

**Modelos mistos aditivos
semiparamétricos de contornos elípticos**

Germán Mauricio Ibacache Pulgar

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientador: Prof. Dr. Gilberto Alvarenga Paula

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da
CAPES/CNPq

São Paulo, Setembro 2009

Modelos mistos aditivos semiparamétricos de contornos elípticos

Este exemplar corresponde à redação
final da tese devidamente corrigida
e defendida por Germán Mauricio Ibacache Pulgar
e aprovada pela Comissão Julgadora.

São Paulo, Setembro de 2009

Banca Examinadora:

- Prof. Dr. Gilberto Alvarenga Paula (orientador) - IME-USP.
- Prof. Dra. Viviana Giampaoli - IME-USP.
- Prof. Dr. Ronaldo Dias - UNICAMP.
- Prof. Dra. Patricia Cristina Gimenez - UNMP.
- Prof. Dr. Manuel Jesus Galea Rojas - UV.

Dedico este trabalho com muito amor a minha esposa Bernardita, a meu filhinho Jordan, e a minha mãe Myrella, pois neles encontrei a força para percorrer este caminho. Obrigado por tudo.

Em memória de Carmen e Ruth, com muita saudade.

Sumário

Agradecimentos	viii
Abreviaturas e símbolos	xiii
Lista de Figuras	xiv
Lista de Tabelas	xvii
1 Introdução	1
1.1 Motivação	1
1.2 Descrição dos objetivos	3
1.3 Organização dos capítulos	4
1.4 Distribuições de contornos elípticos	5
2 Modelo misto aditivo semiparamétrico de contornos elípticos	11
2.1 Introdução	11
2.2 Especificação do modelo	13
2.3 Representação matricial	15
2.4 Modelo normal	16
2.5 Modelo elíptico	17
2.6 Conclusões do capítulo	20
3 Estimação de máxima verossimilhança penalizada	21

3.1	Introdução	21
3.2	Critério da verossimilhança penalizada	23
3.2.1	Função de penalidade	25
3.2.2	Logaritmo da função de verossimilhança penalizada	27
3.3	Função escore penalizada	28
3.4	Matriz de informação de Fisher penalizada	31
3.5	Processo de estimação	33
3.5.1	Efeitos fixos	35
3.5.2	Componentes de variância	41
3.5.3	Processo iterativo	41
3.5.4	Efeitos aleatórios	42
3.5.5	Erro padrão	44
3.6	Modelo misto modificado normal	46
3.7	Os parâmetros de suavização	47
3.7.1	Validação cruzada	48
3.7.2	Validação cruzada generalizada	49
3.8	Seleção de modelos	49
3.9	Conclusões do capítulo	50
4	Influência local baseada na função de verossimilhança penalizada	51
4.1	Introdução	51
4.2	Método de influência local	55
4.3	Derivação da curvatura	58
4.3.1	Matriz de informação	58
4.3.2	Matriz de perturbação	61
4.4	Conclusões do capítulo	66
5	Modelo linear parcial t-Student univariado	67
5.1	Introdução	67
5.2	Especificação do modelo	69

5.3	Processo de estimação	71
5.3.1	Função escore	72
5.3.2	Matriz de informação de Fisher	73
5.3.3	Processo iterativo	74
5.3.4	Erro padrão	76
5.4	Análise de resíduos	77
5.5	Validação cruzada generalizada	79
5.6	Diagnóstico de influência local	79
5.6.1	Matriz de informação observada	80
5.6.2	Matriz de perturbação	81
5.7	Conclusões do capítulo	83
6	Modelo misto semiparamétrico t-Student	
	multivariado	84
6.1	Introdução	84
6.2	Especificação do modelo	86
6.3	Processo de estimação	89
6.3.1	Função escore	90
6.3.2	Matriz de informação de Fisher	91
6.3.3	Efeitos fixos e componentes de variância	92
6.3.4	Erro padrão	95
6.3.5	Shrinkage e valor ajustado	97
6.4	Análise de resíduos	97
6.4.1	Resíduo marginal	98
6.4.2	Resíduo condicional	99
6.4.3	Resíduos padronizados	100
6.5	Validação cruzada generalizada	101
6.6	Qualidade do ajuste: distâncias transformadas	102
6.7	Diagnóstico de influência local	103
6.7.1	Matriz de Informação observada	103
6.7.2	Esquemas de perturbação	105

6.8	Conclusões do capítulo	108
7	Aplicações	109
7.1	Dados AFP	109
7.1.1	Modelo proposto	110
7.1.2	Ajustando os modelos normal e t-Student univariado	111
7.1.3	Detectando observações aberrantes	113
7.1.4	Diagnóstico de influência	115
7.1.5	Análise confirmatória	120
7.2	Dados de glucose	121
7.2.1	Modelo proposto	122
7.2.2	Ajustando os modelos normal e t-Student multivariado	124
7.2.3	Detectando observações aberrantes	127
7.2.4	Diagnóstico de influência	128
7.3	Dados de pressão ocular	131
7.3.1	Modelo proposto	132
7.3.2	Ajustando os modelos normal e t-Student multivariado	133
7.3.3	Detectando observações aberrantes	134
7.3.4	Diagnóstico de influência	136
7.4	Conclusões do capítulo	140
8	Considerações finais	141
8.1	Discussão	141
8.2	Implementação computacional	142
8.3	Perspectivas de trabalhos futuros	142
A	Matrizes de incidência e suavização	145
A.1	Construção das matrizes	145
A.2	Programas em linguagem MATLAB	147
B	Prova de alguns resultados	149
B.1	Prova de (2.12)	149

B.2 Prova de (3.30)	150
C Derivadas da log-verossimilhança penalizada	152
C.1 Derivadas de primeiro ordem	152
C.2 Derivadas de segunda ordem	154
C.3 Matriz de informação de Fisher penalizada	158
C.3.1 Prova de (3.18)	159
C.3.2 Prova de (3.19)	161
C.3.3 Prova de (3.20)	162
C.3.4 Prova de (3.22) e (3.23)	165
C.3.5 Prova de (3.22) e (3.23)	169
D Dados de fosfato	170
E Dados pressão ocular	171
Referências Bibliográficas	172

Agradecimentos

Gostaria de agradecer, em primeiro lugar, a Deus, por me oferecer a possibilidade de continuar minha formação acadêmica e por iluminar o meu caminho com a Sua Palavra.

Ao meu orientador Prof. Dr. Gilberto Alvarenga Paula pela confiança em mim depositada e especialmente pelo constante apoio durante e após o desenvolvimento deste trabalho. Gostaria de dizer também, que me sinto honrado de ter sido orientado pelo professor Gilberto.

Ao professor Manuel Galea Rojas, um grande mestre e amigo, pelo seu constante incentivo [durante toda a minha formação acadêmica], pelos seus conselhos, e por acreditar e confiar em mim.

Desejo agradecer, em forma muito especial, a minha esposa Bernardita e meu filho Jordan Matheus, pelo amor, ternura, paciência, compressão, companheirismo, força, apoio e o tempo que me deram nos momentos mais difíceis pelos quais passei para que este trabalho se completasse [vocês mais do que ninguém sabem o valor deste trabalho]. Na verdade, faltariam as palavras para expressar o grande amor e gratidão que sinto por vocês. Simplesmente, obrigado por existir na minha vida.

À pessoa responsável de tudo que sou hoje: Myrella, minha mãe. Obrigado pelo amor, pelos carinhos, as palavras de conforto, as orações a Deus [tanto pelas madrugadas quanto pelas noites] pedindo para me abençoar, e por sempre confiar em mim, mesmo nos momentos mais difíceis da minha vida. "Mãe, esta vitória também é sua".

Gostaria de agradecer e destacar o constante apoio de minha família nesta longa jornada. Em especial, agradeço: a meu avô Julio e a minha avó Carmen

(em memória) [obrigado pelas suas orações e constante apoio espiritual durante toda a minha vida, e por ter me fornecido os princípios básicos e fundamentais para minha formação moral]; a Ruth (em memória) [simplesmente quero dizer que suas últimas palavras dirigidas a mim, naquela tarde cinzenta para seu corpo mas radiante para seu espírito, hoje são uma realidade: "Deus te abençoara"]; a Macarena [obrigado por ser minha "irmã" e estar sempre do meu lado, me substituindo em meus compromissos pessoais: "esta vitória também é de você"]; a Omar e Sergio [obrigado pelo permanente apoio durante todos estes anos e pelas palavras de estímulo que me animaram a seguir adiante]; a Willians, Fresia e Milsa [obrigado pelo carinho e constante apoio nesta jornada]; a Hector, Julio, Augusto e Romelia [obrigado pelo carinho sempre sincero que de vocês recebi]; a Carolina [obrigado pelo carinho e palavras de conforto]; a meu sogro Eduardo [obrigado pelo carinho e incentivo]; a minha cunhada Isabel e família [obrigado por nos ajudar e o apoiar durante todos estes anos nos momentos em que mais precisávamos]; e a Custodia (Toya) [obrigado pelos ensinamentos transmitidos na minha infância].

Quero agradecer, de forma muito especial, a duas grandes famílias aqui no Brasil. A primeira, composta por Tito Marcial Ibacache Gallardo e Maria Jose Silva dos Santos, por nos ajudar nos momentos em que mais precisávamos. Obrigado pela amizade e carinho. E a segunda, a Igreja Evangélica Pentecostal Missão Chilena, representada pelo querido Pastor Ricardo Lopez Meneses, esposa e filhos, pelo apoio espiritual e por ter me facilitado a moradia no meu primeiro ano de estadia no Brasil. Agradeço também aos irmãos e irmãs da Igreja: Maria Echeverria; Geraldo Leite de Moraes (em memória), esposa e filha; Zenon Ajata e esposa; Segundo Bobadilla e esposa; Maria Bobadilla e filho; e Ambart Covarruvias.

Agradeço aos professores Heleno Bolfarine e Nelson Tanaka pela grande ajuda prestada. Aos professores do Departamento de Estatística do IME-USP Fábio Prates Machado, Viviana Giampaoli, Silvia Ferrari, Heleno Bolfarine, Júlia Pavan Soler e Gilberto Paula, por participar da minha formação acadêmica. E aos professores componentes da minha banca, Patricia Cristina Gimenez e Ronaldo Dias, pelas sugestões e comentários para o melhoramento deste trabalho de tese.

Obviamente que não posso deixar de agradecer aos meus colegas e compatriotas Cristian [obrigado pelo companheirismo e lealdade durante todos os anos da minha formação acadêmica], Marco [obrigado pelos ensinamentos transmitidos durante o Mestrado e pelo carinho da sua família], Felipe, Mario e Pedro, e suas respectivas famílias, pelo apoio e amizade. Aproveito para agradecer também aos meus colegas de turma e de disciplinas Tatiane (BRA), Elsa (BRA), Gilson (BRA), Nubia (COL) e Germán (COL) pela amizade e companheirismo, e aos funcionários do departamento de Estatística e da CPG, Cecilia (chefe da secretaria do departamento), Pinho, Marilucia, Alessandra e Emerson, por sempre me atender com prontidão.

Agradeço às Universidades de Valparaíso e de São Paulo por me oferecer a oportunidade de obter a minha formação acadêmica, e o apoio financeiro da CAPES e CNPq durante o desenvolvimento deste trabalho. Finalmente, agradeço a todos aqueles que colaboraram de uma forma ou de outra para a realização deste trabalho.

São Paulo, Brasil

14 de Setembro de 2009

G.M.I.P.

Resumo

Neste trabalho estendemos os modelos mistos semiparamétricos propostos por Zhang et al. (1998) para uma classe mais geral de modelos, a qual denominamos modelos mistos aditivos semiparamétricos com erros de contornos elípticos. Com essa nova abordagem, flexibilizamos a curtose da distribuição dos erros possibilitando a escolha de distribuições com caudas mais leves ou mais pesadas do que as caudas da distribuição normal padrão. Funções de verossimilhança penalizadas são aplicadas para a obtenção das estimativas de máxima verossimilhança com os respectivos erros padrão aproximados. Essas estimativas, sob erros de caudas pesadas, são robustas no sentido da distância de Mahalanobis contra observações aberrantes. Curvaturas de influência local são obtidas segundo alguns esquemas de perturbação e gráficos de diagnóstico são propostos. Exemplos ilustrativos são apresentados em que ajustes sob erros normais são comparados, através das metodologias de sensibilidade desenvolvidas no trabalho, com ajustes sob erros de contornos elípticos.

Abstract

In this work we extend the models proposed by Zhang et al. (1998) to a more general class of models, known as semiparametric additive mixed models with elliptical errors in order to allow distributions with heavier or lighter tails than the normal ones. Penalized likelihood equations are applied to derive the maximum likelihood estimates which appear to be robust against outlying observations in the sense of the Mahalanobis distance. In order to study the sensitivity of the penalized estimates under some usual perturbation schemes in the model or data, the local influence curvatures are derived and some diagnostic graphics are proposed. Motivating examples preliminarily analyzed under normal errors are reanalyzed under some appropriate elliptical errors. The local influence approach is used to compare the sensitivity of the model estimates.

Abreviaturas e Símbolos

\mathbb{R}^n	Espaço Euclidiano n -dimensional
$\mathbf{A} \in \mathbb{R}^{n \times p}$	Matriz real de ordem $(n \times p)$
\mathbf{A}^{-1}	Inversa da matriz \mathbf{A}
\mathbf{A}^T	Transposta da matriz \mathbf{A}
$\text{rg}(\mathbf{A})$	Posto da matriz \mathbf{A}
$ \mathbf{A} $	Determinante da matriz \mathbf{A}
$\text{tr}\{\mathbf{A}\}$	Traza da matriz \mathbf{A}
$\ \mathbf{a}\ $	Norma do vetor \mathbf{a}
$\mathbf{X}, \mathbf{Z}, \mathbf{D}, \mathbf{V}, \boldsymbol{\Sigma}, \dots$	Matrizes reais
$\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{b}, \dots$	Vetores (aleatórios ou de constantes)
$\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \dots$	Vetores de parâmetros
$\varphi_{\mathbf{y}}(\cdot)$	Função geradora de funções característica de \mathbf{y}
$\psi_{\mathbf{y}}(\cdot)$	Função característica de \mathbf{y}
$f_{\mathbf{y}}(\cdot)$	Função densidade de \mathbf{y}
$g(\cdot)$	Função geradora de densidades
$E(\mathbf{y})$	Valor esperado de \mathbf{y}
$\text{Var}(\mathbf{y})$	Variância de \mathbf{y}
$\text{Cov}(\mathbf{y})$	Covariância de \mathbf{y}
$\mathcal{E}l_n(\cdot, \cdot, \cdot)$ ou $\mathcal{E}l_n(\cdot, \cdot)$	Distribuição elíptica n -variada
$S_n(\cdot)$	Distribuição esférica n -variada
$\mathcal{N}_n(\cdot, \cdot)$	Distribuição normal n -variada
χ_m^2	Distribuição χ^2 com m graus de liberdade (g.l.)
$\mathcal{F}_{(v_1, v_2)}$	Distribuição \mathcal{F} com v_1 e v_2 graus de liberdade
t-Student	Distribuição t-Student univariada
t -Student	Distribuição t -Student n -variada
\sim	Distribuição
\approx	Aproximado
$\hat{\boldsymbol{\theta}}$	Estimativa (ou estimador) do vetor de parâmetros $\boldsymbol{\theta}$

Lista de Figuras

1.1	Gráficos da distribuição normal bivariada padrão: densidade conjunta (a) e contornos da densidade (b).	9
1.2	Gráficos da distribuição t-Student bivariada padrão com 4 graus de liberdade: densidade conjunta (a) e contornos da densidade (b).	9
1.3	Gráficos da distribuição de Cauchy bivariada padrão: densidade conjunta (a) e contornos da densidade (b).	10
7.1	Diagrama de dispersão: rentabilidades da AFP <i>Cuprum</i> contra as rentabilidades do IPSA (a), rentabilidades contra o tempo (b).	110
7.2	Gráficos dos ajustes não paramétricos (linha contínua) aos dados de AFP sob o modelo normal e t-Student com $\nu = 4$ graus de liberdade.	113
7.3	Gráficos normais de probabilidade para a distância transformada sob os modelos normal (a) e t-Student (b) ajustados aos dados de AFP.	113
7.4	Gráficos de índices para a distância $\hat{\delta}_i$ sob os modelos normal (a) e t-Student (b) ajustados aos dados de AFP, e entre os pesos estimados e a distância $\hat{\delta}_i$ sob o modelo t-Student (c).	114
7.5	Gráficos de índices de C_i para $\hat{\beta}$ sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de AFP.	115
7.6	Gráficos de índices de C_i para \hat{f} sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de AFP.	116
7.7	Gráficos de índices de C_i para $\hat{\phi}$ sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de AFP.	116

7.8	Gráficos de índices de C_i para $\hat{\beta}$ sob perturbação de escala para os modelos normal e t-Student ajustados aos dados de AFP.	117
7.9	Gráficos de índices de C_i para \hat{f} sob perturbação de escala para os modelos normal e t-Student ajustados aos dados de AFP.	117
7.10	Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação de escala para os modelos normal e t-Student ajustados aos dados de AFP.	117
7.11	Gráficos de índices de C_i para $\hat{\beta}$ sob perturbação do IPSA para os modelos normal e t-Student ajustados aos dados de AFP.	118
7.12	Gráficos de índices de C_i para \hat{f} sob perturbação do IPSA para os modelos normal e t-Student ajustados aos dados de AFP.	118
7.13	Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação do IPSA para os modelos normal e t-Student ajustados aos dados de AFP.	119
7.14	Gráficos de índices: resíduos padronizados (a), pontos de alavanca (b), distâncias de Cook para $\hat{\beta}$ (c) e medida DFITS para \hat{f} (d), referentes ao ajuste do modelo normal aos dados de AFP.	119
7.15	Gráficos de perfis (a) e das médias (b) de fosfato inorgânico no plasma.	123
7.16	Gráficos normais de probabilidade das distâncias transformadas sob os modelos normal e t-Student com 7 graus de liberdade ajustados aos dados de glucose.	125
7.17	Gráfico de perfis ajustados (–) sob o modelo t-Student com 7 graus de liberdade para os 20 pacientes do estudo de dados de glucose (da esquerda para a direita).	126
7.18	Gráficos de índices para a distância $\hat{\delta}_i$ sob os modelos normal (a) e t-Student (b) ajustados aos dados de glucose, e entre os pesos estimados e a distância $\hat{\delta}_i$ sob o modelo t-Student (c).	127
7.19	Gráficos de índices de C_i para \hat{f} sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de glucose.	128
7.20	Gráficos de índices de C_i para $\hat{\lambda}$ sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de glucose.	129

7.21	Gráficos de índices de C_i para $\hat{\phi}$ sob ponderação de casos para os modelos normal e t -Student ajustados aos dados de glucose.	129
7.22	Gráficos de índices de C_i para \hat{f} sob perturbação de escala para os modelos normal e t -Student ajustados aos dados de glucose.	130
7.23	Gráficos de índices de C_i para $\hat{\lambda}$ sob perturbação de escala para os modelos normal e t -Student ajustados aos dados de glucose.	130
7.24	Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação de escala para os modelos normal e t -Student ajustados aos dados de glucose.	130
7.25	Gráficos de perfis (a) e das médias (b) da pressão ocular do olho esquerdo.	132
7.26	Gráficos normais de Probabilidades das distâncias transformadas sob os modelos normal e t -Student com 9 graus de liberdade ajustados aos dados sobre pressão ocular.	135
7.27	Gráficos de índices para a distância $\hat{\delta}_i$ sob os modelos normal (a) e t -Student (b) ajustados aos dados sobre pressão ocular, e entre os pesos estimados e a distância $\hat{\delta}_i$ sob o modelo t -Student (c).	136
7.28	Gráficos de índices de C_i para \hat{f} sob ponderação de casos para os modelos normal e t -Student ajustados aos dados sobre pressão ocular.137	
7.29	Gráficos de índices de C_i para $\hat{\lambda}$ sob ponderação de casos para os modelos normal e t -Student ajustados aos dados sobre pressão ocular.137	
7.30	Gráficos de índices de C_i para $\hat{\phi}$ sob ponderação de casos para os modelos normal e t -Student ajustados aos dados sobre pressão ocular.138	
7.31	Gráficos de índices de C_i para \hat{f} sob perturbação de escala para os modelos normal e t -Student ajustados aos dados sobre pressão ocular.138	
7.32	Gráficos de índices de C_i para $\hat{\lambda}$ sob perturbação de escala para os modelos normal e t -Student ajustados aos dados sobre pressão ocular.139	
7.33	Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação de escala para os modelos normal e t -Student ajustados aos dados sobre pressão ocular.139	

Lista de Tabelas

1.1	Algumas distribuições pertencentes à classe das distribuições de contornos elípticos. c é uma constante de normalização.	8
3.1	Expressões das quantidades v_i para algumas distribuições elípticas. .	30
3.2	$(u^* + 1)$ -ésima etapa do algoritmo <i>backfitting</i> na $(u + 1)$ -ésima etapa do processo iterativo score de Fisher sob o modelo misto aditivo semiparamétrico elíptico.	40
5.1	$(u^* + 1)$ -ésima etapa do algoritmo <i>backfitting</i> na $(u + 1)$ -ésima etapa do processo iterativo score de Fisher sob o modelo linear parcial t-Student univariado.	75
6.1	$(u^* + 1)$ -ésima etapa do algoritmo <i>backfitting</i> na $(u + 1)$ -ésima etapa do processo iterativo score de Fisher sob o modelo misto semiparamétrico t-Student multivariado.	94
7.1	Valores estimados do logaritmo da função de verossimilhança penalizada e do critério de informação de Schwarz sob o modelo t-Student para diferentes graus de liberdade ajustado aos dados de AFP.	112
7.2	Estimativas de máxima verossimilhança penalizada para os modelos normal e t-Student com $\nu = 4$ graus de liberdade ajustados aos dados de AFP.	112

7.3	Mudanças relativas (em parênteses) nas estimativas de máxima verossimilhança penalizada para os modelos normal e t-Student com $\nu = 4$ graus de liberdade ajustados aos dados de AFP.	121
7.4	Valores estimados do logaritmo da função de verossimilhança penalizada e do critério de informação de Schwarz sob o modelo t-Student multivariado para diferentes graus de liberdade ajustado aos dados de glucose.	124
7.5	Estimativas de máxima verossimilhança penalizada para os modelos normal e t-Student multivariado com $\nu = 7$ graus de liberdade ajustados aos dados de glucose.	125
7.6	Valores estimados do logaritmo da função de verossimilhança penalizada e do critério de informação de Schwarz sob o modelo t-Student para diferentes graus de liberdade ajustado aos dados sobre pressão ocular.	134
7.7	Estimativas de máxima verossimilhança penalizada para os modelos normal e t-Student com $\nu = 9$ graus de liberdade ajustados aos dados sobre pressão ocular.	135
A.1	Descrição dos valores da variável t_i para o modelo de regressão não paramétrico clássico.	145
D.1	Medições de fosfato inorgânico no plasma obtidas a partir de amostras de sangue de 20 pacientes obesos, registradas após da aplicação oral de uma dose de glucose padrão (mg/dd).	170
E.1	Medições da pressão ocular do olho esquerdo de 20 pacientes registradas em intervalos de tempo de três horas.	171

Capítulo 1

Introdução

1.1 Motivação

É sabido que os modelos estatísticos permitem incorporar relações matemáticas existentes entre dados e parâmetros de interesse. As formas mais simples dessas relações são a linearidade e a suposição paramétrica. Os modelos mistos são um exemplo de tais modelos e são considerados uma ferramenta poderosa na modelagem estatística devido à sua flexibilidade para modelar a estrutura de variância-covariância entre e intraunidades amostrais, à sua capacidade de tratar com dados balanceados como desbalanceados, e à disponibilidade de pacotes confiáveis para o ajuste desses modelos. Esses modelos são frequentemente utilizados em pesquisas relacionadas com esquemas de amostragem longitudinal, espacial e hierárquico, e caracterizam-se por levar em consideração uma estrutura linear entre seus componentes (efeitos fixos e aleatórios) e assumir que os efeitos fixos contribuem de maneira paramétrica no valor esperado da variável resposta. Uma extensa teoria para o caso normal é apresentada em detalhe em Verbeke e Molenberghs (2001). Porém, diversos trabalhos da literatura mostram que muitas vezes a suposição paramétrica imposta sobre os modelos mistos não é apropriada, principalmente quando existem covariáveis que podem contribuir de maneira não paramétrica. Em tais casos, alguns autores têm proposto quantificar os efeitos dessas covariáveis através de funções arbitrárias. Essa situação tem motivado o desenvolvimento de

modelos mais flexíveis, como por exemplo, os modelos mistos estocásticos semiparamétricos para dados longitudinais e os modelos mistos aditivos generalizados. Especificamente, os modelos mistos estocásticos semiparamétricos permitem modelar o valor esperado da variável resposta através de uma função arbitrária do tempo e de funções paramétricas das covariáveis, e flexibilizar a estrutura de variância-covariância incorporando processos estocásticos estacionários e não estacionários para modelar a correlação serial (Zhang et al., 1998). Por sua parte, os modelos mistos aditivos generalizados usam funções não paramétricas aditivas para modelar os efeitos das covariáveis e efeitos aleatórios para modelar a superdispersão e a correlação (Lin e Zhang, 1999).

Devido ao evidente surgimento de novos planejamentos para representar fenômenos cada vez mais complexos, surge a necessidade de construir modelos que possam levar em consideração as diversas características desses estudos. Neste trabalho propomos uma nova classe de modelos a qual denominamos modelos mistos aditivos semiparamétricos de contornos elípticos. Esses modelos constituem uma alternativa muito atrativa para explicar a dependência das medidas intraunidades amostrais através dos efeitos das variáveis explicativas que podem contribuir tanto de maneira paramétrica quanto de maneira não paramétrica. Além disso, esses modelos têm a flexibilidade de estender a classe dos erros para outras distribuições elípticas além da distribuição normal, e acomodar observações aberrantes através de distribuições com caudas mais pesadas do que a normal. Portanto, a principal contribuição deste trabalho é propor uma nova classe de modelos, desenvolver processos de estimação e inferência para os parâmetros e métodos de diagnóstico.

