level 3 basic linear algebra subprograms. *ACM Transactions on Mathematical Software* **16**, 1–17.
- [40] Fang, K. T. e Y. T. Zhang, (1990). *Generalized Multivariate Analysis*. Springer-Verlag e Science Press, Berlin e Beijing.
- [41] Fang, K. T.; S. Kotz e K. W. Ng, (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London.
- [42] Fei, Y. e J. Pan, (2003). Influence assessments for longitudinal data in linear mixed models. Em G. Verbeke; G. Molenberghs; M. Aerts e S. Fieuws (Eds.), *Statistical Modelling: Proceedings of the 18th International Workshop on Statistical Modelling*, p. 143–148. Leuven, Bélgica.
- [43] Fernández, C. e M. F. J. Steel, (1999). Multivariate Student- t regression models: pitfalls and inference. *Biometrika* **86**, 153–167.
- [44] Fox, P. A.; A. D. Hall e N. L. Schryer, (1978). The PORT mathematical subroutine library. *ACM Transactions on Mathematical Software* **4**, 104–126.
- [45] Fuller, W. A., (1987). *Measurement Error Models*. Wiley, New York.
- [46] Galea, M.; H. Bolfarine e M. de Castro, (2002). Local influence in comparative calibration models. *Biometrical Journal* **44**, 59–81.
- [47] Galea, M.; H. Bolfarine e F. Vilca, (2005). Local influence in comparative calibration models under elliptical t -distributions. *Biometrical Journal* **47**, 691–706.

- [48] Galea, M.; G. A. Paula e H. Bolfarine, (1997). Local influence in elliptical linear regression models. *The Statistician* **46**, 71–79.
- [49] Galea, M.; G. A. Paula e F. J. A. Cysneiros, (2005). On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters* **73**, 459–467.
- [50] Galea, M.; G. A. Paula e M. Uribe-Opazo, (2003). On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers* **44**, 23–45.
- [51] Galea, M.; M. Riquelme e G. A. Paula, (2000). Diagnostics methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics* **14**, 167–184.
- [52] Garbow, B. S.; J. M. Boyle; J. J. Dongarra e C. B. Moler, (1972). *Matrix Eigensystem Routines: EISPACK Guide Extension*. Lecture Notes in Computer Science, Volume 51. Springer-Verlag, New York.
- [53] Gill, P. S., (2000). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine* **19**, 975–987.
- [54] Golub, G. H. e C. F. van Loan, (1996). *Matrix Computations, 3rd ed.*, The Johns Hopkins University Press, Baltimore.
- [55] Goodall, C. R., (1993). Computation using the QR decomposition. Em C. R. Rao (Ed.), *Handbook of Statistics, Vol. 9*, p. 467–508. Elsevier.
- [56] Gómez, E.; M. A. Gómez-Villegas e J. M. Marín, (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory & Methods* **27**, 589–600.
- [57] Gómez, E.; M. A. Gómez-Villegas e J. M. Marín, (2006). Sequences of elliptical distributions and mixtures of normal distributions. *Journal of Multivariate Analysis* **97**, 295–310.
- [58] Grubbs, F. E., (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association* **43**, 243–264.
- [59] Grubbs, F. E., (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics* **15**, 53–66.