Embora seja bastante conhecida a robustez das estimativas de máxima verossimilhança dos modelos elípticos paramétricos contra observações aberrantes, a extensão para modelos semiparamétricos ainda não foi estudada e é um dos objetivos deste trabalho. Contudo, de forma similar ao caso paramétrico, essa robustez não é diretamente estendida quando perturbações são aplicadas no modelo ou nos dados. Assim, estudos de sensibilidade são importantes e devem ser considerados

em modelos de contornos elípticos. Dessa forma, aplicamos a metodologia de influência local aos modelos mistos aditivos semiparamétricos de contornos elípticos para avaliar a sensibilidade das estimativas de máxima verossimilhança penalizada sob algum esquema de perturbação. No trabalho estendemos as metodologias apresentadas em Zhu et al. (2003) que tratam modelos lineares parciais com erros normais. Exemplos ilustrativos são apresentados em que ajustes sob erros normais são comparados, através das metodologias de sensibilidade desenvolvidas no trabalho, com ajustes sob erros de contornos elípticos.

1.2 Descrição dos objetivos

Os principais objetivos deste trabalho são: (1) estudar os modelos mistos aditivos semiparamétricos sob a classe das distribuições elípticas e (2) aplicar a técnica de diagnóstico de influência local (Cook, 1986) nessa nova classe de modelos. Neste trabalho introduzimos as distribuições elípticas nos modelos mistos aditivos semiparamétricos assumindo que a distribuição conjunta do vetor de respostas, dos efeitos aleatórios e dos erros aleatórios segue uma distribuição elíptica. Consequentemente, obtemos em forma fechada a distribuição marginal do vetor de respostas e, a partir dessa distribuição, fazemos inferências e diagnóstico no modelo. Dentro desse contexto, temos os seguintes objetivos específicos:

- (i) definir a função de verossimilhança penalizada para os modelos mistos semiparamétricos de contornos elípticos e calcular a função score e a matriz de informação de Fisher para os parâmetros envolvidos no modelo;
- (ii) propor e implementar em um ambiente computacional um método para estimar os coeficientes da regressão, as funções não paramétricas, e os componentes de variância do modelo;
- (iii) derivar as ferramentas necessárias para o procedimento de influência local, tais como, a matriz de informação observada e a matriz de perturbações sob diferentes esquemas de perturbação;

- (iv) particularizar os resultados descritos em (i)-(iii) para o modelo linear parcial t-Student univariado e o modelo misto semiparamétrico t-Student multivariado; e,
- (v) avaliar a capacidade do método de influência local para detectar observações influentes, através de três aplicações com dados reais.

1.3 Organização dos capítulos

Este trabalho de tese é organizado da seguinte maneira. O Capítulo 2 visa descrever as características essenciais do modelo misto aditivo semiparamétrico elíptico. O Capítulo 3 descreve em detalhes a idéia e os elementos fundamentais para construir um procedimento adequado que permita estimar todos os parâmetros envolvidos no modelo misto aditivo semiparamétrico elíptico. Especificamente, definimos o logaritmo da função de verossimilhança penalizada e calculamos a função score e a matriz de informação de Fisher. A partir desses resultados, construímos um processo iterativo para estimar os coeficientes da regressão, as funções não paramétricas, e os componentes de variância. Já no final do capítulo discutimos o problema de estimação dos efeitos aleatórios e do parâmetro de suavização, e a seleção de modelos. O Capítulo 4 traz o desenvolvimento do método de influência local para os modelos mistos aditivos semiparamétricos elípticos que são, por sua vez, aplicados nos capítulos seguintes a dados reais. Nesse contexto, derivamos a curvatura normal considerando diferentes esquemas de perturbação. Os Capítulos 5 e 6 particularizam os resultados inferenciais e de diagnóstico obtidos e apresentados nos capítulos anteriores para o modelo linear parcial t-Student univariado e o modelo misto semiparamétrico t-Student multivariado. O Capítulo 7 ilustra os resultados inferenciais e de diagnóstico de influência local através de três aplicações a conjuntos de dados reais. Por fim, é apresentada no Capítulo 8 uma discussão sobre os resultados, as principais conclusões e as perspectivas futuras deste trabalho.

1.4 Distribuições de contornos elípticos

A fim de introduzirmos os modelos mistos aditivos semiparamétricos de contornos elípticos, apresentamos nesta seção a classe de distribuições de contornos elípticos e algumas das principais propriedades teóricas requeridas para o desenvolvimento do presente trabalho. As distribuições de contornos elípticos vêm ocupando um importante e crescente lugar na teoria estatística, e formam uma classe generalizada de famílias de distribuições que preservam a estrutura simétrica da distribuição normal. De fato, a distribuição normal é um elemento particular dessa classe. Além da distribuição normal, outros elementos típicos na classe das distribuições de contornos elípticos são, por exemplo, a distribuição normal contaminada, a distribuição t-Student e a distribuição exponencial potência. Embora o interesse por estas distribuições seja relativamente recente, muitas de suas propriedades têm sido bem estudadas por diferentes autores; veja, por exemplo, Fang e Zhang (1990), Fang et al. (1990) e Arellano (1994).

Definição 1.4.1 *Diz-se que o vetor aleatório $\mathbf{y} \in \mathbb{R}^n$ ($n \geq 2$) segue uma distribuição de contornos elípticos se sua função característica tem a forma*

$$\psi_{\mathbf{y}}(\mathbf{t}) = \exp\{i \mathbf{t}^T \boldsymbol{\mu}\} \varphi(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}), \quad (1.1)$$

em que $\boldsymbol{\mu} \in \mathbb{R}^n$ denota o parâmetro de posição, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ denota o parâmetro de escala (matriz simétrica e positiva semidefinida), $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função geradora de funções características, $i = \sqrt{-1}$ e $\mathbf{t} \in \mathbb{R}^n$.

Se \mathbf{y} tem distribuição elíptica com função característica dada por (1.1), escrevemos $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ ou simplesmente $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Note que isto não significa que \mathbf{y} tenha uma particular distribuição elíptica, só indica que sua distribuição pertence à classe de distribuições elípticas multivariadas. No caso específico em que $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \mathbf{I}_n$, a distribuição de \mathbf{y} é denominada de esférica n -variada e denota-se $\mathbf{y} \sim \mathcal{S}_n(\varphi)$. A seguir são apresentadas algumas das principais propriedades das distribuições de contornos elípticos.

Propriedade 1.4.2 Assuma que $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ com $\text{rg}(\boldsymbol{\Sigma}) = r < n$ (posto). Se o valor esperado e a variância do vetor aleatório \mathbf{y} existem, então

(a) $E(\mathbf{y}) = \boldsymbol{\mu}$ e

(b) $\text{Var}(\mathbf{y}) = \kappa \boldsymbol{\Sigma}$, em que a constante $\kappa > 0$ é dada por

$$\kappa = -2 \varphi'(0) = -2 \left. \frac{d\varphi(u)}{du} \right|_{u=0}, \quad (1.2)$$

com $\varphi(\cdot)$ sendo a função geradora de funções características. Em particular, para a distribuição *t-Student*, temos que $\kappa = \nu/(\nu - 2)$ ($\nu > 2$), em que ν denota o grau de liberdade da distribuição.

A Propriedade 1.4.2 permite caracterizar os modelos mistos aditivos semiparamétricos de contornos elípticos através dos parâmetros de posição e de escala. Nesse último caso, podemos observar que a matriz de variância-covariância do vetor aleatório é definida em termos da constante κ que depende diretamente da distribuição elíptica em consideração. Em particular, se $\mathbf{y} \sim \mathcal{S}_n(\varphi)$, então κ é a variância das distribuições marginais univariadas, e se $\mathbf{y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\kappa = 1$.

Propriedade 1.4.3 Assuma que $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ com $\text{rg}(\boldsymbol{\Sigma}) = r < n$. Se \mathbf{B} é uma matriz ($n \times m$) e $\boldsymbol{\delta}$ é um vetor ($m \times 1$), então

$$\boldsymbol{\delta} + \mathbf{B}^T \mathbf{y} \sim \mathcal{E}l_m(\boldsymbol{\delta} + \mathbf{B}^T \boldsymbol{\mu}, \mathbf{B}^T \boldsymbol{\Sigma} \mathbf{B}, \varphi). \quad (1.3)$$

Em particular, se consideramos a partição

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (1.4)$$

obtemos as seguintes distribuições marginais:

(a) $\mathbf{y}^{(1)} \sim \mathcal{E}l_m(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}, \varphi)$ e

(b) $\mathbf{y}^{(2)} \sim \mathcal{E}l_{(n-m)}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}, \varphi)$.

Salientamos que a importância da Propriedade 1.4.3 baseia-se principalmente no fato de que uma transformação linear de um vetor aleatório com distribuição elíptica também segue uma distribuição elíptica. Além disso, podemos observar que cada elemento do vetor aleatório \mathbf{y} tem uma distribuição marginal elíptica.

Propriedade 1.4.4 *Assuma que $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ com $\boldsymbol{\Sigma} \geq 0$. Se consideramos a partição (1.4), temos que*

$$(\mathbf{y}^{(1)} | \mathbf{y}_0^{(2)}) \sim \mathcal{E}l_m(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}, \varphi_{q(\mathbf{y}_0^{(2)})}), \quad (1.5)$$

em que

$$\begin{aligned} \boldsymbol{\mu}_{1.2} &= \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{y}_0^{(2)} - \boldsymbol{\mu}^{(2)}), \\ \boldsymbol{\Sigma}_{11.2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \quad \text{e} \\ q(\mathbf{y}_0^{(2)}) &= (\mathbf{y}_0^{(2)} - \boldsymbol{\mu}^{(2)})^T \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{y}_0^{(2)} - \boldsymbol{\mu}^{(2)}). \end{aligned}$$

Analogamente,

$$(\mathbf{y}^{(2)} | \mathbf{y}_0^{(1)}) \sim \mathcal{E}l_m(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}, \varphi_{q(\mathbf{y}_0^{(1)})}), \quad (1.6)$$

em que

$$\begin{aligned} \boldsymbol{\mu}_{2.1} &= \boldsymbol{\mu}^{(2)} + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{y}_0^{(1)} - \boldsymbol{\mu}^{(1)}), \\ \boldsymbol{\Sigma}_{22.1} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \quad \text{e} \\ q(\mathbf{y}_0^{(1)}) &= (\mathbf{y}_0^{(1)} - \boldsymbol{\mu}^{(1)})^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{y}_0^{(1)} - \boldsymbol{\mu}^{(1)}). \end{aligned}$$

Em geral, existem situações nas quais o vetor aleatório \mathbf{y} não necessariamente possui uma função densidade. Por exemplo, quando o $\text{rg}(\boldsymbol{\Sigma}) = r (< n)$ (caso singular), a função densidade não existe em todo o espaço \mathbb{R}^n . Porém, sempre é possível definir a função densidade em um espaço de dimensão menor ou igual ao posto da matriz de escala. No entanto, se o $\text{rg}(\boldsymbol{\Sigma}) = n$ (caso não singular), a função densidade existe em relação à medida de Lebesgue sobre todo o espaço \mathbb{R}^n .

Definição 1.4.5 *Assumindo que o $\text{rg}(\Sigma) = n$, temos que a função densidade do vetor aleatório \mathbf{y} assume a forma*

$$f_{\mathbf{y}}(\mathbf{y}) = |\Sigma|^{-1/2} g(u), \quad (1.7)$$

em que $u = (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})$ e g é uma função escalar contínua e diferenciável de $\mathbb{R} \rightarrow [0, \infty]$, tipicamente chamada função geradora de densidades, que satisfaz a condição

$$\int_0^{\infty} u^{-1/2} g(u) du < \infty. \quad (1.8)$$

Se \mathbf{y} tem distribuição elíptica com densidade dada por (1.7), escrevemos $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \Sigma, g)$ ou simplesmente $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \Sigma)$. Na Tabela 1.1 são apresentadas algumas distribuições que pertencem à classe das distribuições elípticas. Nas figuras 1.1-1.3 são apresentados os gráficos das funções densidade e seus correspondentes contornos, das distribuições normal, t-Student e Cauchy multivariadas .

Tabela 1.1: Algumas distribuições pertencentes à classe das distribuições de contornos elípticos. c é uma constante de normalização.

Distribuição	$g(u)$	
Tipo Kotz	$cu^{N-1} \exp\{-ru^s\}$	$r, s > 0$ e $2N + n > 2$
Normal	$c \exp\{-u/2\}$	$u \geq 0$
Exponencial Potência	$c \exp\{-u^s/2\}$	
Pearson tipo VII	$c \left\{1 + \frac{u}{s}\right\}^{-N}$	$N > n/2$ e $s > 0$
t-Student	$c \left\{1 + \frac{u}{s}\right\}^{-(\nu+m)/2}$	$m > 0$
Cauchy	$c \left\{1 + \frac{u}{s}\right\}^{-(\nu+1)/2}$	$s > 0$
Pearson tipo II	$c \{1 - u\}^m$	$m > 0$
Logística	$c \exp\{-u\} / [1 + \exp\{-u\}]^2$	$u \geq 0$
Mistura de escala	$c \int_0^{\infty} t^{-n/2} \exp\{-u/2t\} dG(t)$	$G(t)$: f.d.a.

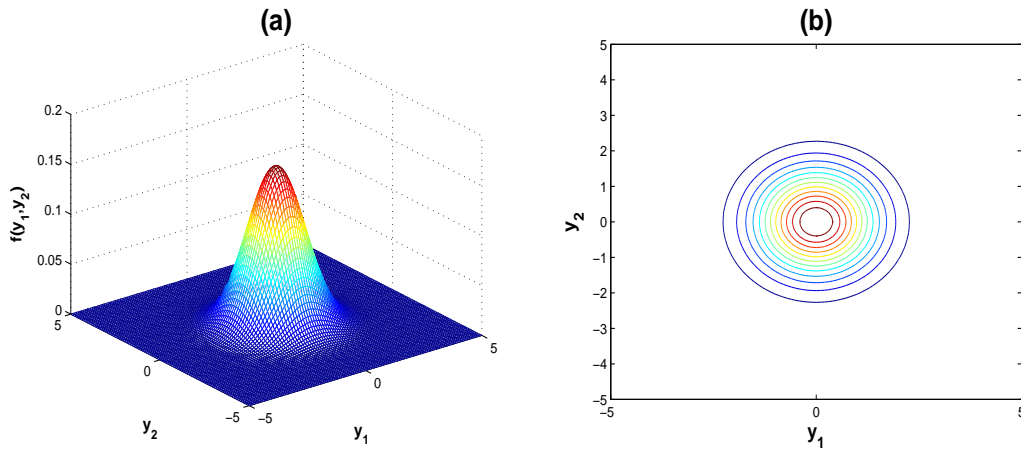


Figura 1.1: Gráficos da distribuição normal bivariada padrão: densidade conjunta (a) e contornos da densidade (b).

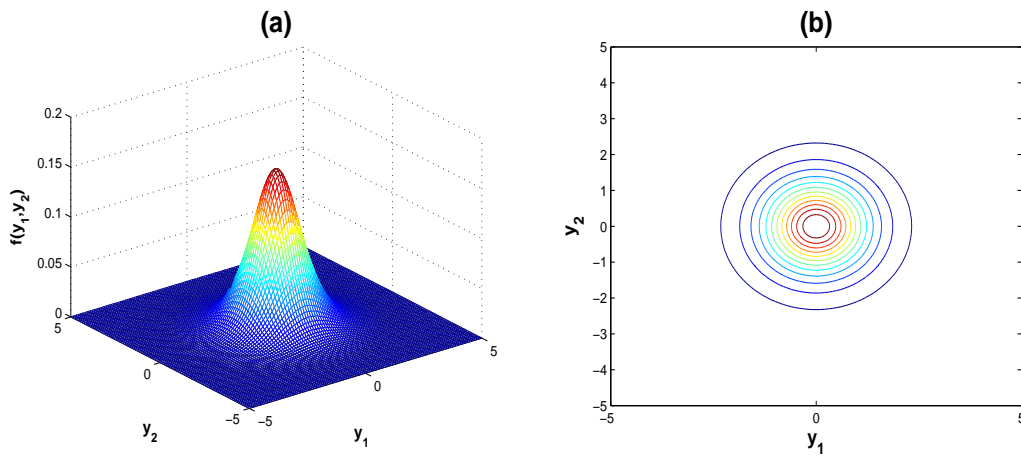


Figura 1.2: Gráficos da distribuição t -Student bivariada padrão com 4 graus de liberdade: densidade conjunta (a) e contornos da densidade (b).

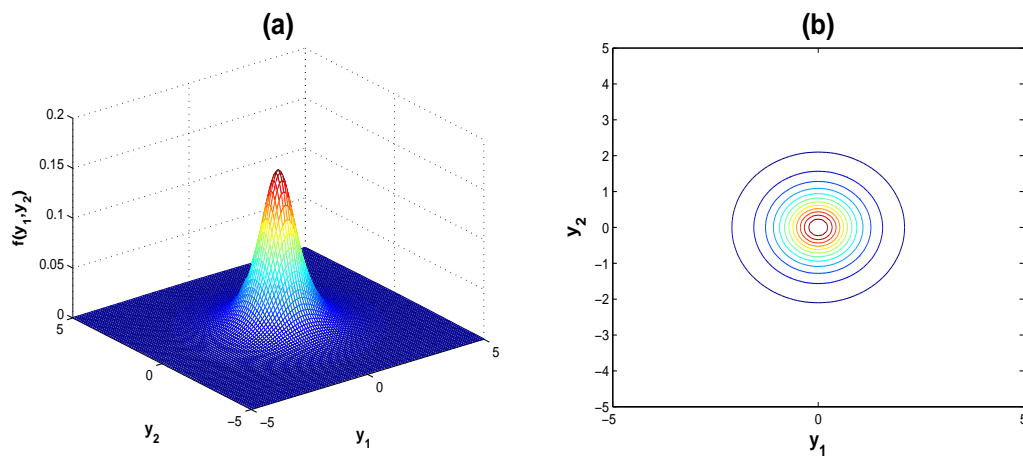


Figura 1.3: Gráficos da distribuição de Cauchy bivariada padrão: densidade conjunta (a) e contornos da densidade (b).

Capítulo 2

Modelo misto aditivo semiparamétrico de contornos elípticos

Neste capítulo introduzimos o modelo misto aditivo semiparamétrico de contornos elípticos. No início do capítulo, Seção 1, apresentamos uma discussão sucinta sobre os principais trabalhos relacionados a esta classe de modelos. Na Seção 2, especificamos o modelo misto aditivo semiparamétrico e listamos alguns casos particulares. A seguir, na Seção 3, apresentamos o modelo misto aditivo semiparamétrico na sua versão matricial. Na Seção 4, definimos os modelos mistos aditivos semiparamétricos no caso normal. Posteriormente, na Seção 5, definimos dois enfoques para introduzir as distribuições de contornos elípticos no modelo misto aditivo semiparamétrico. E no final do capítulo, Seção 6, são apresentadas algumas das principais conclusões.

2.1 Introdução

Os modelos mistos aditivos semiparamétricos de contornos elípticos emergem como uma ferramenta poderosa na modelagem estatística devido à sua flexibilidade para modelar a estrutura de variância-covariância entre e intraunidades amostrais, modelar a dependência das medidas intraunidades amostrais através dos efeitos das variáveis explicativas que podem contribuir de maneira paramétrica quanto de maneira não paramétrica (pelo menos duas variáveis explicativas), à sua ca-

pacidade de tratar com dados balanceados como desbalanceados, e à flexibilidade de estender a classe dos erros para outras distribuições elípticas além da distribuição normal, permitindo, dessa forma, acomodar observações aberrantes através de distribuições com caudas mais pesadas do que a distribuição normal. Tais modelos são frequentemente utilizados em pesquisas relacionadas com esquemas de amostragem longitudinal, espacial e hierárquico.

Alguns trabalhos relacionados com o modelo misto aditivo semiparamétrico são descritos a seguir. Hastie e Tibshirani (1986) introduzem o modelo aditivo no contexto dos modelos lineares generalizados. Nesse modelo, a função linear dos preditores é substituída por uma soma de funções arbitrárias que, usualmente, é estimada através de algum procedimento de suavização; veja também Hastie e Tibshirani (1993). Recentemente, Rigby e Stasinopoulos (2005) propuseram uma nova classe de modelos aditivos generalizados, em que a suposição de família exponencial é relaxada e substituída por uma família de distribuições mais geral. Uma das vantagens desse modelo é que sua parte sistemática permite modelar, além da sua média, funções não paramétricas.

Lin e Zhang (1999) propuseram o modelo misto aditivo generalizado, o qual corresponde a uma extensão do modelo misto generalizado tratado por Breslow e Clayton (1993). Tais modelos são úteis na análise de dados com medidas repetidas quando a variável resposta pertence à família exponencial. Algumas aplicações do modelo misto aditivo generalizado podem ser encontradas em Parise et al. (2001) e Fahrmeir e Lang (2001). Wang (1998) propõe o modelo misto não paramétrico como uma família mais geral de modelos não paramétricos (veja também Gu e Ma, 2005). Outros trabalhos relacionados aos modelos semiparamétricos podem ser encontrados em Heckman (1986), Speckman (1988) e Green e Silverman (1994).

2.2 Especificação do modelo

Seja y_{ij} ($i = 1, \dots, n; j = 1, \dots, m_i$) a j -ésima medição associada ao i -ésimo grupo. O modelo misto aditivo semiparamétrico (MMAS) assume a seguinte forma:

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i + f_1(t_{1ij}) + \dots + f_s(t_{sij}) + \varepsilon_{ij}, \quad (2.1)$$

em que \mathbf{x}_{ij} , \mathbf{z}_{ij} e $\mathbf{t}_{k_i} = (t_{k_{i1}}, \dots, t_{k_{im_i}})^T$ ($k = 1, \dots, s$) são, respectivamente, vetores $(p \times 1)$, $(q \times 1)$ e $(m_i \times 1)$ que contêm valores de variáveis explicativas; $\boldsymbol{\beta}$ é um vetor (efeitos fixos) de parâmetros desconhecido; \mathbf{b}_i é um vetor de variáveis latentes $(q \times 1)$, comumente denominadas efeitos aleatórios, que refletem o comportamento individual do i -ésimo grupo; f_k é uma função univariada arbitrária "suave" que quantifica o efeito da k -ésima variável explicativa que contribui de maneira não paramétrica sobre a variável resposta y_{ij} ; e ε_{ij} é um erro aleatório não observado.

Os efeitos fixos do modelo (2.1) são usados para modelar o valor esperado de y_{ij} e estão representados pelos componentes paramétrico e não paramétrico, respectivamente, associados ao coeficiente de regressão $\boldsymbol{\beta}$, que pertence a um espaço Euclidiano finito dimensional (usualmente \mathbb{R}^p), e às funções f_k ($k = 1, \dots, s$), que pertencem a um espaço infinito dimensional. Neste caso, as funções f_k ($k = 1, \dots, s$) são consideradas como parâmetros com respeito ao valor esperado de y_{ij} e no processo de estimação deve-se especificar o espaço ao qual pertencem. Essa escolha tipicamente depende do domínio da função, de algum conhecimento a priori da forma da função, de restrições para garantir a identificabilidade, ou simplesmente de alguma aplicação específica. Por sua parte, os efeitos aleatórios são incorporados para modelar a estrutura de variância-covariância.

A natureza semiparamétrica do modelo (2.1), com respeito aos modelos paramétricos e não paramétricos, oferece maior flexibilidade na modelagem de relações complexas entre a variável y_{ij} e as covariáveis, mantendo uma relação funcional linear entre seus componentes. Alguns casos particulares do modelo (2.1) tem sido estudado por vários e na maioria desses casos os resultados inferenciais e de diagnóstico baseiam-se na distribuição normal. Alguns exemplos são dados abaixo.

Modelo misto semiparamétrico

Quando $s = 1$ e $\epsilon_{ij} = u_i(t_{ij}) + \varepsilon_{ij}$ ($i = 1, \dots, n; j = 1, \dots, m_i$), com $u_i(t)$ sendo um processo estocástico independente de ε_{ij} , o modelo (2.1) corresponde ao modelo misto semiparamétrico apresentado por Zhang et al. (1998) e assume a forma

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i + f(t_{ij}) + u_i(t_{ij}) + \varepsilon_{ij}, \quad (2.2)$$

em que os efeitos aleatórios e os erros aleatórios seguem uma distribuição normal e os processos estocásticos são definidos como um processo gaussiano. Em particular, quando $\mathbf{b}_i = \mathbf{0}$, o modelo (2.2) reduz-se ao modelo semiparamétrico proposto por Zeger e Diggle (1994), em que os erros aleatórios seguem uma distribuição normal e os processos estocásticos são definidos como processos gaussianos estacionários.

Modelo misto não paramétrico

Quando $s = 1$ e $\boldsymbol{\beta} = \mathbf{0}$, o modelo (2.1) reduz-se ao modelo misto não paramétrico proposto por Wang (1998). Nesse caso, o modelo assume a forma

$$y_{ij} = \mathbf{z}_{ij}^T \mathbf{b}_i + f(t_{ij}) + \epsilon_{ij}, \quad (2.3)$$

($i = 1, \dots, n; j = 1, \dots, m_i$) em que os erros aleatórios e os efeitos aleatórios seguem uma distribuição normal.

Modelo misto

No caso em que $f_k(\cdot) = 0$ ($k = 1, \dots, s$), o modelo (2.1) reduz-se ao modelo misto proposto por Laird e Ware (1982). Nesse caso, o modelo é dado por

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i + \epsilon_{ij}, \quad (2.4)$$

($i = 1, \dots, n; j = 1, \dots, m_i$) em que os efeitos aleatórios e os erros aleatórios seguem uma distribuição normal. Osorio (2006) estende o modelo (2.4) para a classe de distribuições elípticas e desenvolve a técnica de influência local. Lindstrom e Bates (1990) estendem esse modelo para o caso não linear, e posteriormente Ke e Wang (2001) introduzem o modelo misto não linear semiparamétrico.

Modelo linear parcial

Quando $s = 1$ e $\mathbf{b}_i = \mathbf{0}$ ($i = 1, \dots, n$), o modelo (2.1) reduz-se ao modelo linear parcial para dados longitudinais. Nesse caso, o modelo assume a forma

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + f(t_{ij}) + \epsilon_{ij}, \quad (2.5)$$

($j = 1, \dots, m_i$) em que os erros aleatórios seguem uma distribuição normal. Note que para $m_i = 1$ ($i = 1, \dots, n$), o modelo reduz-se ao modelo linear parcial discutido por Heckman (1986), Speckman (1988) e Green e Silverman (1994).

Modelo aditivo

Quando $\boldsymbol{\beta} = \mathbf{0}$ e $\mathbf{b}_i = \mathbf{0}$ ($i = 1, \dots, n$), o modelo (2.1) reduz-se ao modelo aditivo discutido, por exemplo, em Buja et al. (1989). Nesse caso, o modelo é dado por

$$y_{ij} = \sum_{k=1}^s f_k(t_{kij}) + \epsilon_{ij}, \quad (2.6)$$

($j = 1, \dots, m_i$) em que ϵ_{ij} são erros aleatórios com média 0 e variância σ^2 . Veja outros detalhes referentes aos modelos aditivos em Hastie e Tibshirani (1990).

2.3 Representação matricial

Seja $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^T$ o vetor aleatório de respostas observadas ($m_i \times 1$) para o i -ésimo grupo ($i = 1, \dots, n$; $j = 1, \dots, m_i$). Então, o modelo (2.1) pode ser expresso em forma matricial da seguinte maneira:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{N}_{1i} \mathbf{f}_1 + \mathbf{N}_{2i} \mathbf{f}_2 + \dots + \mathbf{N}_{si} \mathbf{f}_s + \boldsymbol{\epsilon}_i \\ &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k + \boldsymbol{\epsilon}_i, \end{aligned} \quad (2.7)$$

em que \mathbf{X}_i é uma matriz de planejamento ($m_i \times p$), com linhas \mathbf{x}_{ij}^T ; \mathbf{Z}_i é uma matriz de planejamento ($m_i \times q$) associada aos efeitos aleatórios, com linhas \mathbf{z}_{ij}^T ; \mathbf{f}_k é um vetor de parâmetros ($r_k \times 1$) (efeitos fixos não paramétricos) definido na forma

$$\mathbf{f}_k = \begin{pmatrix} f_k(t_{k_1}^0) \\ \vdots \\ f_k(t_{k_{r_k}}^0) \end{pmatrix} = \begin{pmatrix} \zeta_{k_1} \\ \vdots \\ \zeta_{k_{r_k}} \end{pmatrix},$$

em que $t_{k_g}^0$ ($k = 1, \dots, s$; $g = 1, \dots, r_k$) representam os valores ordenados e distintos da variável explicativa $t_{k_{ij}}$ ($i = 1, \dots, n$; $j = 1, \dots, m_i$); \mathbf{N}_{ki} é a matriz de incidência ($m_i \times r_k$) associada ao i -ésimo grupo e k -ésima variável explicativa, cujo elemento (j, g)-ésimo é igual à função indicadora $I(t_{k_{ij}} = t_{k_g}^0)$ ($j = 1, \dots, m_i$; $g = 1, \dots, r_k$); e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im_i})^T$ é um vetor de erros aleatórios ($m_i \times 1$) não observáveis. Note que a conexão entre os vetores \mathbf{t}_{k_i} e $\mathbf{t}_{k_i}^0 = (t_{k_1}^0, \dots, t_{k_{r_k}}^0)^T$ recupera-se através da matriz de incidência \mathbf{N}_{ki} ($k = 1, \dots, s$), a qual pode ser construída da mesma forma como é construída a matriz de incidência apresentada por Green e Silverman (1994) na Seção 4.3.1. Veja um exemplo no Apêndice A.

2.4 Modelo normal

No contexto dos modelos mistos clássicos tipicamente assume-se uma formulação hierárquica em que os erros aleatórios e os efeitos aleatórios são independentes e cada um tem distribuição normal. Analogamente, podemos introduzir um enfoque hierárquico para o modelo misto aditivo semiparamétrico (2.7) na forma

$$\begin{cases} \mathbf{y}_i \mid \mathbf{b}_i \sim \mathcal{N}_{m_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k, \mathbf{V}_i), \\ \mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{D}) \quad \text{e} \\ \boldsymbol{\epsilon}_i \sim \mathcal{N}_{m_i}(\mathbf{0}, \mathbf{V}_i), \end{cases} \quad (2.8)$$

em que \mathbf{b}_i e $\boldsymbol{\epsilon}_i$ são independentes. Consequentemente, temos que a distribuição conjunta do vetor aleatório $(\mathbf{y}_i^T, \mathbf{b}_i^T, \boldsymbol{\epsilon}_i^T)^T$ assume a forma

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \sim \mathcal{N}_{m_i^*} \left\{ \begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i & \mathbf{Z}_i \mathbf{D} & \mathbf{V}_i \\ \mathbf{D} \mathbf{Z}_i^T & \mathbf{D} & \mathbf{0} \\ \mathbf{V}_i & \mathbf{0} & \mathbf{V}_i \end{pmatrix} \right\},$$

em que $m_i^* = m_i + q + m_i$. Assim, a inferência clássica é, usualmente, baseada na função de verossimilhança penalizada (isto será discutido em detalhe no capítulo seguinte no contexto das distribuições de contornos elípticos) do modelo marginal

$$\mathbf{y}_i \sim \mathcal{N}_{m_i} \left(\mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k, \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i \right). \quad (2.9)$$

Sob o enfoque hierárquico as matrizes \mathbf{D} ($q \times q$) e \mathbf{V}_i ($m_i \times m_i$) devem ser matrizes quadradas positivas definidas, uma vez que representam às matrizes de variância-covariância dos vetores aleatórios \mathbf{b}_i e $\boldsymbol{\epsilon}_i$, respectivamente. Por outro lado, no modelo marginal (2.9) apenas a matriz $\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i$ deve ser positiva definida e portanto podemos considerar situações em que os componentes de variância da matriz \mathbf{D} sejam negativos. Em geral, a decisão sobre qual modelo adotar depende diretamente da interpretação do problema em consideração. Contudo, a distribuição normal, sendo a mais utilizada na modelagem de muitos fenômenos, tem sido criticada por fornecer estimativas de máxima verossimilhança sensíveis a observações aberrantes. A fim de acomodar tais observações, que eventualmente podem ser influentes nas conclusões finais de um estudo particular, diversos autores têm sugerido o uso das **distribuições de contornos elípticos**. Essas distribuições, além de permitir estender os modelos já desenvolvidos sob erros normais, permitem acomodar as observações aberrantes através de distribuições com caudas mais leves ou mais pesadas do que as caudas da distribuição normal.

2.5 Modelo elíptico

A questão principal, agora, é como definir um modelo misto aditivo semiparamétrico na classe das distribuições elípticas. Em geral há na literatura basicamente duas abordagens que têm sido consideradas para o caso paramétrico. Por um lado, podemos usar uma representação hierárquica elíptica da mesma forma que no modelo normal. Nesse caso, o modelo (2.7), definido desde o ponto de vista hierárquico, supõe que

$$\begin{cases} \mathbf{y}_i \mid \mathbf{b}_i \sim \mathcal{E}l_{m_i} \left(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k, \mathbf{V}_i \right), \\ \mathbf{b}_i \sim \mathcal{E}l_q(\mathbf{0}, \mathbf{D}) \quad \text{e} \\ \boldsymbol{\epsilon}_i \sim \mathcal{E}l_{m_i}(\mathbf{0}, \mathbf{V}_i), \end{cases} \quad (2.10)$$

em que \mathbf{D} ($q \times q$) e \mathbf{V}_i ($m_i \times m_i$) são matrizes de escala positivas definidas, respectivamente. Porém, a representação (2.10) tem a desvantagem, em relação à representação hierárquica no caso normal, que a distribuição conjunta do vetor aleatório $(\mathbf{y}_i^T, \mathbf{b}_i^T, \boldsymbol{\epsilon}_i^T)^T$ não necessariamente pertence à classe das distribuições elípticas, dificultando, dessa forma, a obtenção da distribuição marginal de \mathbf{y}_i . Para os modelos mistos aditivos semiparamétricos com erros assumindo distribuição \mathbf{t} -Student multivariada, podemos considerar a representação hierárquica em dois estágios proposta por Pinheiro et al. (2001). Neste caso, é feita uma composição da distribuição normal e gama para obter uma distribuição \mathbf{t} -Student multivariada; especificamente,

$$\begin{aligned} \mathbf{y}_i \mid (\mathbf{b}_i, v_i) &\sim \mathcal{N}_{m_i} \left(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k, \frac{1}{v_i} \mathbf{V}_i \right), \\ \mathbf{b}_i \mid v_i &\sim \mathcal{N}_q \left(\mathbf{0}, \frac{1}{v_i} \mathbf{D} \right) \quad \text{e} \\ v_i &\sim \text{Gama} \left(\frac{\nu_i}{2}, \frac{\nu_i}{2} \right), \end{aligned}$$

em que $\text{Gama}(\zeta_1, \zeta_2)$ denota a distribuição gama com parâmetros ζ_1 e ζ_2 . Essa representação hierárquica implica que

$$\mathbf{y}_i \sim \mathbf{t}_{m_i} \left(\mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k, \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i, \nu_i \right),$$

em que ν_i ($i = 1, \dots, n$) representa o número de graus de liberdade da distribuição \mathbf{t} -Student e Inserido nesse contexto, podemos estimar os efeitos fixos e aleatórios adaptando o processo de estimação (baseado no algoritmo EM) proposto por Pinheiro et al. (2001). Note que no caso da distribuição \mathbf{t} -Student, a repre-

sentação hierárquica descrita acima não apresenta grandes dificuldades analíticas, enquanto para outras distribuições pertencentes à classe das distribuições de contornos elípticos, a obtenção da distribuição marginal de \mathbf{y}_i não é simples e pode ser necessário utilizar procedimentos de integração numérica. Uma forma possível de fazer frente a essa dificuldade é assumir, inicialmente, que

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \sim \mathcal{E}l_{m_i^*} \left\{ \begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i & \mathbf{Z}_i \mathbf{D} & \mathbf{V}_i \\ \mathbf{D} \mathbf{Z}_i^T & \mathbf{D} & \mathbf{0} \\ \mathbf{V}_i & \mathbf{0} & \mathbf{V}_i \end{pmatrix} \right\}. \quad (2.11)$$

em que $m_i^* = m_i + q + m_i$. A representação (2.11) é bastante conveniente desde o ponto de vista analítico, porém possui a desvantagem de que os efeitos aleatórios não têm a mesma interpretação dos efeitos fixos como no caso hierárquico. Note que neste caso os vetores aleatórios \mathbf{b}_i e $\boldsymbol{\epsilon}_i$ são não correlacionados, mas não necessariamente independentes, com exceção do caso normal, e a distribuição marginal de \mathbf{y}_i pode ser obtida sem precisar de métodos de integração numérica. Com efeito, usando a Propriedade 1.4.3 apresentada no início deste trabalho, podemos mostrar que os vetores \mathbf{y}_i ($i = 1, \dots, n$) seguem uma distribuição marginal na forma (veja a prova deste resultado no Apêndice B)

$$\mathbf{y}_i \sim \mathcal{E}l_{m_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (2.12)$$

cujas funções de densidade associadas assumem a forma

$$f_{\mathbf{y}}(\mathbf{y}_i) = |\boldsymbol{\Sigma}_i|^{-1/2} g(\delta_i), \quad (2.13)$$

em que $g(\cdot)$ é a função geradora de densidades que satisfaz a condição (1.8) e $\delta_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$ é a distância de Mahalanobis, com $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k$ sendo o vetor de posição ($m_i \times 1$) e $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i$ a matriz de escala ($m_i \times m_i$) positiva definida. Vamos, a partir de agora, denominar o modelo definido por (2.7) e (2.12) modelo misto aditivo semiparamétrico de contornos elípticos ou simplesmente modelo misto aditivo semiparamétrico elíptico (MMASE).

O problema de identificabilidade do modelo MMASE, usualmente gerado pela superdispersão das funções não paramétricas f_k ($k = 1, \dots, s$) conjuntamente com os demais parâmetros, é um aspecto que deve ser considerado com muita atenção. Na prática, podemos obter a identificabilidade do modelo (a) impondo condições sobre as funções f_k ($k = 1, \dots, s$), (b) incorporando condições à distribuição dos efeitos aleatórios, ou pela combinação de (a) e (b). Uma interessante discussão referente ao problema de identificabilidade no contexto dos modelos mistos não lineares semiparamétricos é apresentada por Ke e Wang (2001).

2.6 Conclusões do capítulo

Neste capítulo foi apresentado o modelo misto aditivo semiparamétrico na classe das distribuições de contornos elípticos. Assumindo que o vetor de respostas, o vetor dos efeitos aleatórios e o vetor dos erros aleatórios de cada grupo segue uma distribuição conjunta elíptica, derivamos a distribuição marginal do vetor de respostas e sua correspondente função densidade. Nesse contexto, destacamos o fato de que o modelo marginal (2.12) preserva a média do modelo hierárquico (2.11).

Capítulo 3

Estimação de máxima verossimilhança penalizada

Neste capítulo abordamos o problema de estimação dos parâmetros envolvidos no modelo misto aditivo semiparamétrico elíptico. A Seção 1 apresenta uma breve discussão dos principais trabalhos relacionados à estimação e inferência em modelos semiparamétricos. A Seção 2 define o critério que será usado para estimar os parâmetros do modelo; especificamente, discute-se o método da função penalizada e define-se o logaritmo da função de verossimilhança penalizada. As Seções 3 e 4 contêm a função escore e a matriz de informação de Fisher, respectivamente. A Seção 5 dá início à apresentação do procedimento de estimação dos parâmetros. A Seção 6 apresenta uma interessante reparametrização do modelo misto aditivo semiparamétrico. As Seções 7 e 8 apresentam, respectivamente, uma discussão sobre os principais procedimentos de estimação dos parâmetros de suavização e de seleção de modelos. Por fim, discutimos em linhas gerais na Seção 9 os principais resultados do capítulo.

3.1 Introdução

O problema de estimação no contexto dos modelos mistos aditivos semiparamétricos elípticos ainda não foi discutido na literatura. No entanto, vários autores já consideraram o problema de estimação para alguns casos particulares. Por

exemplo, no contexto dos modelos lineares parciais (MLP), Heckman (1986) estima o coeficiente de regressão e a função não paramétrica baseado no critério da verossimilhança penalizada, e mostra que o estimador da função não paramétrica é uma *spline* cúbica natural; veja também Engle et al. (1986) e Rice (1986). Green (1987) estima os efeitos paramétricos e os efeitos não paramétricos baseados no critério da verossimilhança penalizada, e sugere resolver as equações de estimação através do processo iterativo de Newton-Raphson ou score de Fisher. Além disso, Green estabelece algumas condições sobre o modelo para obter a convergência do processo. Speckman (1988) apresenta um método para estimar o coeficiente de regressão e a função não paramétrica baseado em um procedimento de suavização de *kernel*. Robinson (1988) estuda o problema de estimação do coeficiente de regressão e observa que esse estimador é inconsistente quando a função não paramétrica não é parametrizada corretamente. Além disso, Robinson propõe um estimador de mínimos quadrados \sqrt{n} -consistente para o coeficiente de regressão. Opsomer e Ruppert (1999) propõem um estimador *backfitting* \sqrt{n} -consistente para o coeficiente de regressão do modelo MLP.

Zeger e Diggle (1994) propõem um estimador para a função não paramétrica baseado no procedimento de suavização de *kernel*. Por sua parte, Zhang et al. (1998) derivam o estimador do coeficiente de regressão e da função não paramétrica a partir da verossimilhança penalizada, e mostram que o estimador da função não paramétrica conduz a uma *spline* cúbica natural. Além disso, eles estimam os componentes de variância e o parâmetro de suavização simultaneamente usando a verossimilhança restrita. Entretanto, os efeitos aleatórios são estimados através de estimadores empíricos de Bayes. Lin e Zhang (1999) estimam as funções não paramétricas a partir do logaritmo da quase-verossimilhança penalizada e mostram que tais estimadores formam uma *spline* cúbica natural. Além disso, eles estimam os componentes de variância e o parâmetro de suavização simultaneamente baseados na quase-verossimilhança marginal. Fahrmeir e Lang (2001) estudam o modelo misto aditivo generalizado desde o ponto de vista bayesiano e usam o procedimento MCMC para gerar amostras das distribuições a posteriori e

assim estimar a média a posteriori, as medianas, e as bandas de confiança para os quartis. Hastie e Tibshirani (1986) usam o algoritmo *escore local* para ajustar o componente aditivo de um modelo aditivo generalizado e ilustram tal procedimento usando dados de respostas binárias e dados de sobrevivência; veja também Hastie e Tibshirani (1987). Hastie e Tibshirani (1993) estudam os modelos aditivos generalizados em que os coeficientes de regressão variam suavemente de acordo com o valor de outras covariáveis, e mostram, baseados no critério de mínimos quadrados penalizados, que os estimadores das funções não paramétricas formam uma *spline* cúbica natural. Berhane e Tibshirani (1998) usam o processo iterativo de Newton Raphson para maximizar a quase-verossimilhança penalizada e o algoritmo *backfitting* para estimar as funções não paramétricas em um modelo aditivo. Recentemente, Rigby e Stasinopoulos (2005) usam o processo de Newton-Raphson para ajustar um modelo aditivo generalizado para posição, escala e forma, baseados no critério da verossimilhança penalizada.

3.2 Critério da verossimilhança penalizada

Por simplicidade, neste trabalho, vamos assumir que as matrizes de escala \mathbf{D} e \mathbf{V}_i independem dos parâmetros $\boldsymbol{\beta}$ e \mathbf{f}_k ($k = 1, \dots, s$), e que dependem apenas de um número finito de parâmetros (desconhecidos); isto é, vamos assumir que $\mathbf{D} = \mathbf{D}(\boldsymbol{\lambda})$ e $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\gamma})$ são matrizes parametrizadas pelos vetores $\boldsymbol{\lambda} \in \mathbb{R}^{d_\lambda}$ e $\boldsymbol{\gamma} \in \mathbb{R}^{d_\gamma}$, respectivamente (veja em Hand e Crowder, 1996 e Banerjee e Frees, 1997 algumas estruturas para a matriz \mathbf{V}_i). Dessa forma, o vetor de parâmetros a ser estimado no modelo misto aditivo semiparamétrico elíptico será $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}_1^T, \dots, \mathbf{f}_s^T, \boldsymbol{\lambda}^T, \boldsymbol{\gamma}^T)^T$ cujo espaço paramétrico associado é dado por

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{p^*} \mid \boldsymbol{\beta} \in \Theta_\beta, \mathbf{f}_1 \in \Theta_{f_1}, \dots, \mathbf{f}_s \in \Theta_{f_s}, \boldsymbol{\lambda} \in \Theta_\lambda, \boldsymbol{\gamma} \in \Theta_\gamma \right\},$$

em que $\Theta_\beta \subseteq \mathbb{R}^p$, $\Theta_{f_k} \subseteq \mathbb{R}^{r_k}$ ($k = 1, \dots, s$), $\Theta_\lambda \subseteq \mathbb{R}^{d_\lambda}$ e $\Theta_\gamma \subseteq \mathbb{R}^{d_\gamma}$ denotam, respectivamente, os espaços paramétricos associados com os efeitos fixos paramétricos, os efeitos fixos não paramétricos, e os componentes de efeitos aleatórios;

$p^* = p + r + d^*$, com $r = \sum_{k=1}^s r_k$ e $d^* = d_\lambda + d_\gamma$. Logo, o logaritmo da função de verossimilhança pode ser expresso na seguinte forma:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n L_i(\boldsymbol{\theta}), \quad (3.1)$$

em que

$$L_i(\boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log g(\delta_i). \quad (3.2)$$

Em geral, os procedimentos tipicamente utilizados para estimar o vetor de parâmetros finito dimensional em um modelo paramétrico, como por exemplo máxima verossimilhança, fornecem estimadores com boas propriedades. Contudo, a aplicação desse procedimento não tem gerado resultados análogos no contexto semiparamétrico. Porém, diversos autores têm proposto algumas variantes desse procedimento, tentando obter a identificabilidade dos parâmetros e boas propriedades dos estimadores. Estritamente falando, a maximização direta de (3.1) sem impor restrições sobre as funções f_k 's pode gerar um super ajuste e fazer com que $\boldsymbol{\beta}$ seja não identificável; vide, por exemplo, Green (1987).

Um procedimento alternativo baseado no critério da verossimilhança penalizada consiste em incorporar uma função de penalização no logaritmo da função de verossimilhança $L(\boldsymbol{\theta})$, tal que

$$L_p(\boldsymbol{\theta}, \alpha_1, \dots, \alpha_s) = L(\boldsymbol{\theta}) + \sum_{k=1}^s \alpha_k^* J(f_k), \quad (3.3)$$

em que $J(f_k)$ é uma função de penalidade (medida de variação local) imposta sobre a função não paramétrica, f_k , que depende de alguma aplicação específica ou conhecimento a priori, e $\alpha_k^* = \alpha^*(\alpha_k)$ é um termo que depende do parâmetro $\alpha_k \geq 0$ ($k = 1, \dots, s$). Os parâmetros α_k , conhecidos por parâmetros de suavização, regulam a relação entre a fidelidade dos dados, quantificada por valores grandes de $L(\boldsymbol{\theta})$, e a suavidade ou ondulação das curvas estimadas, quantificada por valores pequenos dos termos de penalidade $J_k = \alpha_k^* J(f_k)$. Observamos que o termo de penalização J_k age diretamente sobre a suavidade da função estimada e penaliza

funções com curvatura muito acentuada em diversos pontos, suavizando, deste modo, a estimativa final da função alvo f_k .

3.2.1 Função de penalidade

No contexto semiparamétrico tem-se proposto diferentes funções de penalidade; veja, por exemplo, Good e Gaskins (1971) e Shen (1997). Neste trabalho assumiremos que o termo de penalidade é definido por

$$J_k = \alpha_k^* \int_{a_k}^{b_k} [f_k^{(l)}(t_k)]^2 dt_k, \quad (3.4)$$

em que $f_k^{(l)}(t_k) = \frac{d^l}{dt_k^l} f(t_k)$, $t_k \in [a_k, b_k]$, e a função f_k pertence ao espaço de funções Sobolev definido por

$$\mathcal{W}_2^{(l)}[a_k, b_k] = \left\{ f_k : f_k^{(l)} \in L^2[a_k, b_k], f_k^{(1)}, f_k^{(2)}, \dots, f_k^{(l-1)} \text{ abs. cont.} \right\},$$

em que

$$L^2[a_k, b_k] = \left\{ f : \int_{a_k}^{b_k} f^2(t) dt < \infty \right\}.$$

Em particular, para $l = 2$, o estimador da função não paramétrica f_k que maximiza (3.3) com J_k dado por (3.4), corresponde a uma *spline* cúbica natural com nós ou *knots* (pontos de descontinuidade na terceira derivada) nos pontos t_{kg}^0 ($g = 1, \dots, r_k$). De acordo com Reinsch (1967) e Silverman (1985), é possível mostrar que a curva estimada \hat{f}_k tem as seguintes propriedades:

- (1) \hat{f}_k é um polinômio cúbico em cada intervalo $[t_{k_j}, t_{k_{j+1}}]$;
- (2) para cada ponto t_{k_j} , a curva e suas duas primeiras derivadas são contínuas, embora possa ser descontínua em sua terceira derivada; e,
- (3) em cada intervalo $(-\infty, t_{k_1}]$ e $[t_{k_{r_k}}, \infty)$ a segunda derivada é zero, de modo que \hat{f}_k é linear fora do domínio dos dados.

Em geral, qualquer curva que satisfaz (1) e (2) é chamada *spline* cúbica. Note também que as propriedades descritas acima não são impostas sobre o estimador da função f_k , pois elas emergem automaticamente devido à escolha da função de penalidade (3.4) para $l = 2$. Nesse caso, Green e Silverman (1994) mostram que o termo de penalidade (3.4) pode ser expresso através da forma quadrática

$$\int_{a_k}^{b_k} [f_k^{(2)}(t_k)]^2 dt_k = \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k, \quad (3.5)$$

em que $f_k^{(2)}(t_k) = \frac{d^2}{dt_k^2} f(t_k)$ e \mathbf{K}_k é uma matriz de suavização ($r_k \times r_k$) positiva definida para a k -ésima variável explicativa, que depende apenas dos nós. Note que o termo de penalidade (3.5) mede a ondulação da função f_k . Em particular, para $k = 1$, a matriz de suavização, digamos \mathbf{K} , tem estrutura dada por $\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T$, em que \mathbf{Q} é uma matriz $[n \times (n-2)]$ com entradas q_{ij} ($i = 1, \dots, n$ e $j = 1, \dots, n-1$) dadas por

$$\begin{aligned} q_{j-1,j} &= h_{j-1}^{-1}, \\ q_{j,j} &= -h_{j-1}^{-1} - h_j^{-1} \quad \text{e} \\ q_{j+1,j} &= h_j^{-1}, \end{aligned}$$

com $q_{ij} = 0$ para $|i-j| \geq 0$ ($j, \dots, n-1$). \mathbf{R} é uma matriz simétrica $[(n-2) \times (n-2)]$ com elementos $r_{i,j}$, para $i, j \in [2, n-1]$, dados por

$$\begin{aligned} r_{i,i} &= \frac{1}{3}(h_{i-1} + h_i) \quad i = 1, \dots, n-1, \\ r_{i,i+1} &= r_{i+1,i} = \frac{1}{6}h_i \quad i = 1, \dots, n-2 \quad \text{e} \end{aligned}$$

$r_{i,j} = 0$ para $|i-j| \geq 2$, em que $h_i = t_{i+1} - t_i$ ($i = 1, \dots, n-1$). No caso em que t_g^0 ($g = 1, \dots, r$) correspondam aos valores ordenados e distintos de t_i , a matriz de suavização é construída substituindo t_1, \dots, t_n por t_1^0, \dots, t_r^0 . Uma descrição do procedimento para construir essas matrizes é apresentado no Apêndice A. Maiores detalhes da teoria *spline* aplicada à Estatística em Wegman e Wright (1983).