- [60] Grubbs, F. E., (1983). Grubb's estimator. Em S. Kotz; N. L. Johnson e C. B. Read (Eds.), *Encyclopedia of Statistical Sciences*, Vol. 3, p. 542–548. Wiley.
- [61] He, X.; H. Cui e D. G. Simpson, (2004). Longitudinal data analysis using t -type regression. *Journal of Statistical Planning and Inference* **122**, 253–269.
- [62] Hoaglin, D. C. e R. E. Welsh, (1978). The hat matrix in regression and ANOVA. *The American Statistician* **32**, 17–22.
- [63] Hodgson, D. J.; O. Linton e K. Vorkink, (2002). Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. *Journal of Applied Econometrics* **17**, 617–639.
- [64] Huber, P. J., (1981). *Robust Statistics*. Wiley, New York.
- [65] Huggins, R. M., (1993a). A robust approach to the analysis of repeated measures. *Biometrics* **49**, 715–720.
- [66] Huggins, R. M., (1993b) On the robust analysis of variance components models for pedigree data. *Australian Journal of Statistics* **35**, 43–57.
- [67] Jamshidian, M., (1999). Adaptive robust regression by using a nonlinear regression program. *Journal of Statistical Software* **4**, 1–25.
- [68] Jennrich, R. I. e M. D. Schluchter, (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805–820.
- [69] Johnson, M. E. (1987). *Multivariate Statistical Simulation*. Wiley, New York.
- [70] Johnson, N. L.; S. Kotz e N. Balakrishnan (1994). *Continuous Univariate Distributions*, Vol. 1. Wiley, New York.
- [71] Johnson, N. L.; S. Kotz e N. Balakrishnan (1995). *Continuous Univariate Distributions*, Vol. 2. Wiley, New York.
- [72] Kass, R. E., (1997). Laplace's Method. Em S. Kotz; C. Read e D. Banks (Eds.), *Encyclopedia of Statistical Sciences*, Update, Vol. 1, p. 347–354. Wiley.

- [73] Kowalski, J.; J. R. Mendoza-Blanco; X. M. Tu e L. J. Gleser, (1999). On the difference in inference and prediction between the joint and independent t -error models for seemingly unrelated regressions. *Communications in Statistics - Theory & Methods* **28**, 2119–2140.
- [74] Lachos, V. H.; F. Vilca e M. Galea, (2006). Influence diagnostics for the Grubbs model. Aceito para publicação em *Statistical Papers*.
- [75] Laird, N. M. e J. H. Ware, (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
- [76] Laird, N. M.; N. Lange e D. Stram, (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association* **82**, 97–105.
- [77] Lange, K. e J. S. Sinsheimer, (1993). Normal/Independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* **2**, 175–198.
- [78] Lange, K. L.; R. J. A. Little e J. M. G. Taylor, (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881–896.
- [79] Lawson, C.; R. Hanson; D. Kincaid e F. Krogh, (1979). Basic linear algebra subprograms for Fortran usage. *ACM Transactions on Mathematical Software* **5**, 308–323.
- [80] Lee, S. e L. Xu, (2004). Influence analyses of nonlinear mixed-effects models. *Computational Statistics & Data Analysis* **45**, 321–341.
- [81] Lesaffre, E. e G. Verbeke, (1998). Local influence in linear mixed models. *Biometrics* **54**, 570–582.
- [82] Lindsey, J. K., (1999). Multivariate elliptically contoured distributions for repeated measurements. *Biometrics* **55**, 1277–1280.
- [83] Lindsey, J. K. e P. J. Lindsey, (2005). Multivariate distributions with correlation matrices for nonlinear repeated measurements. *Computational Statistics & Data Analysis* **50**, 720–732.
- [84] Lindstrom, M. J. e D. M. Bates, (1988). Newton-Raphson and EM algorithm for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* **83**, 1014–1022.

- [85] Little, R. J. A., (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics* **37**, 23–38.
- [86] Little, M. J.; G. A. Milliken; W. W. Stroup e R. D. Wolfinger, (1988). *SAS System for Mixed Models* Cary, NC. SAS Institute Inc.
- [87] Liu, C., (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association* **91**, 1219–1227.
- [88] Liu, C., (1997). ML estimation of the multivariate t distribution and the EM algorithm. *Journal of Multivariate Analysis* **63**, 296–312.
- [89] Liu, S., (2000). On local influence for elliptical linear models. *Statistical Papers* **41**, 211–224.
- [90] Liu, S., (2002). Local influence in multivariate elliptical linear regression models. *Linear Algebra and its Applications* **354**, 159–174.
- [91] Magnus, J. R. e H. Neudecker, (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.
- [92] Maronna, R. A., (1976). Robust M -estimators of multivariate location and scatter. *The Annals of Statistics* **4**, 51–67.
- [93] McCulloch, C. E. e S. R. Searle, (2001). *Generalized, Linear, and Mixed Models*. Wiley, New York.
- [94] McLachlan, G. L. e T. Krishnan, (1997). *The EM Algorithm and Extensions*. Wiley, New York.
- [95] McQuarrie, A. D. R. and C. L. Tsai, (1998). *Regression and Time Series Model Selection*. World Scientific Publishing, Singapore.
- [96] Meng, X. e D. B. Rubin, (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267–278.
- [97] Mitchell, A. F. S., (1989). The information matrix, skewness tensor and α -connections for the general multivariate elliptic distribution. *Annals of the Institute of Statistical Mathematics* **41**, 289–304.
- [98] Muirhead, R. J., (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