3.2.2 Logaritmo da função de verossimilhança penalizada

Seja a função $L(\boldsymbol{\theta})$ definida pela equação (3.1), a função de penalidade J_k da forma (3.4) com $\alpha_k^* = -\alpha_k/2$, e a forma quadrática definida pela equação (3.5). Então, o logaritmo da função de verossimilhança penalizada (log-verossimilhança penalizada) associado ao modelo misto aditivo semiparamétrico elíptico pode ser expresso na forma

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \quad (3.6)$$

em que

$$L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = L_i(\boldsymbol{\theta}) - \frac{1}{2n} \sum_{k=1}^s \alpha_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k, \quad (3.7)$$

sendo (a_k, b_k) o intervalo que define o domínio da k -ésima variável explicativa e $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_s)^T$ o vetor de parâmetros de suavização. Note que se α_k for grande, maior importância estaremos dando para o critério de suavização, e portanto as curvas serão mais suaves ou menos onduladas. Agora, se α_k for pequeno (próximo de zero), estaremos dando mais importância para a medida da qualidade do ajuste.

Em geral, a função de verossimilhança penalizada pode ser estudada com propósitos de otimização estabelecendo condições analíticas gerais, tais como continuidade, convexidade, e diferenciabilidade, sobre a função de verossimilhança e a função de penalidade. Nessa direção, alguns autores têm mostrado que existe uma relação de equivalência entre o processo de otimização da função de verossimilhança penalizada e o processo de otimização da função de verossimilhança sujeita a alguma condição específica imposta sobre um termo de penalidade; vide, por exemplo, Schoenberg (1964) no contexto de *spline* polinomial natural.

Na seguinte seção derivamos a função escore penalizada associada ao vetor de parâmetros $\boldsymbol{\theta}$. Os cálculos envolvidos são apresentados no Apêndice C.

3.3 Função escore penalizada

Seja a função escore penalizada definida por

$$\mathbf{U}_p(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}}, \quad (3.8)$$

em que $\partial L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})/\partial \boldsymbol{\theta}$ denota a primeira derivada parcial do logaritmo da função de verossimilhança penalizada em relação ao vetor $\boldsymbol{\theta}$; especificamente, em relação aos vetores de parâmetros $\boldsymbol{\beta}, \mathbf{f}_1, \dots, \mathbf{f}_s$, e $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_{d^*})^T$, com $\tau_\ell = \lambda_\ell$ ($\ell = 1, \dots, d_\lambda$) e $\tau_\ell = \gamma_j$ ($\ell = d_\lambda + 1, \dots, d^*$; $j = 1, \dots, d_\gamma$). Com efeito, assumindo que a função geradora de densidades $g(\cdot)$, definida na Seção 2.5, é contínua e diferenciável, definimos as quantidades

$$v_i(\boldsymbol{\theta}) = -2W_g(\delta_i),$$

em que

$$W_g(\delta_i) = \frac{d}{d\delta_i} \log g(\delta_i) = \frac{g'(\delta_i)}{g(\delta_i)}. \quad (3.9)$$

Usando resultados de diferenciação de matrizes (vide Magnus e Neudecker, 1988) sobre a equação (3.6) em relação aos elementos de $\boldsymbol{\theta}$, temos que a função escore penalizada pode ser expressa na forma

$$\mathbf{U}_p(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{U}_{P_i}(\boldsymbol{\theta}),$$

em que

$$\mathbf{U}_{P_i}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_{P_i}^\beta(\boldsymbol{\theta}) \\ \mathbf{U}_{P_i}^{\mathbf{f}_1}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{U}_{P_i}^{\mathbf{f}_s}(\boldsymbol{\theta}) \\ \mathbf{U}_{P_i}^\tau(\boldsymbol{\theta}) \end{pmatrix}, \quad (3.10)$$

sendo as funções escore parciais associadas aos efeitos fixos do modelo dadas por

$$\begin{aligned} \mathbf{U}_{\mathbf{P}_i}^\beta(\boldsymbol{\theta}) &= \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \\ &= v_i(\boldsymbol{\theta}) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i), \end{aligned} \quad (3.11)$$

$$\begin{aligned} \mathbf{U}_{\mathbf{P}_i}^{\mathbf{f}_k}(\boldsymbol{\theta}) &= \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k} \\ &= v_i(\boldsymbol{\theta}) \mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) - \frac{\alpha_k}{n} \mathbf{K}_k \mathbf{f}_k, \end{aligned} \quad (3.12)$$

($k = 1, \dots, s$) e as funções escore parciais associadas aos componentes de variância dadas por

$$\mathbf{U}_{\mathbf{P}_i}^\tau(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_{\mathbf{P}_i}^{\tau_1}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{U}_{\mathbf{P}_i}^{\tau_{d^*}}(\boldsymbol{\theta}) \end{pmatrix}, \quad (3.13)$$

em que

$$\begin{aligned} \mathbf{U}_{\mathbf{P}_i}^{\tau_\ell}(\boldsymbol{\theta}) &= \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \tau_\ell} \\ &= -\frac{1}{2} \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \} - v_i(\boldsymbol{\theta}) (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right], \end{aligned} \quad (3.14)$$

com $\dot{\boldsymbol{\Sigma}}_i(\ell) = \partial \boldsymbol{\Sigma}_i / \partial \tau_\ell$ ($i = 1, \dots, n$).

As quantidades $v_i(\boldsymbol{\theta})$ que aparecem nas equações acima podem ser interpretadas como pesos e como $g(\delta_i)$ é uma função positiva e decrescente para quase todas as distribuições que pertencem à classe das distribuições elípticas, temos que $v_i(\boldsymbol{\theta}) > 0$, com exceção para as distribuições Kotz, Kotz generalizada, e dupla exponencial. A Tabela 3.1 mostra as expressões de $v_i(\boldsymbol{\theta})$ para algumas distribuições elípticas e podemos observar que no caso da distribuição t-Student, a quantidade $v_i(\boldsymbol{\theta})$ é inversamente proporcional à distância de Mahalanobis $\delta_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$. Dessa forma, o procedimento de estimação tende a atribuir pesos pequenos para as observações aberrantes. No caso da distribuição exponencial potência, o parâmetro ζ é uma medida de curtose; para $-1 < \zeta < 0$ a

distribuição tem caudas mais leves do que as da normal e para $0 < \zeta < 1$ a distribuição tem caudas mais pesadas. Quando $\zeta = 0$ recaímos na distribuição normal, e portanto esse parâmetro pode ser visto como um parâmetro de afastamento da normalidade. Assim, com o objetivo de acomodar observações aberrantes, pode-se usar $0 < \zeta < 1$. Note que tais distribuições conduzem a um processo de estimação menos sensível a observações aberrantes em relação ao modelo normal em que $v_i(\boldsymbol{\theta}) = 1$. Esse processo de estimação será robusto, no sentido da distância de Mahalanobis, quando trata-se de estimar o coeficiente de regressão, e parcialmente robusto quando trata-se de estimar o componente de variância e o componente não paramétrico. Nesse último caso, a parcialidade deve-se ao termo de penalidade imposto sobre a função de verossimilhança que não é ponderado pelos pesos e que apenas depende de $\mathbf{t}_k^0 = (t_{k_1}^0, \dots, t_{k_{r_k}}^0)^T$ ($k = 1, \dots, s$).

Tabela 3.1: Expressões das quantidades v_i para algumas distribuições elípticas.

Distribuição	$v_i(\boldsymbol{\theta}) = -2W_g(\delta_i)$
Normal	1
t-Student	$\frac{\nu_i + m_i}{\nu_i + \delta_i}$
Exponencial Potência	$\frac{1}{1+\zeta} \delta_i^{\frac{1}{1+\zeta}-1}$
Logística I	$2 \tanh\left(\frac{\delta_i}{2}\right)$
Logística II	$\delta_i^{1/2} \tanh\left(\frac{\delta_i^{1/2}}{2}\right)$

Na seguinte seção derivamos a matriz de informação de Fisher penalizada associada ao vetor de parâmetros $\boldsymbol{\theta}$. Essa matriz será utilizada na construção do procedimento de estimação de $\boldsymbol{\theta}$ e na obtenção da matriz de variância-covariância de $\hat{\boldsymbol{\theta}}$. A prova deste resultado e os cálculos algébricos relacionados são apresentados no Apêndice C. Outros detalhes referentes a estes resultados no caso dos modelos mistos lineares elípticos, veja Savalli (2006) e Lange et al. (1989).

3.4 Matriz de informação de Fisher penalizada

Definamos as quantidades $d_{g_i} = E(W_g^2(U_i)U_i)$ e $f_{g_i} = E(W_g^2(U_i)U_i^2)$, com $U_i = \|\mathbf{Z}_i\|^2$ e $\mathbf{Z}_i \sim \mathcal{S}_{m_i}(g)$. A matriz de informação de Fisher penalizada é definida por

$$\mathcal{I}_p(\boldsymbol{\theta}) = -E\left\{\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right\}, \quad (3.15)$$

em que

$$\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$$

denota as segundas derivadas parciais do logaritmo da função de verossimilhança penalizada em relação ao vetor $\boldsymbol{\theta}$, especificamente, em relação aos vetores de parâmetros $\boldsymbol{\beta}$, $\mathbf{f}_1, \dots, \mathbf{f}_s$, e $\boldsymbol{\tau}$. Com efeito, é possível mostrar que a matriz de informação de Fisher penalizada (3.15) para o modelo misto aditivo semiparamétrico elíptico assume a forma bloco diagonal

$$\mathcal{I}_p(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix}, \quad (3.16)$$

em que

$$\begin{aligned} \mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) &= \sum_{i=1}^n \mathcal{I}_{p_i}^{\beta\beta}(\boldsymbol{\theta}) \quad \text{e} \\ \mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta}) &= \sum_{i=1}^n \mathcal{I}_{p_i}^{\tau\tau}(\boldsymbol{\theta}) \end{aligned}$$

são matrizes de ordens $[(p+r) \times (p+r)]$ e $(d^* \times d^*)$, respectivamente, sendo

$$\mathcal{I}_{p_i}^{\beta\beta}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{p_i}^{\beta\beta}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{\beta f_1}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{p_i}^{\beta f_s}(\boldsymbol{\theta}) \\ \mathcal{I}_{p_i}^{f_1 \beta}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{f_1 f_1}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{p_i}^{f_1 f_s}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{p_i}^{f_s \beta}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{f_s f_1}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{p_i}^{f_s f_s}(\boldsymbol{\theta}) \end{pmatrix} \quad (3.17)$$

uma matriz $[(p+r) \times (p+r)]$, com

$$\begin{aligned} \mathcal{I}_{\mathbf{p}_i}^{\beta\beta}(\boldsymbol{\theta}) &= -\mathbb{E}\left\{\frac{\partial^2 L_{\mathbf{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\} \\ &= \frac{4d_{g_i}}{m_i}\mathbf{X}_i^T\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i, \end{aligned} \quad (3.18)$$

$$\begin{aligned} \mathcal{I}_{\mathbf{p}_i}^{\beta f_k}(\boldsymbol{\theta}) &= -\mathbb{E}\left\{\frac{\partial^2 L_{\mathbf{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\mathbf{f}_k^T}\right\} \\ &= \frac{4d_{g_i}}{m_i}\mathbf{X}_i^T\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_{ki}, \end{aligned} \quad (3.19)$$

$$\begin{aligned} \mathcal{I}_{\mathbf{p}_i}^{f_k f_{k'}}(\boldsymbol{\theta}) &= -\mathbb{E}\left\{\frac{\partial^2 L_{\mathbf{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\mathbf{f}_k\partial\mathbf{f}_{k'}^T}\right\} \\ &= \begin{cases} \frac{4d_{g_i}}{m_i}\mathbf{N}_{ki}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_{ki} + \frac{\alpha_k}{n}\mathbf{K}_k & k = k' \\ \frac{4d_{g_i}}{m_i}\mathbf{N}_{ki}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_{k'i} & k \neq k', \quad k, k' = 1, \dots, s, \end{cases} \end{aligned} \quad (3.20)$$

e $\mathcal{I}_{\mathbf{p}}^{\tau\tau}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{I}_{\mathbf{p}_i}^{\tau\tau}(\boldsymbol{\theta})$ uma matriz $(d^* \times d^*)$ em que o (j^*, ℓ^*) -ésimo elemento da matriz $\mathcal{I}_{\mathbf{p}_i}^{\tau\tau}(\boldsymbol{\theta})$ pode ser expresso na forma

$$\begin{aligned} \mathcal{I}_{\mathbf{p}_{i_j^* \ell^*}}^{\tau\tau} &= -\mathbb{E}\left\{\frac{\partial^2 L_{\mathbf{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\tau_{j^*}\partial\tau_{\ell^*}}\right\} \\ &= \frac{b_{i_j^* \ell^*}}{4}\left(\frac{4f_{g_i}}{m_i(m_i+2)} - 1\right) + \\ &\quad \frac{2f_{g_i}}{m_i(m_i+2)}\text{tr}\left\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(j^*)\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell^*)\right\}, \end{aligned} \quad (3.21)$$

em que $b_{i_j^* \ell^*} = \text{tr}\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(j^*)\}\text{tr}\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell^*)\}$ ($j^*, \ell^* = 1, \dots, d^*$). Para algumas distribuições multivariadas pertencentes à classe das distribuições elípticas as quantidades d_{g_i} e f_{g_i} da expressão (3.21) têm uma forma fechada, como é o caso da

distribuição normal, t-Student e exponencial potência. Para outras distribuições, como a distribuição normal contaminada e a distribuição logística tipo I e II, as quantidades (integrais) d_{g_i} e f_{g_i} devem ser calculadas mediante algum método de aproximação.

Por outro lado, é possível mostrar que o vetor de parâmetros associado aos componentes de variância, $\boldsymbol{\tau}$, é ortogonal aos vetores de parâmetros $\boldsymbol{\beta}$ e \mathbf{f}_k ($k = 1, \dots, s$) associados aos efeitos fixos do modelo, respectivamente, isto é, verifica-se que

$$\begin{aligned} \mathcal{I}_p^{\beta\boldsymbol{\tau}}(\boldsymbol{\theta}) &= -\mathbb{E}\left\{\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\tau}^T}\right\} \\ &= \mathbf{0} \end{aligned} \tag{3.22}$$

e

$$\begin{aligned} \mathcal{I}_p^{\mathbf{f}_k\boldsymbol{\tau}}(\boldsymbol{\theta}) &= -\mathbb{E}\left\{\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\mathbf{f}_k\partial\boldsymbol{\tau}^T}\right\} \\ &= \mathbf{0}. \end{aligned} \tag{3.23}$$

A propriedade de ortogonalidade de $\boldsymbol{\tau}$ com relação a $\boldsymbol{\beta}$ e \mathbf{f}_k ($k = 1, \dots, s$) facilita o desenvolvimento do processo iterativo para estimar o vetor de parâmetros $\boldsymbol{\theta}$. Veja maiores detalhes a respeito dos resultados (3.22) e (3.23) no contexto paramétrico em Lange et al. (1989).

Na seguinte seção concentramos nossa discussão no desenvolvimento do processo iterativo para estimar o vetor de parâmetros $\boldsymbol{\theta}$.

3.5 Processo de estimação

Para nosso propósito de otimização vamos supor que o logaritmo da função de verossimilhança penalizada $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ definido por (3.6) é uma função convexa e que satisfaz certas condições de regularidade; veja em Gu (2002) uma discussão dessas condições de regularidade no contexto de regressão não paramétrica. Então, por analogia com o critério de máxima verossimilhança, o valor de $\boldsymbol{\theta}$ que maximiza

$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$, em todo o espaço paramétrico Θ , digamos $\widehat{\boldsymbol{\theta}}$, é chamado estimador de máxima verossimilhança penalizada (EMVP) de $\boldsymbol{\theta}$, e satisfaz a seguinte desigualdade:

$$L_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) \geq \sup_{\boldsymbol{\theta} \in \Theta} L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}). \quad (3.24)$$

Em geral, a existência do estimador de máxima verossimilhança penalizada tem sido discutida por diversos autores, como por exemplo, De Montricher et al. (1975), Tapia e Thompson (1978), e Silverman (1982). No caso específico dos modelos de regressão não paramétricos clássicos, O'Sullivan et al. (1986) mostram que se a verossimilhança penalizada é convexa, então a existência de um único estimador de máxima verossimilhança da função não paramétrica sobre o espaço de funções lineares, garante a existência do estimador de máxima verossimilhança penalizada no espaço de funções de Sobolev; veja também Buja et al. (1989).

Em nosso caso, como a função $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ depende de f_k ($k = 1, \dots, s$) através da avaliação funcional $f_k(\mathbf{t}_{k_g}^0)$ ($g = 1, \dots, r_k$), e \mathbf{f}_k e \mathbf{K}_k , sendo avaliados em $\mathbf{t}_k^0 = (\mathbf{t}_{k_1}^0, \dots, \mathbf{t}_{k_{r_k}}^0)^T$, têm dimensões finitas, temos que, sob certas condições de regularidade, o estimador de máxima verossimilhança penalizada de f_k sob o modelo (2.1) existe e é único, e corresponde a uma *spline* cúbica natural que é completamente determinada pelo vetor finito dimensional \mathbf{t}_k^0 ; ou seja, o estimador de máxima verossimilhança penalizada de f_k é admissível em um espaço finito. A vantagem de usar suavização *spline* é a possibilidade de reduzir um problema infinito dimensional a um problema finito dimensional sob o modelo (2.1). Geralmente os estimadores de máxima verossimilhança (penalizada) não podem ser expressos em forma explícita, e portanto necessita-se de um método iterativo para a obtenção das raízes das equações de máxima verossimilhança (penalizada) associadas. Nos casos em que as duas primeiras derivadas do logaritmo da função de verossimilhança (penalizada) existam, com relação aos parâmetros de interesse, os procedimentos usuais para calcular os estimadores de máxima verossimilhança (penalizada) estão baseados em uma expansão de série Taylor em torno de alguma estimativa inicial. Nesse caso, podemos usar o algoritmo de Newton-Raphson ou algoritmo score de Fisher.

3.5.1 Efeitos fixos

Em geral, para maximizar a verossimilhança penalizada associada a um modelo semiparamétrico existem vários procedimentos que podem ser considerados. Por exemplo, Green (1987) propõe encontrar as soluções das equações de estimação de máxima verossimilhança penalizada de um modelo de regressão semiparamétrico usando o algoritmo de Newton-Raphson. Hastie e Tibshirani (1990) sugerem ajustar um modelo aditivo generalizado através do algoritmo escore de Fisher. Lin e Zhang (1999) usam o algoritmo escore de Fisher para ajustar um modelo misto aditivo generalizado e mostram que os estimadores das funções não paramétricas podem ser obtidos maximizando a quase-verossimilhança duplamente penalizada. Green (1990) propõe maximizar a verossimilhança penalizada usando o algoritmo EM penalizado e o algoritmo One-Step-Late (OSL), o qual corresponde a uma versão modificada do algoritmo EM penalizado (Nychka, 1990). Uma discussão sobre a variância do estimador de máxima verossimilhança penalizada obtido através do algoritmo EM penalizado pode ser encontrada em Segal et al. (1994). Maiores detalhes do algoritmo EM podem ser encontrados em Dempster et al. (1977, 1981), McLachlan e Krishnan (1997) e Laird et al. (1987).

Baseados nesses resultados, neste trabalho sugerimos adaptar o algoritmo escore de Fisher para encontrar soluções para as equações de máxima verossimilhança penalizada e ajustar os componentes paramétrico e não paramétrico mediante a combinação dos procedimentos iterativos escore de Fisher e *backfitting*.

Equações de estimação

Consideremos o logaritmo da função de verossimilhança penalizada associado ao modelo marginal (2.12),

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}),$$

em que $L_{p_i}(\boldsymbol{\theta})$ é dada pela equação (3.7). Suponhamos que o vetor de parâmetros associado aos componentes de variância $\boldsymbol{\tau}$ e o vetor de parâmetros de suavização

α são fixos. Então, o vetor de parâmetros a ser estimado é $(\beta^T, \mathbf{f}_1^T, \dots, \mathbf{f}_s^T)^T$, cujas equações de máxima verossimilhança penalizada são dadas por

$$\begin{aligned} \mathbf{U}_p^\beta(\boldsymbol{\theta}) &= \sum_{i=1}^n v_i(\boldsymbol{\theta}) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \\ \mathbf{U}_p^{\mathbf{f}_1}(\boldsymbol{\theta}) &= \sum_{i=1}^n \left[v_i(\boldsymbol{\theta}) \mathbf{N}_{1i}^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] - \alpha_1 \mathbf{K}_1 \mathbf{f}_1 = \mathbf{0} \quad \text{e} \quad (3.25) \\ &\vdots \\ \mathbf{U}_p^{\mathbf{f}_s}(\boldsymbol{\theta}) &= \sum_{i=1}^n \left[v_i(\boldsymbol{\theta}) \mathbf{N}_{si}^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right] - \alpha_s \mathbf{K}_s \mathbf{f}_s = \mathbf{0}. \end{aligned}$$

É possível mostrar que as soluções para as equações de estimação acima conduzem às estimativas de máxima verossimilhança penalizada, digamos $\widehat{\beta}$ e $\widehat{\mathbf{f}}_k$ ($k = 1, \dots, s$).

Algoritmo escore de Fisher

Assumindo que o vetor de parâmetros de suavização α e a matriz $\boldsymbol{\Sigma}_i$ são fixos, podemos mostrar, usando (3.10) e (3.17), que a $(u + 1)$ -ésima etapa do processo iterativo escore de Fisher que permite resolver (3.25) é dada pela seguinte equação matricial:

$$\begin{aligned} &\begin{pmatrix} \sum_{i=1}^n \mathcal{I}_{p_i}^{\beta\beta}(\boldsymbol{\theta}) & \sum_{i=1}^n \mathcal{I}_{p_i}^{\beta\mathbf{f}_1}(\boldsymbol{\theta}) & \dots & \sum_{i=1}^n \mathcal{I}_{p_i}^{\beta\mathbf{f}_s}(\boldsymbol{\theta}) \\ \sum_{i=1}^n \mathcal{I}_{p_i}^{\mathbf{f}_1\beta}(\boldsymbol{\theta}) & \sum_{i=1}^n \mathcal{I}_{p_i}^{\mathbf{f}_1\mathbf{f}_1}(\boldsymbol{\theta}) & \dots & \sum_{i=1}^n \mathcal{I}_{p_i}^{\mathbf{f}_1\mathbf{f}_s}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \mathcal{I}_{p_i}^{\mathbf{f}_s\beta}(\boldsymbol{\theta}) & \sum_{i=1}^n \mathcal{I}_{p_i}^{\mathbf{f}_s\mathbf{f}_1}(\boldsymbol{\theta}) & \dots & \sum_{i=1}^n \mathcal{I}_{p_i}^{\mathbf{f}_s\mathbf{f}_s}(\boldsymbol{\theta}) \end{pmatrix}^{(u)} \begin{pmatrix} \Phi_0 \\ \Phi_1 \\ \vdots \\ \Phi_s \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{U}_p^\beta(\boldsymbol{\theta}) \\ \mathbf{U}_p^{\mathbf{f}_1}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{U}_p^{\mathbf{f}_s}(\boldsymbol{\theta}) \end{pmatrix}^{(u)}, \quad (3.26) \end{aligned}$$

em que $\Phi_0 = \beta^{(u+1)} - \beta^{(u)}$ e $\Phi_k = \mathbf{f}_k^{(u+1)} - \mathbf{f}_k^{(u)}$ ($k = 1, \dots, s$), para $u = 0, 1, 2, \dots$. Cálculos diretos mostram que os elementos da matriz de informação de Fisher e do vetor escore da equação acima podem ser escritos em uma forma matricial conveniente. Com efeito, podemos escrever

$$\sum_{i=1}^n \mathcal{I}_{\mathbf{p}_i}^{\beta\beta}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{4d_{g_i}}{m_i} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i$$

da seguinte forma:

$$\begin{aligned} \sum_{i=1}^n \frac{4d_{g_i}}{m_i} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i &= \frac{4d_{g_1}}{m_1} \mathbf{X}_1^T \mathbf{W}_1 \mathbf{X}_1 + \dots + \frac{4d_{g_n}}{m_n} \mathbf{X}_n^T \mathbf{W}_n \mathbf{X}_n \\ &= \mathbf{X}^T \mathbf{W}^* \mathbf{X}, \end{aligned}$$

em que $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ e $\mathbf{W}^* = \text{diag}\left(\frac{4d_{g_1}}{m_1} \mathbf{W}_1, \dots, \frac{4d_{g_n}}{m_n} \mathbf{W}_n\right)$ é uma matriz ($n^* \times n^*$), com $\mathbf{W}_i = \boldsymbol{\Sigma}_i^{-1}$ e $n^* = \sum_{i=1}^n m_i$. Analogamente, podemos escrever o elemento

$$\sum_{i=1}^n \mathcal{I}_{\mathbf{p}_i}^{\beta \mathbf{f}_k}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{4d_{g_i}}{m_i} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{ki}$$

na forma

$$\begin{aligned} \sum_{i=1}^n \frac{4d_{g_i}}{m_i} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{ki} &= \frac{4d_{g_1}}{m_1} \mathbf{X}_1^T \mathbf{W}_1 \mathbf{N}_{k1} + \dots + \frac{4d_{g_n}}{m_n} \mathbf{X}_n^T \mathbf{W}_n \mathbf{N}_{kn} \\ &= \mathbf{X}^T \mathbf{W}^* \mathbf{N}_k, \end{aligned}$$

sendo a matriz $\mathbf{N}_k = (\mathbf{N}_{k1}^T, \dots, \mathbf{N}_{kn}^T)^T$. Similarmente, podemos escrever as matrizes

$$\sum_{i=1}^n \mathcal{I}_{\mathbf{p}_i}^{\mathbf{f}_k \mathbf{f}_{k'}}(\boldsymbol{\theta}) = \begin{cases} \sum_{i=1}^n \mathbf{N}_{ki}^T \mathbf{W}^* \mathbf{N}_{ki} + \frac{\alpha_k}{n} \mathbf{K}_k & k = k' \\ \sum_{i=1}^n \mathbf{N}_{ki}^T \mathbf{W}^* \mathbf{N}_{k'i} & k \neq k'. \end{cases}$$

O procedimento é análogo para os elementos do vetor escore. Nessas condições, podemos escrever a equação (3.26) na forma

$$\begin{pmatrix} \mathbf{X}^T \mathbf{W}^* \mathbf{X} & \mathbf{X}^T \mathbf{W}^* \mathbf{N}_1 & \dots & \mathbf{X}^T \mathbf{W}^* \mathbf{N}_s \\ \mathbf{N}_1^T \mathbf{W}^* \mathbf{X} & \mathbf{N}_1^T \mathbf{W}^* \mathbf{N}_1 + \alpha_1 \mathbf{K}_1 & \dots & \mathbf{N}_1^T \mathbf{W}^* \mathbf{N}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{N}_s^T \mathbf{W}^* \mathbf{X} & \mathbf{N}_s^T \mathbf{W}^* \mathbf{N}_1 & \dots & \mathbf{N}_s^T \mathbf{W}^* \mathbf{N}_s + \alpha_s \mathbf{K}_s \end{pmatrix}^{(u)} \begin{pmatrix} \Phi_0 \\ \Phi_1 \\ \vdots \\ \Phi_s \end{pmatrix} \\ = \begin{pmatrix} \mathbf{X}^T \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) \\ \mathbf{N}_1^T \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) - \alpha_1 \mathbf{K}_1 \mathbf{f}_1 \\ \vdots \\ \mathbf{N}_s^T \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) - \alpha_s \mathbf{K}_s \mathbf{f}_s \end{pmatrix}^{(u)},$$

em que $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$ é um vetor ($n^* \times 1$), $\mathbf{W} = \text{diag}(v_1 \mathbf{W}_1, \dots, v_n \mathbf{W}_n)$ é uma matriz ($n^* \times n^*$), e $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_k \mathbf{f}_k$ é um vetor ($n^* \times 1$). Note que as quantidades $v_i = v_i(\boldsymbol{\theta})$ ($i = 1, \dots, n$) e o vetor $\boldsymbol{\mu}$ devem ser avaliadas em $\boldsymbol{\beta} = \boldsymbol{\beta}^{(u)}$ e $\mathbf{f}_k = \mathbf{f}_k^{(u)}$ ($k = 1, \dots, s$). Assim, após algumas manipulações algébricas, é possível escrever a equação matricial acima na forma

$$\begin{pmatrix} \boldsymbol{\beta}^{(u+1)} \\ \mathbf{f}_1^{(u+1)} \\ \vdots \\ \mathbf{f}_s^{(u+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0 \{ \mathbf{r}_{\beta-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \\ \mathbf{S}_1 \{ \mathbf{r}_{\mathbf{f}_1-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \\ \vdots \\ \mathbf{S}_s \{ \mathbf{r}_{\mathbf{f}_s-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \end{pmatrix}, \quad (3.27)$$

em que

$$\mathbf{S}_k = \begin{cases} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^* & k = 0 \\ (\mathbf{N}_k^T \mathbf{W}^* \mathbf{N}_k + \alpha_k \mathbf{K}_k)^{-1} \mathbf{N}_k^T \mathbf{W}^* & k = 1, \dots, s \end{cases}$$

são matrizes de suavização e

$$\mathbf{r}_{\vartheta-}^{(u,u+1)} = \begin{cases} (\mathbf{I}_{n^*} - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \sum_{k=1}^s \mathbf{N}_k \mathbf{f}_k^{(u+1)} & \vartheta = \beta \\ (\mathbf{I}_{n^*} - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(u+1)} - \sum_{l=1, k \neq l}^s \mathbf{N}_k \mathbf{f}_k^{(u+1)} & \vartheta = \mathbf{f}_k, \end{cases}$$

são resíduos parciais que permitem ajustar o componente paramétrico e o componente aditivo do modelo atualizando a matriz $\widetilde{\mathbf{W}}^{(u)} = \mathbf{I}_{n^*} - \mathbf{W}^{*-1} \mathbf{W}^{(u)}$ e o vetor $\boldsymbol{\mu}^{(u)}$ em cada etapa do processo. Nas expressões acima podemos observar que tanto as matrizes de suavização quanto os resíduos parciais, dependem da distribuição elíptica através das quantidades $v_i(\boldsymbol{\theta})$ e d_{g_i} . Note também que, em geral, as matrizes \mathbf{D} e \mathbf{V}_i são desconhecidas e devem ser estimadas através de algum processo de estimação. Da mesma forma, quando o vetor de parâmetros de suavização é desconhecido, devemos estimá-lo através de algum método apropriada, como por exemplo, o método de validação cruzada generalizada (este tópico será discutido na seção 3.7). Porém, em alguns casos, o pesquisador escolhe valores específicos para esses parâmetros de acordo com a sua conveniência.

Algoritmo backfitting

Em geral, resolver a equação (3.27) de maneira direta não é muito apropriado desde o ponto de vista prático e portanto sugerimos aproximar sua solução, como é usual no ajuste de um modelo aditivo geral, através do algoritmo *backfitting* (Gauss-Seidel). Embora o algoritmo *backfitting* seja uma técnica iterativa que fornece dificuldades adicionais no desenvolvimento da teoria assintótica, o método tem sido refinado e estendido para modelos mais complexos; vide, por exemplo, Green (1985), Stone (1986), Hastie e Tibshirani (1987), e Buja et al. (1989). Geralmente, as estimativas finais geradas pelo algoritmo *backfitting* podem depender dos valores iniciais das funções ou dos critérios de convergência estabelecidos. Nesse contexto, Buja et al. (1989) mostram que o algoritmo *backfitting* coincide com o método iterativo de Gauss-Seidel e estabelecem algumas condições de regularidade para garantir a consistência das equações de estimação (normais) e a convergência

do algoritmo; veja também outros detalhes em Berhane e Tibshirani (1998) que mostram que o algoritmo sempre converge quando os suavizadores envolvidos no processo formam uma *spline* cúbica. No caso de suavizadores simétricos e com autovalores no intervalo $[0, 1]$ que não apresentam *concurvity*¹, o algoritmo converge para uma única solução, independente dos valores iniciais das funções usadas no processo iterativo. No entanto, se os suavizadores apresentam *concurvity*, o algoritmo converge a alguma solução da equação, e os valores iniciais das funções determinam as soluções finais do processo.

Tabela 3.2: $(u^* + 1)$ -ésima etapa do algoritmo *backfitting* na $(u + 1)$ -ésima etapa do processo iterativo escore de Fisher sob o modelo misto aditivo semiparamétrico elíptico.

-
-
- (i) Inicie o processo iterativo com $\boldsymbol{\beta}^{(u)} = \boldsymbol{\beta}^{(u,0)}$ e $\mathbf{f}_k^{(u)} = \mathbf{f}_k^{(0,0)}$, $k = 1, \dots, s$
- (ii) Para $k = 1, \dots, s, 1, \dots, s, \dots$ e $u^* = 0, 1, 2, \dots$ calcular
- (ii')
- $$\mathbf{r}_{\beta^-}^{(u, u^*)} = (\mathbf{I}_{n^*} - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \sum_{k=1}^s \mathbf{N}_k \mathbf{f}_k^{(u, u^*)} \quad \text{e}$$
- $$\boldsymbol{\beta}^{(u+1, u^*+1)} = \mathbf{S}_0 \{ \mathbf{r}_{\beta^-}^{(u, u^*)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} .$$
- (ii'')
- $$\mathbf{r}_{\mathbf{f}_k^-}^{(u, u^*)} = (\mathbf{I}_{n^*} - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(u+1, u^*+1)} - \sum_{l=1, l \neq k}^s \mathbf{N}_l \mathbf{f}_l^{(u, u^*)} \quad \text{e}$$
- $$\mathbf{f}_k^{(u+1, u^*+1)} = \mathbf{S}_k \{ \mathbf{r}_{\mathbf{f}_k^-}^{(u, u^*)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} .$$
- (iii) Repita (i) e (ii) até atingir a convergência desejada .
-
-

¹Conceito análogo ao conceito de colinearidade usado na teoria dos modelos lineares

3.5.2 Componentes de variância

Assumindo que o vetor de parâmetros de suavização $\boldsymbol{\alpha}$ é fixo, podemos obter a estimativa de máxima verossimilhança penalizada de $\boldsymbol{\tau} = (\boldsymbol{\lambda}^T, \boldsymbol{\gamma}^T)^T$ através do seguinte processo iterativo:

$$\boldsymbol{\tau}^{(u+1)} = \arg \max_{\boldsymbol{\tau}} \left\{ L_p^c(\boldsymbol{\beta}_\tau^{(u+1)}, \mathbf{f}_{1\tau}^{(u+1)}, \dots, \mathbf{f}_{s\tau}^{(u+1)}, \boldsymbol{\tau}, \boldsymbol{\alpha}) \right\}, \quad (3.28)$$

em que $L_p^c(\boldsymbol{\beta}_\tau^{(u+1)}, \mathbf{f}_{1\tau}^{(u+1)}, \dots, \mathbf{f}_{s\tau}^{(u+1)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ denota o logaritmo da função de verossimilhança penalizada concentrada de $\boldsymbol{\tau}$ definido por

$$\begin{aligned} L_p^c(\boldsymbol{\beta}_\tau^{(u+1)}, \mathbf{f}_{1\tau}^{(u+1)}, \dots, \mathbf{f}_{s\tau}^{(u+1)}, \boldsymbol{\tau}, \boldsymbol{\alpha}) &= \sum_{i=1}^n \left\{ -\frac{1}{2} \log |\boldsymbol{\Sigma}_{\tau_i}| + \log g(\tilde{\delta}_{\tau_i}) \right\} \\ &\quad - \frac{1}{2} \sum_{k=1}^s \alpha_k \mathbf{f}_{k\tau}^{(u+1)T} \mathbf{K}_k \mathbf{f}_{k\tau}^{(u+1)}, \end{aligned}$$

em que $\tilde{\delta}_{\tau_i} = \boldsymbol{\varepsilon}_{\tau_i}^T \boldsymbol{\Sigma}_{\tau_i}^{-1} \boldsymbol{\varepsilon}_{\tau_i}$, com $\boldsymbol{\varepsilon}_{\tau_i} = \left(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_\tau^{(u+1)} - \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_{k\tau}^{(u+1)} \right)$ e $u = 0, 1, \dots$. Note que o subíndice τ foi incorporado para denotar a dependência com respeito ao vetor de parâmetros $\boldsymbol{\tau}$. A maximização de (3.28) pode ser efetuada, por exemplo, através do algoritmo de quase-Newton, escore de Fisher, EM, ou secante multivariado. Tais algoritmos conduzem aproximadamente à estimativa de máxima verossimilhança penalizada de $\boldsymbol{\tau}$, digamos $\hat{\boldsymbol{\tau}}$. No caso específico do algoritmo escore de Fisher, temos que

$$\boldsymbol{\tau}^{(u+1)} = \boldsymbol{\tau}^{(u)} + \left\{ \mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta})^{(u)} \right\}^{-1} \mathbf{U}_p^\tau(\boldsymbol{\theta})^{(u)}, \quad (3.29)$$

com $\mathbf{U}_p^\tau(\boldsymbol{\theta})$ e $\mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta})$ definidas pelas equações (3.13) e (3.21), respectivamente.

3.5.3 Processo iterativo

Finalmente, o processo iterativo conjunto para obter a estimativa de máxima verossimilhança penalizada de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}_1^T, \dots, \mathbf{f}_s^T, \boldsymbol{\tau}^T)^T$, baseado no algoritmo escore de Fisher, é dado por

$$\begin{pmatrix} \boldsymbol{\beta}^{(u+1)} \\ \mathbf{f}_1^{(u+1)} \\ \vdots \\ \mathbf{f}_s^{(u+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0 \{ \mathbf{r}_{\beta-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \\ \mathbf{S}_1 \{ \mathbf{r}_{f_1-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \\ \vdots \\ \mathbf{S}_s \{ \mathbf{r}_{f_s-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \end{pmatrix}$$

e

$$\boldsymbol{\tau}^{(u+1)} = \boldsymbol{\tau}^{(u)} + \left\{ \mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta})^{(u)} \right\}^{-1} \mathbf{U}_p^\tau(\boldsymbol{\theta})^{(u)}, \quad u = 0, 1, \dots$$

Para iniciar o processo iterativo descrito acima, valores iniciais $\boldsymbol{\beta}^{(0)}$, $\mathbf{f}_k^{(0)}$ ($k = 1, \dots, s$) e $\boldsymbol{\tau}^{(0)}$ devem ser fornecidos. Em particular, para os modelos não gaussianos elípticos, podemos considerar as estimativas obtidas a partir do modelo normal para iniciar o processo.

3.5.4 Efeitos aleatórios

Em muitas aplicações práticas é preciso estimar os efeitos aleatórios. Nesse sentido, pode-se usar o fato de que a média condicional de \mathbf{b}_i , dado o valor observado \mathbf{y}_i , segue uma distribuição elíptica. Especificamente falando, a partir da distribuição conjunta

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{pmatrix} \sim \mathcal{E}_{\ell(m_i+q)} \left\{ \begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i & \mathbf{Z}_i \mathbf{D} \\ \mathbf{D} \mathbf{Z}_i^T & \mathbf{D} \end{pmatrix} \right\},$$

podemos mostrar, mediante a Propriedade 1.4.4 das distribuições elípticas apresentada na Seção 1.4, que a distribuição condicional de $\mathbf{b}_i | \mathbf{y}_i$ é da forma

$$\mathbf{b}_i | \mathbf{y}_i \sim \mathcal{E}_{\ell_q} \left\{ \mathbf{D} \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k \right), \mathbf{D} - \mathbf{D} \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \mathbf{D} \right\}. \quad (3.30)$$

Veja a prova desse resultado no Apêndice B. Daí segue que, para $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i$ fixa, o estimador de Bayes empírico dos efeitos aleatórios \mathbf{b}_i é dado por

$$\begin{aligned}\widehat{\mathbf{b}}_i &= \mathbb{E}\{\widehat{\mathbf{b}}_i \mid \mathbf{y}_i\} \\ &= \mathbf{D} \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} - \sum_{k=1}^s \mathbf{N}_{ki} \widehat{\mathbf{f}}_k \right),\end{aligned}\tag{3.31}$$

em que $\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}}_1, \dots, \widehat{\mathbf{f}}_s$ correspondem às estimativas de máxima verossimilhança penalizada. Assim, o vetor de efeitos aleatórios estimados é dado por

$$\widehat{\mathbf{b}} = (\widehat{\mathbf{b}}_1^T, \dots, \widehat{\mathbf{b}}_n^T)^T.\tag{3.32}$$

Para uma revisão do processo de estimação de \mathbf{b}_i nos modelos mistos lineares normais veja Harville (1976, 1977) e Laird e Ware (1982), entre outros; para os modelos mistos lineares elípticos, veja Savalli et al. (2006); para os modelos mistos semiparamétricos normais veja Zhang et al. (1998) e Fung et al. (2002); e, finalmente, para os modelos mistos aditivos generalizados veja Lin e Zhang (1999).

Consequentemente, o valor ajustado para o i -ésimo grupo pode ser expresso na forma

$$\begin{aligned}\widehat{\mathbf{y}}_i &= \mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{Z}_i \widehat{\mathbf{b}}_i + \sum_{k=1}^s \mathbf{N}_{ki} \widehat{\mathbf{f}}_k \\ &= \widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1} \left(\mathbf{X}_i \widehat{\boldsymbol{\beta}} + \sum_{k=1}^s \mathbf{N}_{ki} \widehat{\mathbf{f}}_k \right) + (\mathbf{I}_{m_i} - \widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1}) \mathbf{y}_i.\end{aligned}$$

Observamos que $\widehat{\mathbf{y}}_i$ pode ser interpretada como uma média ponderada entre o perfil da população $(\mathbf{X}_i \widehat{\boldsymbol{\beta}} + \sum_{k=1}^s \mathbf{N}_{ki} \widehat{\mathbf{f}}_k)$ e os dados observados \mathbf{y}_i , com pesos $\widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1}$ e $(\mathbf{I}_{m_i} - \widehat{\mathbf{V}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1})$, respectivamente. Note que quando a variabilidade residual \mathbf{V}_i é maior do que a variabilidade intraunidades amostrais $\boldsymbol{\Sigma}_i$, maior ponderação será dada ao perfil ajustado $(\mathbf{X}_i \widehat{\boldsymbol{\beta}} + \sum_{k=1}^s \mathbf{N}_{ki} \widehat{\mathbf{f}}_k)$. Por outro lado, quando a variabilidade intraunidades experimentais for maior do que a variabilidade residual, maior ponderação será atribuída ao valor observado \mathbf{y}_i .

3.5.5 Erro padrão

Na teoria paramétrica clássica é bem conhecido o fato de que se as estimativas de máxima verossimilhança são obtidas via o algoritmo de Newton-Raphson, escore de Fisher ou EM, as matrizes de variância-covariância assintóticas dessas estimativas podem ser obtidas a partir da inversa da matriz de informação observada ou esperada (veja, por exemplo, Lehmann, 1999; Louis, 1982 e Meng e Rubin, 1991). Da mesma maneira, na teoria não paramétrica ou semiparamétrica, vários autores têm discutido o problema de estimar as matrizes de variância-covariância das estimativas de máxima verossimilhança penalizada. No caso específico do modelo não paramétrico clássico,

$$y_i = f(t_i) + \epsilon_i \quad (i = 1, \dots, n), \quad (3.33)$$

em que $t_1 < \dots < t_n$ e os ϵ_i são erros aleatórios independentes com distribuição $\mathcal{N}(0, \phi)$, Wahba (1983) e Silverman (1985) propõem estimar a matriz de variância-covariância da estimativa de máxima verossimilhança $\hat{\mathbf{f}} = (f(t_1), \dots, f(t_n))^T$ usando a matriz de variância-covariância a posteriori de $\hat{\mathbf{f}}$ sob o modelo bayesiano

$$f(t) = A + Bt + \alpha^{-1/2} \int_0^t \mathcal{W}(u) du, \quad (3.34)$$

em que A e B são variáveis aleatórias que têm distribuições uniformes impróprias (a priori não informativa) no intervalo $(-\infty, \infty)$, respectivamente, e $\mathcal{W}(u)$ é um processo *Wiener* padrão². Da mesma maneira, Zhang et al. (1998) propõem estimar as matrizes de variância-covariância das estimativas de máxima verossimilhança penalizada sob o modelo misto semiparamétrico normal, usando as correspondentes matrizes de variância-covariância a posteriori sob o modelo bayesiano (3.34). Contudo, podemos observar que, em ambos os dois trabalhos, as matrizes de variância-covariância das estimativas de máxima verossimilhança penalizada, ainda que obtidas a partir das matrizes de variância-covariância a posteriori de uma distribuição

²As vezes alguns autores consideram o modelo bayesiano finito dimensional proposto por Green e Silverman (1994).

normal multivariada, correspondem à inversa da matriz de informação observada calculada a partir da função de verossimilhança penalizada, uma vez que essa função é tratada como uma verossimilhança usual; dito de outra forma, as matrizes de variância-covariância podem ser obtidas a partir das matrizes de segundas derivadas parciais da verossimilhança penalizada (Segal et al., 1994).

Motivados pelos resultados desses autores e pelo fato de nós utilizar o algoritmo score de Fisher para obter as estimativas de máxima verossimilhança, neste trabalho sugerimos estimar as matrizes de variância-covariância assintóticas dessas estimativas (o erro padrão) baseados na inversa da matriz de informação de Fisher penalizada $\mathcal{I}_p^{-1}(\boldsymbol{\theta})$ definida na Seção 3.15. Dessa forma, a matriz de variância-covariância assintótica estimada de $\widehat{\boldsymbol{\theta}}$ é dada por

$$\widehat{\text{Cov}}(\widehat{\boldsymbol{\theta}}) \approx \mathcal{I}_p^{-1}(\widehat{\boldsymbol{\theta}}). \quad (3.35)$$

Em particular,

$$\begin{aligned} \widehat{\text{Cov}}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}}) &\approx \mathcal{I}_p^{\beta\mathbf{f}^{-1}}(\widehat{\boldsymbol{\theta}}) && \text{e} \\ \widehat{\text{Cov}}(\widehat{\boldsymbol{\tau}}) &\approx \mathcal{I}_p^{\tau\tau^{-1}}(\widehat{\boldsymbol{\theta}}), \end{aligned}$$

em que $\widehat{\mathbf{f}} = (\widehat{\mathbf{f}}_1, \dots, \widehat{\mathbf{f}}_s)$ denota a estimativa de máxima verossimilhança penalizada. No contexto dos modelos lineares parciais, Heckman (1986) prova a consistência e normalidade assintótica do estimador do coeficiente de regressão e mostra que o viés é assintoticamente desprezível (veja também Green, 1987). Por sua parte, Zhang et al. (1998) apresentam uma interessante discussão sobre o comportamento assintótico dos estimadores de máxima verossimilhança penalizada no caso normal, e derivam expressões fechadas para as matrizes de variância-covariância desses estimadores desde o ponto de vista frequentista e bayesiano. E, recentemente, Durban et al. (1999) apresentam um método, computacionalmente atrativo, que permite aproximar o erro padrão da estimativa do coeficiente de regressão no contexto dos modelos aditivos semiparamétricos; veja também Flanders et al. (2005).

Na seguinte Seção descrevemos um procedimento para representar o modelo misto aditivo semiparamétrico como um modelo misto puramente paramétrico quando os estimadores das funções não paramétricas conduzem a uma *spline* cúbica. Esse procedimento foi discutido inicialmente por Green (1987) no contexto dos modelos semiparamétricos e posteriormente por Zhang et al. (1998) no contexto dos modelos mistos semiparamétricos gaussianos.

3.6 Modelo misto modificado normal

Quando usamos suavização *spline* para estimar as funções não paramétricas, existem algumas conexões com o modelo misto usual (Wang, 1998). Com efeito, de acordo com Green (1987), Zhang et al. (1998) e Lin e Zhang (1999), os componentes não paramétricos do modelo (2.7), \mathbf{f}_k ($k = 1, \dots, s$), podem ser representados através de uma transformação linear 1-1 da seguinte maneira:

$$\mathbf{f}_k = \mathbf{T}_k \boldsymbol{\delta}_k + \mathbf{B}_k \mathbf{a}_k, \quad (3.36)$$

em que $\boldsymbol{\delta}_k \in \mathbb{R}^2$, $\mathbf{a}_k \in \mathbb{R}^{(v_k-2)}$, $\mathbf{B}_k = \mathbf{L}_k(\mathbf{L}_k^T \mathbf{L}_k)^{-1}$, $\mathbf{L}_k \in \mathbb{R}^{q_k \times (q_k-2)}$ é uma matriz de posto completo que satisfaz $\mathbf{K}_k = \mathbf{L}_k \mathbf{L}_k^T$ e $\mathbf{L}_k^T \mathbf{T}_k = \mathbf{0}$, com $\mathbf{T}_k = (\mathbf{1}_{r_k} \mathbf{t}_k^0) \in \mathbb{R}^{r_k \times 2}$. Logo, usando a igualdade $\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k = \mathbf{a}_k^T \mathbf{a}_k$, podemos escrever o logaritmo da função de verossimilhança penalizada (3.6) para o caso gaussiano na forma

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = cte. - \frac{1}{2} \sum_{i=1}^n \ln |\boldsymbol{\Sigma}_i| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) - \frac{1}{2} \sum_{k=1}^s \sigma_k \mathbf{a}_k^T \mathbf{a}_k,$$

em que $\sigma_k = 1/\alpha_k$. Substituindo (3.36) na equação (2.7), temos que as respostas observadas são modeladas da seguinte maneira:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \sum_{k=1}^s \left(\mathbf{N}_{ki} \mathbf{T}_k \boldsymbol{\delta}_k + \mathbf{N}_{ki} \mathbf{B}_k \mathbf{a}_k \right) + \boldsymbol{\epsilon}_i, \quad (3.37)$$

em que $(\boldsymbol{\beta}^T, \boldsymbol{\delta}_1^T, \dots, \boldsymbol{\delta}_s^T)^T$ corresponde ao vetor de parâmetros associado aos efeitos fixos do modelo e $\mathbf{b}_* = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T, \mathbf{a}_1^T, \dots, \mathbf{a}_s^T)^T$ é o vetor de efeitos aleatórios. No-

temos que a conexão entre o modelo semiparamétrico (2.7) e o modelo misto (3.37) se estabelece através de um modelo com efeitos fixos e aleatórios puramente paramétricos. A importância dessa conexão baseia-se principalmente na possibilidade de estimar as funções não paramétricas ajustando um modelo paramétrico. Especificamente, podemos estimar \mathbf{f}_k da forma

$$\widehat{\mathbf{f}}_k = \mathbf{T}_k \widehat{\boldsymbol{\delta}}_k + \mathbf{B}_k \widehat{\mathbf{a}}_k, \quad (3.38)$$

em que $\widehat{\mathbf{f}}_k$ é uma combinação linear das estimativas dos efeitos fixos $\widehat{\boldsymbol{\delta}}_k$ e dos efeitos aleatórios $\widehat{\mathbf{a}}_k$ obtidas pelo ajuste do modelo misto paramétrico (3.37) usando os métodos de estimação tradicionais e as ferramentas computacionais já desenvolvidas. Além disso, podemos estimar o parâmetro de suavização incorporando seu inverso como um componente de variância extra em no modelo misto modificado (3.37). Na seguinte seção apresentamos alguns métodos que permitem obter estimativas para os parâmetros de suavização. Especificamente, consideraremos o método de validação cruzada e o método de validação cruzada generalizada. Salientamos que esta discussão trata apenas o problema de estimação desses parâmetros para alguns casos particulares do modelo proposto neste trabalho.

3.7 Os parâmetros de suavização

Nas seções anteriores consideramos os parâmetros de suavização $\alpha_1, \dots, \alpha_s$ fixos para fazer inferência para as funções não paramétricas f_1, \dots, f_s . Porém, na prática esses parâmetros devem ser estimados a partir dos dados. No caso em que usamos um procedimento de suavização *spline* é usual usar, por exemplo, o método de validação cruzada (Wahba e Wold, 1975) ou o método de validação cruzada generalizada (Craven e Wahba, 1979; Eubank, 1988). Detalhes referentes aos métodos usados para estimar os parâmetros de suavização no contexto dos modelos aditivos podem ser encontrados, por exemplo, em Buja et al. (1989), Opsomer e Ruppert (1998), Rigby e Stasinopoulos (2005) e Hastie e Tibshirani (1990), dentre outros.