- [99] Nel, D. G., (1980). On matrix differentiation in statistics. *South African Statistical Journal* **14**, 137–193.
- [100] Ng, S. H.; T. Krishnan e G. J. McLachlan, (2004). The EM algorithm. Em J. E. Gentle; W. Härdle e Y. Mori (Eds.), *Handbook of Computational Statistics: Concepts and Methods*, p. 137–168. Springer.
- [101] Nobre, J. S., (2004). *Métodos de Diagnóstico para Modelos Lineares Mistos*. Dissertação de Mestrado, Departamento de Estatística, Universidade de São Paulo.
- [102] Nobre, J. S. e J. M. Singer, (2006). Fixed and random effects leverage for influence analysis in linear mixed models. *Submetido*.
- [103] Osorio, F.; G. A. Paula e M. Galea, (2006). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics & Data Analysis* (por aparecer).
- [104] Ouwens, M. J. N.; F. E. S. Tan e M. P. F. Berger, (2001). Local influence to detect influential data structures for generalized linear mixed models. *Biometrics* **57**, 1166–1172.
- [105] Pan, J. X. e K. T. Fang, (2002). *Growth Curve Models and Statistical Diagnostics*. Springer, New York.
- [106] Pan, J. X.; K. T. Fang e D. von Rosen, (1997). Local influence assessment in the growth curve model with unstructured covariance. *Journal of Statistical Planning and Inference*. **62**, 263–278.
- [107] Pan, J. X. e P. Bai, (2003). Local influence analysis in the growth curve model with Rao’s simple covariance structure. *Journal of Applied Statistics*. **7**, 771–781.
- [108] Paula, G. A., (1999). Leverage in inequality-constrained regression models. *The Statistician* **48**, 529–538.
- [109] Pendergast, J. F. e J. D. Broffitt, (1985). Robust estimation in growth curve models. *Communications in Statistics - Theory & Methods* **14**, 1919–1939.
- [110] Pinheiro, J. C. e D. M. Bates, (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics & Computing* **6**, 289–296.
- [111] Pinheiro, J. C. e D. M. Bates, (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York.

- [112] Pinheiro, J. C.; C. Liu e Y. N. Wu, (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics* **10**, 249–276.
- [113] Poon, W. e Y. S. Poon, (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society, Series B* **61**, 51–61.
- [114] Portela, J. e M. A. Gómez-Villegas, (2004). Implementation of a robust bayesian method. *Journal of Statistical Computation & Simulation* **74**, 235–248.
- [115] Potthoff, R. F. e S. N. Roy, (1964). A generalized multivariate analysis of a variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.
- [116] Pregibon, D., (1981). Logistic regression diagnostics. *The Annals of Statistics* **9**, 705–724.
- [117] R Development Core Team, (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- [118] Richardson, A. M., (1997). Bounded influence estimation in the mixed linear model. *Journal of the American Statistical Association* **92**, 154–161.
- [119] Richardson, A. M. e A. H. Welsh, (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics* **51**, 1429–1439.
- [120] Rogers, W. H. e J. W. Tuckey, (1972). Understanding some long-tailed distributions. *Statistica Neerlandica* **26**, 211–226.
- [121] Rosa, G. J. M.; C. R. Padovani e D. Gianola, (2003). Robust linear mixed models with Normal/Independent distributions and Bayesian MCMC implementation. *Biometrical Journal* **45**, 573–590.
- [122] Rosa, G. J. M.; C. R. Padovani e D. Gianola, (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *Journal of Applied Statistics* **31**, 855–873.
- [123] Ross, W. H., (1987). The geometry of case deletion and the assessment of influence in nonlinear regression. *The Canadian Journal of Statistics* **15**, 91–103.