3.7.1 Validação cruzada

Consideremos o modelo não paramétrico (3.33). Validação cruzada (VC) é um método usado para estimar o erro de predição para um modelo ajustado aos dados. Tal método, usa parte dos dados para estimar o modelo e o restante para avaliar se ele é adequado ou não. De acordo com Green e Silverman (1994), quando a estimativa da função não paramétrica f é uma *spline* cúbica natural, a expressão do escore de validação cruzada é dada por

$$VC(\alpha) = n^{-1} \sum_{i=1}^n \left(\frac{y_i - \hat{f}(t_i, \alpha)}{1 - h_{ii}(\alpha)} \right)^2,$$

em que $\hat{\mathbf{f}}(\alpha) = \mathbf{H}(\alpha) \mathbf{y}$, sendo $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\hat{\mathbf{f}}(\alpha) = (\hat{f}(t_1, \alpha), \dots, \hat{f}(t_n, \alpha))^T$ a estimativa de f , para um valor de α fixo, e $h_{ii}(\alpha)$ o i -ésimo elemento diagonal da matriz $\mathbf{H}(\alpha)$, denominada matriz *hat* ou matriz de influência, definida na forma

$$\mathbf{H}(\alpha) = (\mathbf{I}_n + \alpha \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T)^{-1}, \quad (3.39)$$

sendo as matrizes \mathbf{Q} e \mathbf{R} definidas na Subseção 3.2.1. Green e Silverman (1994) fornecem um algoritmo eficiente para obter os elementos $h_{ii}(\alpha)$ e um procedimento alternativo para calcular $\text{tr}\{\mathbf{H}(\alpha)\}$ baseado nos autovalores da matriz *hat*. A idéia básica de validação cruzada é escolher o valor de α que minimize $VC(\alpha)$. Como não há garantia de que a função $VC(\alpha)$ tenha um único mínimo, cuidados devem ser tomados com sua minimização. Uma rede de procura é, provavelmente, o melhor caminho a ser seguido. Além disso, qualquer método de minimização que for utilizado envolverá o cálculo de $VC(\alpha)$ para um número de valores de α e, por isso, é importante que se utilize um método eficiente para o cálculo de $VC(\alpha)$. Silverman (1984) propõe um método de validação cruzada aproximado e mostra, através de um estudo de simulação, que tal método tem boas propriedades estatísticas. Hall e Titterington (1987), propõem e comparam dois métodos para escolher o parâmetro de suavização, o primeiro baseado em uma medida de risco, e o outro, baseado em uma medida da qualidade do ajuste do modelo.

3.7.2 Validação cruzada generalizada

Validação cruzada generalizada (VCG) é uma forma modificada de validação cruzada, e é um método popular para encontrar o parâmetro de suavização. A idéia básica de VCG é substituir $1 - h_{ii}(\alpha)$ pela sua média, $1 - n^{-1}\text{tr}\{\mathbf{H}(\alpha)\}$. Assim, o escore de validação cruzada generalizada é construído por analogia com validação cruzada. Dessa forma temos que,

$$\text{VCG}(\alpha) = n^{-1} \frac{\sum_{i=1}^n (y_i - \hat{f}(t_i, \alpha))^2}{(1 - n^{-1}\text{tr}\{\mathbf{H}(\alpha)\})^2}.$$

Da mesma maneira que no caso de validação cruzada, a escolha do parâmetro de suavização para validação cruzada generalizada é realizada pela minimização do escore $\text{VCG}(\alpha)$. No caso específico em que todos os $h_{ii}(\alpha)$ são iguais, o método de validação cruzada generalizada seria idêntico ao método de validação cruzada. O'Sullivan et al. (1986) mostra, com base em argumentos assintóticos e resultados de simulação, que validação cruzada generalizada tem um comportamento satisfatório desde o ponto de vista do critério do erro quadrático médio ponderado (veja também Wahba, 1985; Wecker e Ansley, 1983; Wang, 1998 e Kohn et al. 1991).

3.8 Seleção de modelos

Nesta seção trataremos de maneira sucinta a seleção de modelos na classe de modelos MMASE. Devido a que os parâmetros do modelo são estimados via máxima verossimilhança penalizada, convém utilizarmos, em analogia com o caso paramétrico, o critério de informação de Akaike (AIC) (Akaike, 1974), ou algumas das suas variantes, como por exemplo, o critério de informação de Schwarz (SIC) (Schwarz, 1978); veja também Burnham e Anderson (1998) e Pauler (1998). O critério de informação de Schwarz definido na classe dos modelos MMASEs é dado por

$$\text{SIC}(\hat{\boldsymbol{\theta}}) = -2L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) + p^* \log n, \quad (3.40)$$

em que $L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\alpha})$ corresponde ao logaritmo da função de verossimilhança penalizada, definido na equação (3.6), avaliado na estimativa de máxima verossimilhança penalizada de $\boldsymbol{\theta}$, $p^* = p + r + d^*$ corresponde o número total de parâmetros no modelo, com $r = \sum_{k=1}^s r_k$ e $d^* = d_\lambda + d_\gamma$, e n denota o tamanho da amostra. Note que maximizar o logaritmo da função de verossimilhança penalizada é equivalente a minimizar o critério de informação de Schwarz. Logo, do ponto de vista prático, escolhemos, dentre os modelos considerados, aquele que apresente o menor valor de $\text{SIC}(\hat{\boldsymbol{\theta}})$. O critério de informação de Schwarz também pode ser utilizado para escolher os graus de liberdade da distribuição t-Student (no caso univariado ou multivariado), e na seleção dos parâmetros de suavização, sendo escolhida a combinação dos parâmetros que minimiza o valor de $\text{SIC}(\hat{\boldsymbol{\theta}})$. Outros detalhes sobre o critério de informação de Schwarz no contexto dos modelos não paramétricos, semiparamétricos e aditivos, podem ser encontrados, por exemplo, em Hurvich et al. (1998), Simonoff e Tsai (1999) e Durban et al. (2003), e Rigby e Stasinopoulos (2005), respectivamente.

3.9 Conclusões do capítulo

A contribuição principal deste capítulo foi o desenvolvimento de um processo iterativo para estimar os parâmetros do modelo misto aditivo semiparamétrico elíptico. Especificamente, a partir do logaritmo da função de verossimilhança penalizada do modelo marginal, calculamos as funções score e as matrizes de informação de Fisher associadas aos parâmetros do modelo. Usando esses resultados, derivamos o processo iterativo score de Fisher e *backfitting* para estimar o coeficiente de regressão e as funções não paramétricas. O processo iterativo score de Fisher também foi proposto para estimar os componentes de variância e o estimador empírico de Bayes para estimar os efeitos aleatórios. Em analogia com os modelos paramétricos, sugerimos estimar o erro padrão dos estimadores usando a matriz de informação de Fisher. O critério de informação de Akaike foi proposto como uma alternativa para a seleção de modelos e os parâmetros de suavização.

Capítulo 4

Influência local baseada na função de verossimilhança penalizada

Neste capítulo concentramos nossa discussão no desenvolvimento do método de influência local para os modelos mistos aditivos semiparamétricos elípticos. Iniciamos o capítulo descrevendo alguns dos principais trabalhos relacionados à aplicação desta metodologia em modelos paramétricos. A Seção 2 descreve o método de influência local baseado na verossimilhança penalizada. A Seção 3 apresenta os resultados obtidos da derivação da curvatura normal. Especificamente, a matriz de informação observada de Fisher e a matriz de perturbações para diferentes esquemas de perturbação. No final, a Seção 4 contém uma discussão sobre os principais resultados apresentados neste capítulo.

4.1 Introdução

Uma etapa importante na modelagem estatística é verificar possíveis afastamentos das suposições estabelecidas sobre o modelo, bem como a existência de observações discrepantes com alguma interferência desproporcional sobre os resultados derivados do ajuste do modelo. Na literatura estatística essa etapa é conhecida como análise de diagnóstico. Inserido neste contexto, tem-se desenvolvido diversos procedimentos para detectar a presença de observações discrepantes. Dentre as técnicas mais usadas encontra-se a análise de resíduos e a eliminação de

casos. A análise de resíduos sugere o uso de uma inspeção gráfica dos resíduos padronizados. Eliminação de casos propõe avaliar o impacto de cada observação sobre as estimativas da regressão através da retirada individual de cada observação do conjunto de dados. Uma observação é influente se o efeito de excluí-la do conjunto de dados produz diferenças significativas na análise. Nesse sentido, a distância de Cook tem recebido uma atenção especial. A análise de influência baseada nos resíduos e eliminação de casos foi proposta, inicialmente, para o modelo de regressão paramétrico. Alguns trabalhos relacionados são os seguintes: Cox e Snell (1968) apresentam uma forma geral de definir resíduos e propõem um método para encontrar seus dois primeiros momentos; Cook (1977) propõe um importante procedimento para detectar observações influentes baseado na idéia de eliminação de casos; Belsley et al. (1980) discutem a padronização dos resíduos e apresentam algumas medidas de diagnóstico dirigidas à identificação de dados influentes; Cook e Weisberg (1982) desenvolvem algumas medidas de diagnóstico para o modelo de regressão baseadas nos resíduos e na distância de Cook; Hawkins (1980) e Rousseeuw e Leroy (1987) tratam o problema da identificação de observações aberrantes; Caroni (1987) apresenta análise de resíduos e de influência para o modelo linear multivariado; Paula e Peres (1988) discutem a eliminação de pontos em modelos lineares generalizados com parâmetros restritos; Christensen et al. (1992) desenvolvem a idéia de eliminação de casos para modelos com efeito misto; Cordeiro e Paula (1992) estendem a técnica de eliminação de casos para modelos cuja distribuição não pertence à família exponencial; Banerjee e Frees (1997) apresentam um procedimento de diagnóstico de influência baseado na eliminação de observações em modelos lineares longitudinais; Galea et al. (2000) estudam a técnica de eliminação de casos em modelos elípticos multivariado; Zhu et al. (2001) estudam a técnica de eliminação de casos para modelos com dados incompletos; e Díaz et al. (2003) desenvolvem diagnóstico de influência para o modelo de regressão linear multivariado elíptico. Veja também outros trabalhos relacionados em Gnanadesikan (1977), Polasek (1984), Chatterjee e Hadi (1988), Barnett e Lewis (1994), Peña (2005) e Tan et al. (2001).

A análise de influência baseada nos resíduos e na técnica de eliminação de casos desenvolvida para o modelo de regressão paramétrico, também tem sido estendida para os modelos de regressão não paramétrico e semiparamétrico. Alguns trabalhos relacionados são os seguintes: Eubank (1984) estuda as propriedades da matriz de alavancas no contexto de regressão não paramétrica e deriva algumas medidas de diagnóstico usando suavização tipo *spline*; Silverman (1985) apresenta uma discussão sobre o uso de resíduos em regressão não paramétrica usando suavização tipo *spline*; Eubank (1985) deriva as propriedades de algumas medidas de diagnóstico para regressão não paramétrica, baseadas nos pontos de alavanca e resíduos studentizados; Eubank e Gunst (1986) derivam algumas medidas de diagnóstico para a classe de estimadores de mínimos quadrados penalizados sob um ponto de vista bayesiano; Eubank e Thomas (1993) propõem alguns testes e gráficos de diagnóstico para detectar heteroscedasticidade em regressão não paramétrica usando suavização por *splines* (veja também Gu, 1992); Kim (1996) estuda resíduos, alavanca e algumas distâncias tipo Cook usando suavização por *spline*; Wei (2004) apresenta algumas medidas de diagnóstico de influência e robustez para suavização por *spline*; Kim et al. (2002) apresentam algumas medidas de diagnóstico de influência, como funções dos resíduos e alavancas para as estimativas do componente paramétrico e não paramétrico propostas por Speckman sobre o modelo de regressão semiparamétrico (linear parcial). Recentemente, Fung et al. (2002) apresentam um importante trabalho no qual estendem as medidas de diagnóstico baseadas nos resíduos e na distância de Cook para as estimativas de máxima verossimilhança penalizada sob o modelo linear com efeito misto semiparamétrico normal. As fórmulas obtidas para a análise de influência apresentadas por Fung et al. (2002) são expressões generalizadas das fórmulas desenvolvidas para os modelos paramétrico e não paramétrico.

Como alternativa à análise de influência baseada na eliminação de pontos, Cook (1986) desenvolve um método de diagnóstico mais inovador na área de regressão, conhecido como influência local. A idéia básica do método de influência local é avaliar a influência conjunta das observações quando pequenas perturbações são

introduzidas no modelo ou nos dados, ao invés da avaliação da influência induzida pela retirada individual ou conjunta dessas observações. Se o fato de introduzir uma pequena perturbação no modelo gera mudanças significativas nos resultados da análise, então isso é uma evidência de alguma dificuldade. Sugere-se medir a sensibilidade da análise frente a mudanças no modelo através de algum tipo de derivada. Cook propõe usar a curvatura da superfície do afastamento da verossimilhança que é essencialmente equivalente a usar a segunda derivada do afastamento da verossimilhança. Na atualidade, o método de influência local tem recebido uma atenção especial na comunidade Estatística envolvida na análise de diagnóstico de influência e tem-se intensificado o estudo desse método para diversos modelos estatísticos. Por exemplo, Paula (1993) propõe uma nova medida de influência local para modelos lineares com parâmetros restritos; Galea et al. (1997) desenvolvem a técnica de influência local para o modelo de regressão linear elíptico; Lesaffre e Verbeke (1998) nos modelos lineares mistos; Ouwens et al. (2001) nos modelos lineares generalizados mistos; Pan e Fang (2002) nos modelos de curva de crescimento; Díaz-García et al. (2003) no modelo de regressão linear multivariado elíptico; Villegas (2002) no contexto dos modelos lineares generalizados; Ibacache (2004) no modelo de regressão multivariado normal com erros aleatórios equicorrelacionados; Lu e Song (2006) no modelo com variáveis latentes proibito; Liu (2000) e Liu (2002) no modelo linear elíptico multivariado; Liu (2004) no modelo de séries de tempo heteroscedástico condicional elíptico; Galea et al. (2005) no modelo de calibração comparativa t-Student; e Osorio (2006) no modelo linear misto elíptico. Veja também outros detalhes sobre o método em Billor e Loynes (1993), Fung e Kwan (1997), Cook (1997), e Poon e Poon (1999). Por outro lado, o método de influência local tem sido também estendido para os modelos não paramétricos e semiparamétricos. Por exemplo, Thomas (1991) desenvolve a técnica de influência local para o parâmetro de suavização selecionado pelo método de validação cruzada no modelo de regressão não paramétrico e Zhu et al. (2003) estendem a análise de influência local para as estimativas de máxima verossimilhança penalizada derivadas do modelo linear parcial normal univariado.

4.2 Método de influência local

Nesta seção desenvolvemos o método de influência local baseado na função de verossimilhança penalizada. Nosso propósito é avaliar a sensibilidade das estimativas de máxima verossimilhança penalizada quando introduzimos pequenas perturbações no modelo ou nos dados. Por simplicidade, vamos assumir que o vetor de parâmetros de suavização, $\boldsymbol{\alpha}$, é fixo.

Consideremos o modelo misto aditivo semiparamétrico elíptico (2.12), em que o logaritmo da função de verossimilhança penalizada é expresso na forma

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \quad (4.1)$$

sendo $L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ a contribuição da i -ésima observação definida pela equação (3.7). Suponhamos que $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})$ é uma versão perturbada de $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ que depende do vetor de perturbações $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ de dimensão $(n \times 1)$, restrito ao subespaço Euclidiano aberto $\Omega \subset \mathbb{R}^n$, e assumimos que existe um vetor $\boldsymbol{\omega}_0$ de não perturbação que satisfaz $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}_0) = L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$. Supor ainda que $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança penalizada obtida ao maximizar $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$, e $\hat{\boldsymbol{\theta}}_\omega$ a estimativa de máxima verossimilhança penalizada obtida ao maximizar $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})$. Uma forma de comparar $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_\omega$ é medir a distância entre essas estimativas através da função de afastamento da verossimilhança, definida como

$$DV(\boldsymbol{\omega}) = 2 \left[L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) - L_p(\hat{\boldsymbol{\theta}}_\omega, \boldsymbol{\alpha}) \right] \geq 0.$$

A idéia de influência local é estudar o comportamento de $DV(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0$. O procedimento consiste em escolher uma direção unitária arbitrária, $\boldsymbol{\ell}$, e então considerar o gráfico de $DV(\boldsymbol{\omega}_0 + a \boldsymbol{\ell})$ versus a , para $a \in \mathbb{R}$. Esse gráfico é chamado de linha projetada. Note que $DV(\boldsymbol{\omega}_0 + a \boldsymbol{\ell})$ tem um mínimo local em $a = 0$, uma vez que $DV(\boldsymbol{\omega}_0) = 0$. Cada linha projetada pode ser caracterizada através da curvatura normal $C_\ell(\boldsymbol{\theta})$ em torno de $a = 0$. Cook considera a direção $\boldsymbol{\ell}_{max}$ correspondente à maior curvatura $C_{\ell_{max}}(\boldsymbol{\theta})$. Eventualmente um gráfico de índices

de ℓ_{max} pode revelar as observações que sob pequenas perturbações exercem uma notável influência sobre $DV(\boldsymbol{\omega})$.

De acordo com Cook (1986), equação (11), a curvatura normal na direção $\boldsymbol{\ell}$ é dada por

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

em que o vetor $\boldsymbol{\ell} \in \Omega$ e $\|\boldsymbol{\ell}\| = 1$, e

$$\ddot{\mathbf{L}}_p = \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

e

$$\boldsymbol{\Delta}_p = \left. \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Note que $-\ddot{\mathbf{L}}_p = -\ddot{\mathbf{L}}_p(\boldsymbol{\theta})$ denota a matriz de informação de Fisher observada penalizada e $\boldsymbol{\Delta}_p = \boldsymbol{\Delta}_p(\boldsymbol{\theta})$ é a matriz de perturbação penalizada. $C_{\boldsymbol{\ell}}(\boldsymbol{\theta})$ representa a curvatura normal sob a estimativa de $\boldsymbol{\theta}$ após perturbar o modelo $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$. Eventualmente valores grandes da curvatura $C_{\boldsymbol{\ell}}(\boldsymbol{\theta})$ podem indicar a presença de uma alta sensibilidade na estimativa induzida pelas perturbações na direção $\boldsymbol{\ell}$. De acordo com Poon e Poon (1999) podemos usar a curvatura normal conformal (curvatura invariante sob transformações uniformes de escala) definida por

$$B_{\boldsymbol{\ell}}(\boldsymbol{\theta}) = \frac{C_{\boldsymbol{\ell}}(\boldsymbol{\theta})}{2\|\boldsymbol{\ell}^T \boldsymbol{\Delta}_p^T \ddot{\mathbf{L}}_p^{-1} \boldsymbol{\Delta}_p \boldsymbol{\ell}\|_F}, \quad (4.2)$$

em que $\|\cdot\|_F$ denota a norma Frobenius definida por $\|\mathbf{A}\|_F = \{\text{tr}(\mathbf{A}^T \mathbf{A})\}^{1/2}$ para uma matriz \mathbf{A} . Essa curvatura caracteriza-se por permitir que para qualquer direção $\boldsymbol{\ell}$ verifica-se $0 \leq B_{\boldsymbol{\ell}}(\boldsymbol{\theta}) \leq 1$.

Em certas situações estamos interessados em avaliar a influência local sobre o subconjunto de parâmetros, digamos $\boldsymbol{\theta}_1$, de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$. Nesse caso, usamos

$$DV(\boldsymbol{\omega}) = 2 \left[L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) - L_p(\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}, \hat{\boldsymbol{\theta}}_2(\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}), \boldsymbol{\alpha}) \right],$$

em que $\widehat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{1\omega})$ denota a estimativa de máxima verossimilhança penalizada derivada do modelo não perturbado assumindo que $\boldsymbol{\theta}_1$ é fixo. De acordo com a partição do vetor $\boldsymbol{\theta}$, podemos particionar a matriz Hessiana na forma

$$\ddot{\mathbf{L}}_p = \begin{pmatrix} \ddot{\mathbf{L}}_{p11} & \ddot{\mathbf{L}}_{p12} \\ \ddot{\mathbf{L}}_{p21} & \ddot{\mathbf{L}}_{p22} \end{pmatrix}$$

e, conseqüentemente, podemos definir uma matriz \mathbf{G}_{11} da seguinte maneira:

$$\mathbf{G}_{11} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{p22}^{-1} \end{pmatrix}.$$

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

$$C_{\boldsymbol{\ell}}(\boldsymbol{\theta}) = -2\boldsymbol{\ell}^T \boldsymbol{\Delta}_p^T \left\{ \ddot{\mathbf{L}}_p^{-1} - \mathbf{G}_{11} \right\} \boldsymbol{\Delta}_p \boldsymbol{\ell},$$

e a direção da maior curvatura corresponde ao autovetor associado ao maior autovalor de $\mathbf{B}_1 = \boldsymbol{\Delta}_p^T \left\{ \ddot{\mathbf{L}}_p^{-1} - \mathbf{G}_{11} \right\} \boldsymbol{\Delta}_p$. Analogamente, podemos estar interessados em avaliar a influência local sobre $\boldsymbol{\theta}_2$. Neste caso, definimos uma matriz \mathbf{G}_{22} da seguinte maneira:

$$\mathbf{G}_{22} = \begin{pmatrix} \ddot{\mathbf{L}}_{p11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Portanto, a curvatura normal para $\boldsymbol{\theta}_2$ na direção $\boldsymbol{\ell}$, assume a forma

$$C_{\boldsymbol{\ell}}(\boldsymbol{\theta}) = -2\boldsymbol{\ell}^T \boldsymbol{\Delta}_p^T \left\{ \ddot{\mathbf{L}}_p^{-1} - \mathbf{G}_{22} \right\} \boldsymbol{\Delta}_p \boldsymbol{\ell},$$

e a direção da maior curvatura corresponde ao autovetor associado ao maior autovalor de $\mathbf{B}_2 = \boldsymbol{\Delta}_p^T \left\{ \ddot{\mathbf{L}}_p^{-1} - \mathbf{G}_{22} \right\} \boldsymbol{\Delta}_p$. Na prática, a análise de influência local reduz-se a encontrar o maior autovalor absoluto da matriz \mathbf{B} (\mathbf{B}_1 ou \mathbf{B}_2), $C_{\boldsymbol{\ell}_{max}}$, e seu correspondente autovetor, $\boldsymbol{\ell}_{max}$.

Finalmente, seja $\mathbf{e}_{i,n}$ um vetor $n \times 1$ cujo i -ésimo elemento é igual a 1 e os restantes elementos iguais a zero. Escobar e Meeker (1992) propuseram estudar a

curvatura normal na direção $\ell = \mathbf{e}_{i,n}$. De acordo com Escobar e Meeker (1992), essa curvatura é definida por

$$C_i = 2|\mathbf{e}_{i,n}^T \mathbf{B} \mathbf{e}_{i,n}| = 2|b_{ii}|, \quad (4.3)$$

em que $\mathbf{B} = \Delta_p^T \ddot{\mathbf{L}}_p^{-1} \Delta_p$ com b_{ii} sendo o i -ésimo elemento da diagonal principal da matriz \mathbf{B} , para $i = 1, \dots, n$. Essa medida é chamada medida de influência local total da i -ésima observação. Verbeke e Molenberghs (2000) sugerem considerar a i -ésima observação influente se $C_i > 2\bar{C}$, para $\bar{C} = \sum_{i=1}^n C_n/s$.

4.3 Derivação da curvatura

A seguir calculamos a matriz de informação observada penalizada, $-\ddot{\mathbf{L}}_p(\boldsymbol{\theta})$, e a matriz $\Delta_p(\boldsymbol{\theta})$ para diferentes esquemas de perturbação. Consideremos o vetor de parâmetros de suavização, $\boldsymbol{\alpha}$, fixo, e denotemos as primeiras e segundas derivadas parciais das matrizes de escala \mathbf{D} e \mathbf{V}_i em relação aos vetores $\boldsymbol{\lambda}$ e $\boldsymbol{\gamma}$, respectivamente, da seguinte forma: $\dot{\mathbf{D}}(l) = \partial \mathbf{D} / \partial \lambda_l$, $\ddot{\mathbf{D}}(j, l) = \partial^2 \mathbf{D}_i / \partial \lambda_j \partial \lambda_l$, $\dot{\mathbf{V}}(l) = \partial \mathbf{V}_i / \partial \gamma_l$ e $\ddot{\mathbf{V}}(j, l) = \partial^2 \mathbf{V}_i / \partial \gamma_j \partial \gamma_l$.

4.3.1 Matriz de informação

Seja $\mathbf{f} = (\mathbf{f}_1^T, \dots, \mathbf{f}_s^T)^T$. A matriz de informação observada penalizada associada ao modelo misto aditivo semiparamétrico elíptico (a obtenção deste resultado é descrito em detalhe no apêndice C) assume a forma

$$-\ddot{\mathbf{L}}_p(\boldsymbol{\theta}) = -\sum_{i=1}^n \ddot{\mathbf{L}}_{p_i}(\boldsymbol{\theta}), \quad (4.4)$$

em que

$$\ddot{\mathbf{L}}_{p_i}(\boldsymbol{\theta}) = \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \ddot{\mathbf{L}}_{p_i}^{\beta\beta} & \ddot{\mathbf{L}}_{p_i}^{\beta f} & \ddot{\mathbf{L}}_{p_i}^{\beta\lambda} & \ddot{\mathbf{L}}_{p_i}^{\beta\gamma} \\ \ddot{\mathbf{L}}_{p_i}^{f\beta} & \ddot{\mathbf{L}}_{p_i}^{ff} & \ddot{\mathbf{L}}_{p_i}^{f\lambda} & \ddot{\mathbf{L}}_{p_i}^{f\gamma} \\ \ddot{\mathbf{L}}_{p_i}^{\lambda\beta} & \ddot{\mathbf{L}}_{p_i}^{\lambda f} & \ddot{\mathbf{L}}_{p_i}^{\lambda\lambda} & \ddot{\mathbf{L}}_{p_i}^{\lambda\gamma} \\ \ddot{\mathbf{L}}_{p_i}^{\gamma\beta} & \ddot{\mathbf{L}}_{p_i}^{\gamma f} & \ddot{\mathbf{L}}_{p_i}^{\gamma\lambda} & \ddot{\mathbf{L}}_{p_i}^{\gamma\gamma} \end{pmatrix},$$

com

$$\begin{aligned}\ddot{\mathbf{L}}_{\mathbf{P}_i}^{\beta\beta} &= \left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= 2\mathbf{X}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[2W'_g(\hat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \widehat{\boldsymbol{\Sigma}}_i \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i,\end{aligned}\quad (4.5)$$

$$\ddot{\mathbf{L}}_{\mathbf{P}_i}^{ff} = \left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f} \partial \mathbf{f}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (4.6)$$

$$\left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \mathbf{f}_{k'}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{cases} 2\mathbf{N}_{ki}^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[2W'_g(\hat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \widehat{\boldsymbol{\Sigma}}_i \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{N}_{ki} - \frac{\alpha_k}{n} \mathbf{K}_k & k = k' \\ 2\mathbf{N}_{ki}^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[2W'_g(\hat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \widehat{\boldsymbol{\Sigma}}_i \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{N}_{k'i} & k \neq k' \end{cases}$$

$$\left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \mathbf{f}_{k'}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 2\mathbf{N}_{ki}^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[2W'_g(\hat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \widehat{\boldsymbol{\Sigma}}_i \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{N}_{ki} - \frac{\alpha_k}{n} \mathbf{K}_k,$$

$$\ddot{\mathbf{L}}_{\mathbf{P}_i}^{\lambda\lambda} = \left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (4.7)$$

$$\begin{aligned}\left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \lambda_j \partial \lambda_l} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \left[\dot{\mathbf{D}}(j) \mathbf{Z}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) - \ddot{\mathbf{D}}(j, l) \right] \mathbf{Z}_i^T + \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \\ &\quad \left[W'_g(\hat{\delta}_i) \dot{\mathbf{D}}(j) \mathbf{Z}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) - W_g(\hat{\delta}_i) \ddot{\mathbf{D}}(j, l) + W_g(\hat{\delta}_i) \right. \\ &\quad \left. \dot{\mathbf{D}}(j) \mathbf{Z}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) + W_g(\hat{\delta}_i) \dot{\mathbf{D}}(l) \mathbf{Z}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(j) \right] \mathbf{Z}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i,\end{aligned}$$

$$\ddot{\mathbf{L}}_{\mathbf{P}_i}^{\gamma\gamma} = \left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (4.8)$$

$$\begin{aligned}
\left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \gamma_j \partial \gamma_l} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \hat{\boldsymbol{\Sigma}}_i^{-1} \left[\dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(l) - \ddot{\mathbf{V}}(j, l) \right] \\
&+ \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[W'_g(\hat{\delta}_i) \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(l) - W_g(\hat{\delta}_i) \ddot{\mathbf{V}}(j, l) \right. \\
&+ W_g(\hat{\delta}_i) \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(l) + W_g(\hat{\delta}_i) \dot{\mathbf{V}}(l) \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \left. \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i, \\
\ddot{\mathbf{L}}_{p_i}^{\beta f} &= \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \mathbf{f}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{4.9}
\end{aligned}$$

$$\begin{aligned}
\left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \mathbf{f}_k^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= 2 \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[2 W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{N}_{ki}, \\
\ddot{\mathbf{L}}_{p_i}^{\beta \lambda} &= \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad \ddot{\mathbf{L}}_{p_i}^{\beta \gamma} = \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{4.10}
\end{aligned}$$

$$\begin{aligned}
\left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \lambda_l} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= 2 \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(j) \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i, \\
\left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \gamma_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= 2 \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i, \\
\ddot{\mathbf{L}}_{p_i}^{f \lambda} &= \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f} \partial \boldsymbol{\lambda}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad \ddot{\mathbf{L}}_{p_i}^{f \gamma} = \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f} \partial \boldsymbol{\gamma}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{4.11}
\end{aligned}$$

$$\begin{aligned}
\left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \lambda_l} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= 2 \mathbf{N}_{ki}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(j) \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i, \\
\left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \gamma_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= 2 \mathbf{N}_{ki}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i,
\end{aligned}$$

$$\ddot{\mathbf{L}}_{P_i}^{\gamma\lambda} = \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\lambda}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{e} \quad (4.12)$$

$$\begin{aligned} \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \gamma_j \partial \lambda_l} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \left[\hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T \right] \\ &+ \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[W'_g(\hat{\delta}_i) \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T + W_g(\hat{\delta}_i) \dot{\mathbf{V}}(j) \right. \\ &\left. \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T + W_g(\hat{\delta}_i) \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i, \end{aligned}$$

para $j = 1, \dots, d_\lambda$ e $l = 1, \dots, d_\gamma$, $W_g(\delta_i)$ e $W'_g(\delta_i)$ avaliadas em $\hat{\delta}_i = \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i$, com $\hat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \sum_{k=1}^s \mathbf{N}_{ki} \hat{\mathbf{f}}_k$.

4.3.2 Matriz de perturbação

Nesta seção estudaremos quatro esquemas de perturbação; especificamente, ponderação de casos, perturbação na matriz de escala, perturbação nas variáveis explicativas, e perturbação nas variáveis resposta. Detalhes sobre esses esquemas de perturbação podem ser encontrados em Zhu e Lee (2003), Osorio (2006) e Osorio et al. (2007). A matriz de perturbação penalizada associada ao modelo misto aditivo semiparamétrico elíptico assume a forma

$$\Delta_p(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \begin{pmatrix} \Delta_p^\beta(\boldsymbol{\theta}) \\ \Delta_p^{\mathbf{f}_1}(\boldsymbol{\theta}) \\ \vdots \\ \Delta_p^{\mathbf{f}_s}(\boldsymbol{\theta}) \\ \Delta_p^\lambda(\boldsymbol{\theta}) \\ \Delta_p^\gamma(\boldsymbol{\theta}) \end{pmatrix}, \quad (4.13)$$

em que $\Delta_p^\beta(\boldsymbol{\theta}) = \partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T \in \mathbb{R}^{p \times n}$, $\Delta_p^{\mathbf{f}_k}(\boldsymbol{\theta}) = \partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) / \partial \mathbf{f}_k \partial \boldsymbol{\omega}^T \in \mathbb{R}^{r_k \times n}$ ($k = 1, \dots, s$), $\Delta_p^\lambda(\boldsymbol{\theta}) = \partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) / \partial \boldsymbol{\lambda} \partial \boldsymbol{\omega}^T \in \mathbb{R}^{d_\lambda \times n}$, e $\Delta_p^\gamma(\boldsymbol{\theta}) = \partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) / \partial \boldsymbol{\gamma} \partial \boldsymbol{\omega}^T \in \mathbb{R}^{d_\gamma \times n}$, sendo os elementos de cada matriz avaliados em

$\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Recentemente, Zhu et al. (2007) propuseram correções nos esquemas de perturbação utilizados pela metodologia de influência local e mostraram que essas correções podem levar a resultados diferentes em estudos longitudinais desbalanceados. Assim, os resultados que apresentamos neste trabalho apresentam as propriedades ótimas estudadas por Zhu et al. (2007) apenas no caso balanceado.

Ponderação de casos

Através deste esquema de perturbação desejamos avaliar a contribuição individual de cada observação sob o processo de estimação. Neste caso, as contribuições individuais recebem ponderações diferentes. Seja $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ com $0 \leq \omega_i \leq 1$, o vetor de perturbação, e $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ o vetor de não perturbação. Então, o logaritmo da função de verossimilhança penalizada do modelo (2.12) perturbado é dado por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i L_i(\boldsymbol{\theta}) - \frac{1}{2} \sum_{k=1}^s \alpha_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k, \quad (4.14)$$

em que $L_i(\boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log(\delta_i)$ denota a contribuição individual da i -ésima observação no logaritmo da função de verossimilhança não penalizada. Diferenciando $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})$ em relação a $\boldsymbol{\theta}$ e ω_i , obtemos que

$$\begin{aligned} \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2W_g(\widehat{\delta}_i) \mathbf{X}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i, \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f}_k \partial \omega_i} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2W_g(\widehat{\delta}_i) \mathbf{N}_{ki}^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i, \quad k = 1, \dots, s, \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \lambda_l \partial \omega_i} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{2} \text{tr} \{ \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T \} - W_g(\widehat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i \end{aligned}$$

e

$$\left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \gamma_j \partial \omega_i} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{1}{2} \text{tr} \{ \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \} - W_g(\widehat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i.$$

O esquema de ponderação de casos generaliza a idéia de eliminação de casos, fornecendo uma boa aproximação de diagnóstico global, sem ter que reestimar os parâmetros quando uma observação é excluída do conjunto de dados. Alternativamente, podemos considerar o esquema de perturbação

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \left(L_i(\boldsymbol{\theta}) - \frac{1}{2n} \sum_{k=1}^s \alpha_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k \right), \quad (4.15)$$

e assim avaliar a contribuição individual penalizada de cada observação sobre o processo de estimação. Neste caso,

$$\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f}_k \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -2W_g(\hat{\delta}_i) \mathbf{N}_{ki}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\epsilon}}_i - \frac{\alpha_k}{2n} \mathbf{K}_k \hat{\mathbf{f}}_k, \quad k = 1, \dots, s.$$

Perturbação na matriz de escala

Este esquema de perturbação permite avaliar a influência que exercem as observações na estrutura de escala e no vetor dos componentes de variância. Seja $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ o vetor de perturbação com $\omega_i > 0$, e $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ o vetor de não perturbação. Assumindo $\omega_i^{-1} \boldsymbol{\Sigma}_i$ no lugar de $\boldsymbol{\Sigma}_i$, temos que o logaritmo da função de verossimilhança penalizada do modelo (2.12) perturbado é dado por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) = \sum_{i=1}^n L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) - \frac{1}{2} \sum_{k=1}^s \alpha_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k, \quad (4.16)$$

em que $L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \frac{1}{2} m_i \log \omega_i + \log g(\delta_{i\omega})$ representa a contribuição da i -ésima observação no modelo perturbado, com $\delta_{i\omega} = \omega_i \boldsymbol{\epsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i$. Diferenciando $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})$ em relação a $\boldsymbol{\theta}$ e ω_i , obtemos que

$$\begin{aligned} \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2 \{W'_g(\hat{\delta}_i) \hat{\delta}_i + W_g(\hat{\delta}_i)\} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\epsilon}}_i, \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f}_k \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2 \{W'_g(\hat{\delta}_i) \hat{\delta}_i + W_g(\hat{\delta}_i)\} \mathbf{N}_{ki}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\epsilon}}_i, \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \lambda_l \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\{W'_g(\hat{\delta}_i) \hat{\delta}_i + W_g(\hat{\delta}_i)\} \hat{\boldsymbol{\epsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\epsilon}}_i \end{aligned}$$

e

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \gamma_j \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\{W'_g(\hat{\delta}_i) \hat{\delta}_i + W_g(\hat{\delta}_i)\} \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i.$$

Um esquema de perturbação alternativo que permite avaliar a interferência das observações na matriz de escala dos efeitos aleatórios é considerar $\mathbf{D}_{i\omega} = \omega_i \mathbf{D}$. Dessa forma, a perturbação na matriz \mathbf{D} pode ser analisada como uma perturbação na matriz de escala $\boldsymbol{\Sigma}_i$. Isso também é válido para a matriz \mathbf{V}_i . Veja maiores detalhes em Osorio (2006).

Perturbação aditiva na variável explicativa

O fato de introduzir perturbações nas variáveis explicativas pode ter um impacto indesejável sobre as estimativas. Em particular, Fuller (1987) mostra que a presença de erros de medição nas covariáveis podem gerar uma estimativa para o coeficiente de regressão altamente viesado. Seja $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{im_i})^T$ o vetor de perturbação, e $\boldsymbol{\omega}_0 = (0, \dots, 0)^T \in \mathbb{R}^{n^*}$ o vetor de não perturbação, com $n^* = \sum_{i=1}^n m_i$. Consideremos $\mathbf{x}_{it\omega} = \mathbf{x}_{it} + \boldsymbol{\omega}_i$ a t -ésima coluna da matriz \mathbf{X}_i , para $t = 1, \dots, p$, com $\mathbf{x}_{it} \in \mathbb{R}^{m_i}$ e $\boldsymbol{\omega}_i \in \mathbb{R}^{m_i}$. A matriz de planejamento perturbada fica expressa na forma $\mathbf{X}_{i\omega} = \mathbf{X}_i + \boldsymbol{\omega}_i \mathbf{z}_t^T$, em que $\mathbf{z}_t \in \mathbb{R}^p$ denota um vetor com um 1 na t -ésima posição e zeros nas demais posições. Então, logaritmo da função de verossimilhança penalizada do modelo (2.12) perturbado é da forma

$$L_P(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) = \sum_{i=1}^n L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) - \frac{1}{2} \sum_{k=1}^s \alpha_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k,$$

em que $L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log g(\delta_{i\omega})$, com $\delta_{i\omega} = \boldsymbol{\varepsilon}_{i\omega}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_{i\omega}$, $\boldsymbol{\varepsilon}_{i\omega} = \boldsymbol{\varepsilon}_i - \mathbf{z}_t^T \boldsymbol{\beta} \boldsymbol{\omega}_i$. Diferenciando $L_P(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})$ em relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos que

$$\begin{aligned} \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= 4W'_g(\hat{\delta}_i) \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\beta}}_t - 2W_g(\hat{\delta}_i) \{\mathbf{X}_i^T \boldsymbol{\beta}_t - \mathbf{z}_t \hat{\boldsymbol{\varepsilon}}_i^T\} \hat{\boldsymbol{\Sigma}}_i^{-1}, \\ \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f}_k \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{N}_{ki}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{4W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i\} \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\beta}}_t, \end{aligned}$$

$$\left. \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \lambda_l \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = 2 \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(j) \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i\} \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\beta}}_t$$

e

$$\left. \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \gamma_j \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \{2W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i\} \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\beta}}_t,$$

em que $\hat{\boldsymbol{\beta}}_t$ denota o t -ésimo elemento do vetor $\hat{\boldsymbol{\beta}}$. Podemos considerar um esquema de perturbação no qual a matriz de planejamento \mathbf{X}_i seja substituída pela matriz perturbada $\mathbf{X}_{i\omega} = \mathbf{X}_i + \mathbf{W}_i$, assumindo que $\mathbf{W}_i = (\omega_{ij})$ é uma matriz de perturbação ($m_i \times p$). Nesse caso, a matriz de não perturbação é $\mathbf{W}_0 = \mathbf{0}$.

Perturbação na variável resposta

Através deste esquema de perturbação desejamos avaliar a sensibilidade das estimativas quando são introduzidas pequenas perturbações nos componentes de cada vetor de respostas. Seja $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{im_i})^T \in \mathbb{R}^{m_i}$ o vetor de perturbação, e $\boldsymbol{\omega}_0 = \mathbf{0}$ ($\in \mathbb{R}^{n^*}$) o vetor de não perturbação. Consideremos a perturbação do vetor de respostas na forma $\mathbf{y}_{i\omega} = \mathbf{y}_i + \boldsymbol{\omega}_i$. Então, o logaritmo da função de verossimilhança penalizada do modelo (2.12) perturbado é dado por

$$L_P(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) = \sum_{i=1}^n L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) - \frac{1}{2} \sum_{k=1}^s \alpha_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k,$$

em que $L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log g(\delta_{i\omega})$, com $\delta_{i\omega} = \boldsymbol{\varepsilon}_{i\omega}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_{i\omega}$, $\boldsymbol{\varepsilon}_{i\omega} = \mathbf{y}_{i\omega} - \boldsymbol{\mu}_i$. Diferenciando $L_P(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})$ em relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos que

$$\begin{aligned} \left. \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{4W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i\} \hat{\boldsymbol{\Sigma}}_i^{-1}, \\ \left. \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f}_k \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\mathbf{N}_{ki}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{4W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i\} \hat{\boldsymbol{\Sigma}}_i^{-1}, \\ \left. \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \lambda_l \partial \boldsymbol{\omega}_i^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(l) \mathbf{Z}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{2W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i\} \hat{\boldsymbol{\Sigma}}_i^{-1} \end{aligned}$$

e

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \gamma_j \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \{2W'_g(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W_g(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i\} \hat{\boldsymbol{\Sigma}}_i^{-1}.$$

4.4 Conclusões do capítulo

Neste capítulo foi desenvolvido o método de influência local para os modelos mistos aditivos semiparamétricos elípticos baseados na função de verossimilhança penalizada. Usando resultados de diferenciação de matrizes, derivamos a matriz de informação observada de Fisher e as matrizes de perturbações. Os esquemas de perturbação abordados neste capítulo foram ponderação de casos, perturbação na matriz de escala, perturbação nas variáveis explicativas, e perturbação nas variáveis respostas. Todas essas matrizes assumem expressões matriciais que tornam relativamente simples a implementação computacional do método de influência local para os modelos mistos aditivos semiparamétricos elípticos, em qualquer software que permita a manipulação de matrizes.

Capítulo 5

Modelo linear parcial t-Student univariado

Este capítulo descreve o modelo linear parcial ou semiparamétrico t-elíptico, que é um caso particular do modelo MMASE. Iniciamos nossa descrição fornecendo alguns dos principais trabalhos relacionados com o modelo linear parcial e sua aplicabilidade em diferentes áreas de pesquisa. Em seguida, especificamos o modelo e derivamos a função escore e a matriz de informação de Fisher baseados na verossimilhança penalizada. Logo, apresentamos o vetor de resíduos padronizados e o método de validação cruzada generalizada. Finalmente, particularizamos o método de influência local descrito no Capítulo 4 para esta classe de modelos.

5.1 Introdução

O modelo linear parcial ou semiparamétrico (MLP) é uma extensão do modelo linear clássico e caracteriza-se por incluir, além de um componente paramétrico explicando a variável resposta, um componente não paramétrico associado a alguma covariável. Esses modelos têm sido aplicados em diversas áreas. No contexto de estudos longitudinais, por exemplo, tem-se mostrado que os níveis de uma variável resposta dependem do tempo em forma não paramétrica. Alguns trabalhos relacionados com os modelos MLPs são os seguintes. Heckman (1986) prova a consistência e normalidade assintótica do estimador do coeficiente de regressão e mostra que o viés é assintoticamente desprezível. Além disso, Heckman

mostra que o estimador do coeficiente de regressão e da função não paramétrica são estimadores de Bayes quando assumimos uma distribuição a priori apropriada, e apresenta uma aplicação no contexto de análise de variância para um esquema de amostragem balanceado. Rice (1986) estuda alguns critérios de convergência e mostra que o viés do estimador do coeficiente de regressão pode dominar, assintoticamente, a variância sob um esquema de amostragem desbalanceado. Green (1987) estuda o comportamento assintótico dos estimadores de máxima verossimilhança penalizada e fornece definições apropriadas para a função desvio, os graus de liberdade, e os resíduos. Além disso, apresenta aproximações quadráticas para todas as estatísticas propostas; veja também Green (1985). Speckman (1988) compara o viés e a variância dos estimadores do modelo MLP e propõe um novo estimador que apresenta um viés assintoticamente de ordem menor; veja também Shiau e Wahba (1988). Heckman (1988) deriva dois estimadores minimax para o coeficiente de regressão e mostra que cada um desses estimadores apresenta um erro quadrático médio igual a n^{-1} quando as covariáveis estão altamente correlacionadas (não necessariamente de forma exata). Pitrun et al. (2006) desenvolvem alguns testes de hipóteses para provar a não linearidade no modelo MLP. Bianco et al. (2006) estudam o problema de teste de hipóteses para o coeficiente de regressão e analisam sua distribuição assintótica. Ma et al. (2006) estudam os modelos MLPs heteroscedásticos e propõem uma família de estimadores consistentes. Além disso, eles estudam suas propriedades assintóticas. Liang (2006) estuda alguns aspectos inferenciais sob o modelo MLP e propõe dois testes para avaliar a linearidade do componente não paramétrico. Com relação ao desenvolvimento de métodos de diagnóstico, Thomas (1991) propõe uma medida de diagnóstico de influência local para a estimativa do parâmetro de suavização em regressão *spline*. Kim et al. (2002) apresentam algumas medidas de influência baseadas nos resíduos e pontos de alavanca para detectar observações influentes nas estimativas do componente paramétrico e não paramétrico sugeridas por Speckman. E, recentemente, Zhu et al. (2003) estendem a análise de influência local para avaliar a sensibilidade das estimativas de máxima verossimilhança penalizada derivadas do MLP gaussiano.

5.2 Especificação do modelo

Seja y_i ($i = 1, \dots, n$) a resposta associada à i -ésima unidade experimental. O modelo linear parcial (MLP) univariado assume a forma

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i) + \epsilon_i, \quad (5.1)$$

em que \mathbf{x}_i é um vetor ($p \times 1$) de variáveis explicativas, $\boldsymbol{\beta}$ é um vetor ($p \times 1$) de parâmetros desconhecidos, $f(t)$ é uma função suave duas vezes diferenciável que depende da covariável t , e $\epsilon_1, \dots, \epsilon_n$ são erros aleatórios independentes. Em termos matriciais, o modelo linear parcial pode ser expresso na forma

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \boldsymbol{\epsilon}, \quad (5.2)$$

em que $\mathbf{y} = (y_1, \dots, y_n)^T$ é o vetor de respostas ($n \times 1$), \mathbf{X} é uma matriz de planejamento ($n \times p$) cuja i -ésima linha é \mathbf{x}_i^T , \mathbf{N} é uma matriz de incidência ($n \times r$) cujo elemento (j, ℓ) -ésimo é igual à função indicadora $I(t_i) = t_{\ell}^0$ ($i = 1, \dots, n$; $\ell = 1, \dots, r$) com t_1^0, \dots, t_r^0 sendo os valores ordenados e distintos de t_i , $\mathbf{f} = (f(t_1^0), \dots, f(t_r^0))^T$ e $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ é um vetor de erros aleatórios ($n \times 1$).

Assumindo que ϵ_i segue uma distribuição t-Student univariada com parâmetro de posição 0, parâmetro de escala ϕ , e graus de liberdade ν_i , temos que

$$y_i \sim t_1(\mu_i, \phi, \nu_i), \quad (5.3)$$

cuja função densidade é dada por

$$f_y(y_i) = \frac{\phi^{-1/2} \Gamma(\frac{1+\nu_i}{2})}{(\pi \nu_i)^{1/2} \Gamma(\frac{\nu_i}{2})} (1 + \nu_i^{-1} \phi^{-1} \delta_i)^{-\left(\frac{1+\nu_i}{2}\right)},$$

em que $\Gamma(\cdot)$ denota a função gama, $\delta_i = \phi^{-1}(y_i - \mu_i)^2$, $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{n}_i^T \mathbf{f}$, e \mathbf{n}_i^T é a i -ésima linha da matriz de incidência \mathbf{N} ($i = 1, \dots, n$). De acordo com Lange et al. (1989), temos que

$$\begin{aligned} E(y_i) &= \mu_i \quad (\nu_i > 0) \quad \text{e} \\ \text{Var}(y_i) &= \frac{\nu_i}{\nu_i - 2} \phi \quad (\nu_i > 2). \end{aligned}$$

Por simplicidade, vamos assumir que $\nu_i = \nu$ ($i = 1, \dots, n$) é fixo. Dessa forma, o vetor de parâmetros a ser estimado no modelo linear parcial t-Student univariado será

$$\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \phi)^T,$$

cujo espaço paramétrico associado é dado por

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{p^*} \mid \boldsymbol{\beta} \in \Theta_{\beta}, \mathbf{f} \in \Theta_{\mathbf{f}}, \phi \in \Theta_{\phi} \right\},$$

em que $\Theta_{\beta} \subseteq \mathbb{R}^p$ e $\Theta_{\mathbf{f}} \subseteq \mathbb{R}^r$ denotam, respectivamente, os espaços associados aos componente paramétrico e não paramétrico; $\Theta_{\phi} = \mathbb{R}^+$; e $p^* = p + r + 1$. Logo, o logaritmo da função de verossimilhança de $\boldsymbol{\theta}$ pode ser expresso na forma

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n L_i(\boldsymbol{\theta}), \quad (5.4)$$

em que

$$L_i(\boldsymbol{\theta}) = \log \left\{ \frac{\Gamma(\frac{\nu+1}{2})}{(\phi\nu)^{1/2} \Gamma(\frac{\nu}{2})} \right\} - \frac{1}{2} \log \phi - \left(\frac{\nu+1}{2} \right) \log \left\{ 1 + \frac{\delta_i}{\nu} \right\}. \quad (5.5)$$

Como mencionamos no Capítulo 3, a maximização direta de (5.4) sem estabelecer restrições para a função f pode fazer com que $\boldsymbol{\beta}$ seja não identificável. Para fazer frente a essa dificuldade, consideramos um procedimento alternativo que consiste em incorporar uma função de penalidade do tipo (3.5) no logaritmo da função de verossimilhança $L(\boldsymbol{\theta})$. Dessa forma, o estimador da função f conduz a uma *spline* cúbica natural com nós nos pontos t_1^0, \dots, t_r^0 . Usando o mesmo critério (veja Seção 3.2.2) temos que o logaritmo da função de verossimilhança pe-

nalizada associado ao modelo linear parcial t-Student univariado pode ser expresso na forma

$$L_p(\boldsymbol{\theta}, \alpha) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \alpha), \quad (5.6)$$

em que

$$\begin{aligned} L_{p_i}(\boldsymbol{\theta}, \alpha) &= L_i(\boldsymbol{\theta}) - \frac{\alpha}{2n} \int_a^b [f^{(2)}(t)]^2 dt \\ &= L_i(\boldsymbol{\theta}) - \frac{\alpha}{2n} \mathbf{f}^T \mathbf{K} \mathbf{f}, \end{aligned} \quad (5.7)$$

$f_k^{(2)}(t) = \frac{d^2}{dt^2} f(t)$ com $t \in [a, b]$, f é uma função que pertence ao espaço de funções de Sovolev definido em (3.2.1), α é o parâmetro de suavização que regula a qualidade do ajuste e a ondulação da estimativa de f , e \mathbf{K} é a matriz de suavização positiva definida dada pela equação (2.3) de Green e Silverman (1994). Dessa forma, o valor de $\boldsymbol{\theta}$ que maximiza $L_p(\boldsymbol{\theta}, \alpha)$, em todo o espaço paramétrico Θ , digamos $\hat{\boldsymbol{\theta}}$, é chamado de estimador de máxima verossimilhança penalizada (EMVP) de $\boldsymbol{\theta}$, e satisfaz

$$L_p(\hat{\boldsymbol{\theta}}, \alpha) \geq \sup_{\boldsymbol{\theta} \in \Theta} L_p(\boldsymbol{\theta}, \alpha).$$

Na seguinte seção derivamos o procedimento para estimar o vetor de parâmetros $\boldsymbol{\theta}$ associado ao modelo linear parcial t-Student univariado.