- [124] Rousseeuw, P. J. e A. Leroy, (1987). *Robust Regression and Outlier Detection* Wiley, New York.
- [125] Savalli, C.; G. A. Paula e F. J. A. Cysneiros, (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling* **6**, 59–76.
- [126] Shi, L. e M. M. Ojeda, (2004). Local influence in multilevel regression for growth curves. *Journal of Multivariate Analysis* **91**, 282–304.
- [127] Shyr, J. Y. e L. J. Gleser, (1986). Inference about comparative precision in linear structural relationships. *Journal of Statistical Planning and Inference* **14**, 339–358.
- [128] Smith, D. M.; W. H. Robertson e P. J. Diggle, (1996). Oswald: Object-Oriented Software for the Analysis of Longitudinal Data in *S. Technical Report MA 96/192*, Department of Mathematics and Statistics, University of Lancaster, United Kingdom.
- [129] St. Laurent, R. T. e R. D. Cook, (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association* **87**, 985–990.
- [130] St. Laurent, R. T. e R. D. Cook, (1993). Leverage, local influence and curvature in nonlinear regression. *Biometrika* **80**, 99–106.
- [131] Stromberg, A. J., (2004). Why write statistical software? The case of robust statistical methods. *Journal of Statistical Software* **10**, 1–8.
- [132] Tan, F. E. S.; M. J. N. Ouwens; e M. P. F. Berger, (2001). Detection of influential observations in longitudinal mixed effects regression models. *The Statistician* **50**, 271–284.
- [133] Theobald, C. M. e J. R. Mallinson, (1978). Comparative calibration, linear structural relationships and congeneric measurements. *Biometrics* **34**, 39–45.
- [134] Tsai, C. e X. Wu, (1992). Assessing local influence in linear regression models with first-order autoregressive or heteroscedastic error structure. *Statistics & Probability Letters* **14**, 247–252.
- [135] Verbeke, G. e G. Molenberghs, (2001). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

- [136] Vonesh, E. F. e V. M. Chinchilli, (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York.
- [137] Wei, B. C.; Y. Q. Hu e W. K. Fung, (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25–37.
- [138] Welsh, A. H. e A. M. Richardson, (1997). Approaches to the robust estimation of mixed models. Em G. S. Maddala e C. R. Rao (Eds.), *Handbook of Statistics, Vol. 15*, p. 343–384. Elsevier Science.
- [139] Welsh, R. E., (1983). Leverage. Em S. Kotz; N. L. Johnson e C. B. Read (Eds.), *Encyclopedia of Statistical Sciences, Vol. 4*, p. 610–611. Wiley.
- [140] West, M., (1987). On scale mixture of normal distributions. *Biometrika* **74**, 646–648.
- [141] Wu, C. F. J., (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**, 95–103.
- [142] Wu, X. e Z. Luo, (1993). Second-order approach to local influence. *Journal of the Royal Statistical Society, Series B* **55**, 929–936.
- [143] Xiang, L.; S. Tse e A. H. Lee, (2002). Influence diagnostics for generalized linear mixed models: applications to clustered data. *Computational Statistics and Data Analysis* **40**, 759–774.
- [144] Xiang, L.; A. H. Lee e S. Tse, (2003). Assessing local cluster influence in generalized linear mixed models. *Journal of Applied Statistics* **30**, 349–359.
- [145] Zewotir, T. e J. S. Galpin, (2005). Influence diagnostics for linear mixed models. *Journal of Data Science* **3**, 153–177.
- [146] Zhu, H. e S. Lee, (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B* **63**, 111–126.
- [147] Zhu, H. e S. Lee, (2003). Local influence for generalized linear mixed models. *The Canadian Journal of Statistics* **31**, 293–309.
- [148] Zhu, H.; S. Lee; B. C. Wei e J. Zhou, (2001). Case-deletion measures for models with incomplete data. *Biometrika* **88**, 727–737.
- [149] Zhu, H. e H. Zhang, (2004). A diagnostic procedure based on local influence. *Biometrika* **91**, 579–589.