5.3 Processo de estimação

O problema de estimação do coeficiente de regressão, $\boldsymbol{\beta}$, e da função não paramétrica, f , sob o modelo linear parcial, tem sido discutido por vários autores. Por exemplo, Heckman (1986) estima o coeficiente de regressão e a função não paramétrica no modelo MLP baseado no critério da verossimilhança penalizada, e mostra que o estimador da função não paramétrica é uma *spline* cúbica natural; veja também Engle et al. (1986) e Rice (1986). Green (1987) estima os efeitos paramétricos e os efeitos não paramétricos do modelo MLP baseado no critério da verossimilhança penalizada, e sugere resolver as equações de estimação através do

processo iterativo de Newton-Raphson ou escore de Fisher. Speckman (1988) apresenta um método para estimar o coeficiente de regressão e a função não paramétrica no modelo MLP baseado em um procedimento de suavização tipo *kernel*. Robinson (1988) estuda o problema de estimação do coeficiente de regressão e observa que esse estimador é inconsistente quando a função não paramétrica não é parametrizada corretamente. Além disso, Robinson propõe um estimador de mínimos quadrados \sqrt{n} -consistente para o coeficiente de regressão; veja também Opsomer e Ruppert (1999). He e Shi (1996) consideram o problema de estimação sob o modelo MLP baseado em aproximações *B-spline*; veja também He et al. (2005) no contexto dos modelos MLPs generalizados para dados longitudinais. Hamilton e Truong (1997) consideram o problema de estimação sob o modelo MLP baseado em técnicas de ajuste polinomial local. No contexto dos modelos MLPs para dados longitudinais, He et al. (2002) aproximam a função não paramétrica usando regressão *spline*, e estimam o parâmetro de regressão e o coeficiente *spline* através de estimadores M. E, recentemente, Gannaz (2007) desenvolve um procedimento de estimação baseado em uma expansão tipo *wavelet* da função não paramétrica do modelo MLP gaussiano; veja também Chang e Qu (2004).

5.3.1 Função escore

Assumindo que o logaritmo da função de verossimilhança (5.6) é regular e que todas as derivadas parciais com relação a $\boldsymbol{\beta}$, \mathbf{f} e ϕ existem, temos que a função escore penalizada para o vetor de parâmetros $\boldsymbol{\theta}$ é dada por

$$\begin{aligned} \mathbf{U}_p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta}} \\ &= \begin{pmatrix} \mathbf{U}_p^\beta(\boldsymbol{\theta}) \\ \mathbf{U}_p^f(\boldsymbol{\theta}) \\ \mathbf{U}_p^\phi(\boldsymbol{\theta}) \end{pmatrix}, \end{aligned} \quad (5.8)$$

em que

$$\begin{aligned}
\mathbf{U}_p^\beta(\boldsymbol{\theta}) &= \frac{1}{\phi} \mathbf{X}^T \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}), \\
\mathbf{U}_p^f(\boldsymbol{\theta}) &= \frac{1}{\phi} \mathbf{N}^T \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) - \alpha \mathbf{K} \mathbf{f} \quad \text{e} \\
\mathbf{U}_p^\phi(\boldsymbol{\theta}) &= -\frac{n}{2\phi} + \frac{1}{2\phi^2} (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}),
\end{aligned} \tag{5.9}$$

com $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f}$, $\mathbf{W} = \text{diag}\{v_1, v_2, \dots, v_n\}$, e $v_i = v_i(\boldsymbol{\theta}) = (1 + \nu)/(\nu + \delta_i)$. Como os pesos v_i são inversamente proporcionais às distâncias δ_i , as observações com valores grandes para δ_i tendem a receber pesos pequenos no processo de estimação. Portanto, as estimativas de máxima verossimilhança penalizada sob o modelo MLP t-Student univariado são menos sensíveis a observações aberrantes em relação ao modelo normal univariado em que $v_i(\boldsymbol{\theta}) = 1$ ($i = 1, \dots, n$).

5.3.2 Matriz de informação de Fisher

Particularizando os resultados apresentados na equação (3.17), temos que a matriz de informação de Fisher ($p^* \times p^*$) para o modelo linear parcial t-Student assume a seguinte forma bloco diagonal:

$$\mathcal{I}_p(\boldsymbol{\theta}) = -\mathbb{E} \left\{ \frac{\partial^2 L_p(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} = \begin{pmatrix} \mathcal{I}_p^{\beta f}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_p^{\phi\phi}(\boldsymbol{\theta}) \end{pmatrix}, \tag{5.10}$$

em que

$$\begin{aligned}
\mathcal{I}_p^{\beta f}(\boldsymbol{\theta}) &= \frac{1}{\phi} \begin{pmatrix} \mathbf{X}^T \mathbf{W}^* \mathbf{X} & \mathbf{X}^T \mathbf{W}^* \mathbf{N} \\ \mathbf{N}^T \mathbf{W}^* \mathbf{X} & \mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \phi \mathbf{K} \end{pmatrix} \quad \text{e} \\
\mathcal{I}_p^{\phi\phi}(\boldsymbol{\theta}) &= \frac{n}{4\phi^2} (3c_\nu - 1),
\end{aligned}$$

com $\mathbf{W}^* = c_\nu \mathbf{I}_n$ e $c_\nu = (\nu + 1)/(\nu + 3)$. Pode-se mostrar que o vetor de parâmetros $(\boldsymbol{\beta}, \mathbf{f})$ é ortogonal ao parâmetro de escala ϕ ; vide Lange et al. (1989).

5.3.3 Processo iterativo

Suponhamos que os parâmetros α e ϕ são fixos. De acordo com Green (1987) e os resultados apresentados na Subseção 3.5.1, as equações de estimação

$$\mathbf{U}_p^\beta(\boldsymbol{\theta}) = \mathbf{0} \quad \text{e} \quad \mathbf{U}_p^f(\boldsymbol{\theta}) = \mathbf{0},$$

podem ser resolvidas através do seguinte algoritmo score de Fisher:

$$\begin{pmatrix} \mathbf{X}^T \mathbf{W}^* \mathbf{X} & \mathbf{X}^T \mathbf{W}^* \mathbf{N} \\ \mathbf{N}^T \mathbf{W}^* \mathbf{X} & \mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \phi \mathbf{K} \end{pmatrix}^{(u)} \begin{pmatrix} \boldsymbol{\beta}^{(u+1)} - \boldsymbol{\beta}^{(u)} \\ \mathbf{f}^{(u+1)} - \mathbf{f}^{(u)} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_p^\beta(\boldsymbol{\theta}) \\ \mathbf{U}_p^f(\boldsymbol{\theta}) \end{pmatrix}^{(u)}.$$

Após algumas manipulações algébricas obtemos

$$\begin{pmatrix} \boldsymbol{\beta}^{(u+1)} \\ \mathbf{f}^{(u+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0 \{ \mathbf{r}_{\beta-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \\ \mathbf{S}_1 \{ \mathbf{r}_{f-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \end{pmatrix}, \quad (5.11)$$

em que

$$\mathbf{S}_k = \begin{cases} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^* & k = 0 \\ (\mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}^* & k = 1, \end{cases}$$

são matrizes de suavização que geram as estimativas de $\boldsymbol{\beta}$ e f ,

$$\mathbf{r}_{\vartheta-}^{(u,u+1)} = \begin{cases} (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{N} \mathbf{f}^{(u+1)} & \vartheta = \beta \\ (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(u+1)} & \vartheta = f, \end{cases}$$

são os resíduos parciais que permitem ajustar os componentes paramétrico e não paramétrico, respectivamente. Note que,

$$\begin{aligned} \mu_i^{(u)} &= \mathbf{x}_i^T \boldsymbol{\beta}^{(u)} + \mathbf{n}_i^T \mathbf{f}^{(u)}, \\ \boldsymbol{\mu}^{(u)} &= \mathbf{X} \boldsymbol{\beta}^{(u)} + \mathbf{N} \mathbf{f}^{(u)}, \\ \widetilde{\mathbf{W}}^{(u)} &= \mathbf{I}_n - \mathbf{W}^{*-1} \mathbf{W}^{(u)} \quad \text{e} \end{aligned}$$

$$\mathbf{W}^{(u)} = \text{diag}\{v_1^{(u)}, v_2^{(u)}, \dots, v_n^{(u)}\},$$

em que $v_i^{(u)} = (1 + \nu)/(\nu + \delta_i^{(u)})$ e $\delta_i^{(u)} = \phi^{(u)-1}(y_i - \mu_i^{(u)})^2$. Para resolver (5.11) sugerimos usar o algoritmo *backfitting* descrito na Tabela 5.1.

Tabela 5.1: $(u^* + 1)$ -ésima etapa do algoritmo *backfitting* na $(u + 1)$ -ésima etapa do processo iterativo escore de Fisher sob o modelo linear parcial t-Student univariado.

-
-
- (i) Inicie o processo iterativo com $\boldsymbol{\beta}^{(u)} = \boldsymbol{\beta}^{(u,0)}$ e $\mathbf{f}^{(u)} = \mathbf{f}^{(0,0)}$,
- (ii) Para $u, u^* = 0, 1, 2, \dots$ calcular
- (ii') $\mathbf{r}_{\beta_-}^{(u, u^*)} = (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{N} \mathbf{f}^{(u, u^*)}$ e
- $$\boldsymbol{\beta}^{(u+1, u^*+1)} = \mathbf{S}_0 \{ \mathbf{r}_{\beta_-}^{(u, u^*)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} .$$
- (ii'') $\mathbf{r}_{\mathbf{f}_-}^{(u, u^*)} = (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(u+1, u^*+1)}$ e
- $$\mathbf{f}^{(u+1, u^*+1)} = \mathbf{S}_1 \{ \mathbf{r}_{\mathbf{f}_-}^{(u, u^*)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} .$$
- (iii) Repita (i) e (ii) até atingir a convergência desejada .
-
-

Por outro lado, podemos resolver a equação de estimação $U_p^\phi(\boldsymbol{\theta}) = 0$ através do seguinte processo iterativo:

$$\phi^{(u+1)} = \frac{1}{n} Q_v(\boldsymbol{\beta}^{(u)}, \mathbf{f}^{(u)}), \quad (5.12)$$

em que $Q_v(\boldsymbol{\beta}^{(u)}, \mathbf{f}^{(u)}) = (\mathbf{y} - \boldsymbol{\mu}^{(u)})^T \mathbf{W}^{(u)} (\mathbf{y} - \boldsymbol{\mu}^{(u)})$. Assim, o processo iterativo conjunto para obter a estimativa de máxima verossimilhança penalizada de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \phi)^T$, baseado no algoritmo escore de Fisher, é dado por (5.11)-(5.12). O processo deve ser iniciado com valores $\boldsymbol{\beta}^{(0)}$, $\mathbf{f}^{(0)}$ e $\phi^{(0)}$, e deve ser repetido até a convergência. Para iniciar o processo podemos considerar as estimativas do modelo normal.

5.3.4 Erro padrão

Baseados nos argumentos expostos na Subseção (3.5.5), o erro padrão estimado de $\hat{\boldsymbol{\theta}}$ pode ser calculado usando a inversa da matriz de informação de Fisher penalizada definida na equação (5.10), isto é,

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) \approx \mathcal{I}_p^{-1}(\hat{\boldsymbol{\theta}}). \quad (5.13)$$

A partir de (5.10) podemos mostrar, após algumas manipulações algébricas, que a inversa da matriz de informação de Fisher é dada por

$$\mathcal{I}_p^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_p^{\beta f^{-1}}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_p^{\phi \phi^{-1}}(\boldsymbol{\theta}) \end{pmatrix},$$

em que

$$\mathcal{I}_p^{\beta f^{-1}}(\boldsymbol{\theta}) = \begin{pmatrix} (\mathbf{X}^T \mathcal{W}_x \mathbf{X})^{-1} & -\mathbf{E} \\ -\mathbf{E}^T & (\mathbf{N}^T \mathcal{W}_f \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \end{pmatrix}, \quad (5.14)$$

com

$$\begin{aligned} \mathbf{E} &= (\mathbf{X}^T \mathcal{W}_x \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^* \mathbf{N}) (\mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \phi \mathbf{K})^{-1}, \\ \mathcal{W}_x &= \mathbf{W}^* - \mathbf{W}^* \mathbf{N} (\mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}^* \quad e \\ \mathcal{W}_f &= \mathbf{W}^* - \mathbf{W}^* \mathbf{X} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*. \end{aligned}$$

De (5.14) segue que as matrizes de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{f}}$ assumem a forma

$$\begin{aligned} \text{Cov}_A(\hat{\boldsymbol{\beta}}) &\approx (\mathbf{X}^T \mathcal{W}_x \mathbf{X})^{-1} \quad e \\ \text{Cov}_A(\hat{\mathbf{f}}) &\approx (\mathbf{N}^T \mathcal{W}_f \mathbf{N} + \alpha \phi \mathbf{K})^{-1}. \end{aligned}$$

Na seguinte seção calculamos o vetor de resíduos e sua correspondente versão padronizada sob o modelo linear parcial t-Student univariado.

5.4 Análise de resíduos

A análise de diagnóstico baseada nos resíduos tem sido o procedimento mais utilizado para avaliar o ajuste de um modelo e detectar observações aberrantes e potencialmente influentes.

Suponhamos que o parâmetro de suavização, α , o parâmetro de escala, ϕ , e os pesos $v_i(\boldsymbol{\theta}) = (\nu + m_i)/(\nu + \delta_i)$ são fixos ($i = 1, \dots, n$). A partir de (5.9), podemos obter as seguintes equações:

$$\begin{aligned} (\mathbf{X}^T \mathbf{W} \mathbf{X}) \boldsymbol{\beta} &= \mathbf{X}^T \mathbf{W} (\mathbf{y} - \mathbf{N} \mathbf{f}) \\ (\mathbf{N}^T \mathbf{W} \mathbf{N} + \alpha \phi \mathbf{K}) \mathbf{f} &= \mathbf{N}^T \mathbf{W} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}). \end{aligned}$$

Logo, após alguma álgebra, temos que as estimativas de máxima verossimilhança penalizada de $\boldsymbol{\beta}$ e \mathbf{f} assumem a forma

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W}_x \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_x \mathbf{y} \quad \text{e} \quad (5.15)$$

$$\hat{\mathbf{f}} = (\mathbf{N}^T \mathbf{W}_f \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}_f \mathbf{y}, \quad (5.16)$$

em que

$$\mathbf{W}_x = \mathbf{W} - \mathbf{W} \mathbf{N} (\mathbf{N}^T \mathbf{W} \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W} \quad \text{e}$$

$$\mathbf{W}_f = \mathbf{W} - \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}.$$

Agora, substituindo (5.15) e (5.16) em (5.2), e após alguma álgebra, temos que o vetor de valores ajustados $\hat{\mathbf{y}}$ é dado por

$$\begin{aligned} \hat{\mathbf{y}} &= \widehat{\mathbf{E}(\mathbf{y})} \\ &= \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{N} \hat{\mathbf{f}} \\ &= \bar{\mathbf{H}} \mathbf{y}, \end{aligned} \quad (5.17)$$

em que

$$\bar{\mathbf{H}} = (\mathbf{X} \mathbf{N}) \mathbf{C}^{-1} \begin{pmatrix} \mathbf{X}^T \\ \mathbf{N}^T \end{pmatrix} \mathbf{W}, \quad (5.18)$$

com

$$\mathbf{C} = \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{N} \\ \mathbf{N}^T \mathbf{W} \mathbf{X} & \mathbf{N}^T \mathbf{W} \mathbf{N} + \alpha \phi \mathbf{K} \end{pmatrix}.$$

Em analogia com os modelos lineares os elementos diagonais da matriz $\bar{\mathbf{H}}$ são chamados pontos de alavanca e são comumente utilizados na construção de técnicas de diagnóstico. As propriedades dessa matriz para regressão não paramétrica são discutidas por Eubank (1984). Logo, a partir de (5.17), o vetor de resíduos usual pode ser expresso na forma

$$\begin{aligned} \hat{\mathbf{e}} &= \mathbf{y} - \widehat{\mathbf{E}}(\mathbf{y}) \\ &= \mathbf{y} - \hat{\mathbf{y}} \\ &= (\mathbf{I}_n - \bar{\mathbf{H}})\mathbf{y}. \end{aligned} \quad (5.19)$$

Além disso, cálculos diretos mostram que a matriz de variância-covariância de $\hat{\mathbf{e}}$ é dada por

$$\text{Cov}(\hat{\mathbf{e}}) = (\mathbf{I}_n - \bar{\mathbf{H}}) \text{Cov}(\mathbf{y}) (\mathbf{I}_n - \bar{\mathbf{H}})^T,$$

em que $\text{Cov}(\mathbf{y}) = \kappa \phi^{-1} \mathbf{I}_n$, com $\kappa = \nu / (\nu - 2)$. Assim, se substituirmos ϕ e $v_i(\boldsymbol{\theta})$ pelas suas correspondentes estimativas, isto é, $\hat{\phi}$ e $v_i(\hat{\boldsymbol{\theta}})$, temos uma aproximação da matriz de variância-covariância de $\hat{\mathbf{e}}$ e conseqüentemente uma versão padronizada desses resíduos que assume a forma

$$\hat{\mathbf{e}}_i^* = \frac{\hat{e}_i}{\sqrt{\widehat{\text{Var}}(\hat{e}_i)}}, \quad (5.20)$$

em que $\widehat{\text{Var}}(\hat{e}_i)$ é o i -ésimo elemento da diagonal principal da matriz $\widehat{\text{Cov}}(\hat{\mathbf{e}})$ e denota a variância da i -ésima observação.

Maiores detalhes relacionados à análise de resíduos no contexto dos modelos lineares parciais para dados independentes podem ser encontrados em Green e Silverman (1994), Kim et al. (2002), e Zhu et al. (2003).

5.5 Validação cruzada generalizada

Como foi dito na Subseção 3.2.1, o parâmetro de suavização regula a relação entre a fidelidade dos dados e a suavidade da curva estimada, e portanto a escolha de um valor apropriado para esse parâmetro é um problema essencial que deve ser considerado. Baseados em Green e Silverman (1994), Seção 4.4, podemos definir o escore de validação cruzada generalizada sob o modelo linear parcial t-Student na forma

$$\text{VCG}(\alpha) = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\left(1 - n^{-1} \text{tr}\{\bar{\mathbf{H}}(\alpha)\}\right)^2}, \quad (5.21)$$

em que o vetor de valores ajustados $\hat{\mathbf{y}}$ é definido na equação (5.17) e a matriz $\bar{\mathbf{H}}(\alpha) = \bar{\mathbf{H}}$ na equação (5.18). Consequentemente, escolhemos o valor de α que minimize o escore $\text{VCG}(\alpha)$. Embora o critério de validação cruzada generalizada seja um procedimento eficiente para escolher α sob o modelo t-Student univariado, o mesmo está condicionado ao fato de que a matriz de escala ϕ e os pesos $v_i(\boldsymbol{\theta})$ são fixos. Na prática, o parâmetro ϕ e os pesos $v_i(\boldsymbol{\theta})$ devem ser substituídos pelas correspondentes estimativas.

5.6 Diagnóstico de influência local

Nesta seção calculamos a matriz de informação observada penalizada $-\ddot{\mathbf{L}}_p(\boldsymbol{\theta})$ e a matriz de perturbações penalizada $\Delta_p(\boldsymbol{\theta})$, ambas matrizes definidas por (4.4) e (4.13), respectivamente. Os esquemas de perturbação aqui considerados são ponderação de casos, perturbação do parâmetro de escala e perturbação aditiva na covariável associada ao componente paramétrico do modelo.

5.6.1 Matriz de informação observada

Desde (5.6) temos que o logaritmo da função de verossimilhança penalizada é dado por

$$L_p(\boldsymbol{\theta}, \alpha) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \alpha),$$

em que a função $L_{p_i}(\boldsymbol{\theta}, \alpha)$ é definida na equação (5.5). A matriz de informação observada ($p^* \times p^*$) para o modelo linear parcial t-Student univariado assume a forma

$$-\ddot{\mathbf{L}}_p(\boldsymbol{\theta}) = -\frac{\partial^2 L_p(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = - \begin{pmatrix} \ddot{\mathbf{L}}_{p_{\beta\beta}} & \ddot{\mathbf{L}}_{p_{\beta f}} & \ddot{\mathbf{L}}_{p_{\beta\phi}} \\ \ddot{\mathbf{L}}_{p_{f\beta}}^T & \ddot{\mathbf{L}}_{p_{ff}} & \ddot{\mathbf{L}}_{p_{f\phi}} \\ \ddot{\mathbf{L}}_{p_{\phi\beta}}^T & \ddot{\mathbf{L}}_{p_{\phi f}}^T & \ddot{\mathbf{L}}_{p_{\phi\phi}} \end{pmatrix},$$

em que

$$\begin{aligned} \ddot{\mathbf{L}}_{p_{\beta\beta}}(\boldsymbol{\theta}) &= -\frac{1}{\phi} \mathbf{X}^T \mathbf{D}(\mathbf{a}) \mathbf{X}, \\ \ddot{\mathbf{L}}_{p_{\beta f}}(\boldsymbol{\theta}) &= -\frac{1}{\phi} \mathbf{X}^T \mathbf{D}(\mathbf{a}) \mathbf{N}, \\ \ddot{\mathbf{L}}_{p_{\beta\phi}}(\boldsymbol{\theta}) &= \frac{2}{\phi^2} \mathbf{X}^T \mathbf{b}, \\ \ddot{\mathbf{L}}_{p_{ff}}(\boldsymbol{\theta}) &= -\frac{1}{\phi} \mathbf{N}^T \mathbf{D}(\mathbf{a}) \mathbf{N}, \\ \ddot{\mathbf{L}}_{p_{f\phi}}(\boldsymbol{\theta}) &= \frac{2}{\phi^2} \mathbf{N}^T \mathbf{b} \quad \text{e} \\ \ddot{\mathbf{L}}_{p_{\phi\phi}}(\boldsymbol{\theta}) &= \frac{1}{\phi^2} \left\{ \frac{n}{2} + \boldsymbol{\delta}^T \mathbf{D}(\mathbf{c}) \boldsymbol{\delta} - \frac{1}{\phi} \mathbf{e}^T \mathbf{D}(\mathbf{v}) \mathbf{e} \right\}, \end{aligned}$$

com $\mathbf{D}(\mathbf{a}) = \text{diag}(a_1, \dots, a_n)$, $\mathbf{b} = (b_1, \dots, b_n)^T$, $\mathbf{D}(\mathbf{c}) = \text{diag}(c_1, \dots, c_n)$, $a_i = \{v_i - 4W'(\delta_i) \delta_i\}$, $b_i = \{W(\delta_i) + W'(\delta_i) \delta_i\} \epsilon_i$, $c_i = W'(\delta_i)$, $\epsilon_i = (y_i - \mu_i)$,

$$\begin{aligned} W(\delta_i) &= -\frac{1}{2} \left\{ \frac{\nu + 1}{\nu + \delta_i} \right\} \quad \text{e} \\ W'(\delta_i) &= \frac{1}{2} \left\{ \frac{\nu + 1}{(\nu + \delta_i)^2} \right\}, \quad i = 1, \dots, n. \end{aligned}$$

5.6.2 Matriz de perturbação

A matriz $\Delta_p(\boldsymbol{\theta})$ ($p^* \times n$) para cada esquema de perturbação é definida na forma

$$\Delta_p(\boldsymbol{\theta}) = \left. \frac{\partial^2 L_p(\boldsymbol{\theta}; \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \quad (5.22)$$

em que $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança penalizada e $\boldsymbol{\omega}_0$ o vetor de não perturbação. A seguir são apresentadas as expressões da matriz $\Delta_p(\boldsymbol{\theta})$ para os três esquemas de perturbação.

Ponderação de casos

Para avaliar a contribuição individual das observações podemos atribuir diferentes ponderações a cada uma delas no logaritmo da função de verossimilhança penalizada. Com efeito, seja o logaritmo da função de verossimilhança penalizada para o modelo perturbado definido por

$$L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i L_i(\boldsymbol{\theta}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}, \quad (5.23)$$

em que $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ é o vetor de pesos, com $0 \leq \omega_i \leq 1$ ($i = 1, \dots, n$). Nesse caso, o vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^T \in \mathbb{R}^n$. Usando resultados de diferenciação sobre $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$ obtemos

$$\begin{aligned} \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{2}{\hat{\phi}} W(\hat{\delta}_i) \hat{\epsilon}_i \mathbf{x}_i, \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{2}{\hat{\phi}} W(\hat{\delta}_i) \hat{\epsilon}_i \mathbf{n}_i \quad \text{e} \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \phi \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{2\hat{\phi}} - \frac{1}{2\hat{\phi}} W(\hat{\delta}_i) \hat{\delta}_i, \end{aligned}$$

em que $\hat{\epsilon}_i = y_i - \hat{\mu}_i$, com $\hat{\mu}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \mathbf{n}_i^T \hat{\mathbf{f}}$ ($i = 1, \dots, n$).

Perturbação de escala

A perturbação de escala é introduzida ao considerar $y_i \sim t_1(\mu_i, \omega_i^{-1} \phi, \nu)$, em que $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ é o vetor de perturbação, com $\omega_i > 0$ ($i = 1, \dots, n$). Nesse caso, o vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^T \in \mathbb{R}^n$ tal que $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}) = L_p(\boldsymbol{\theta}, \alpha)$. Usando resultados de diferenciação sobre $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$ obtemos

$$\begin{aligned} \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{2}{\phi} \{W'(\hat{\delta}_i) \hat{\delta}_i + W(\hat{\delta}_i)\} \hat{\boldsymbol{\epsilon}}_i \mathbf{x}_i^T, \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{2}{\phi} \{W'(\hat{\delta}_i) \hat{\delta}_i + W(\hat{\delta}_i)\} \hat{\boldsymbol{\epsilon}}_i \mathbf{n}_i^T \quad \text{e} \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \phi \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{\phi} \{W'(\hat{\delta}_i) \hat{\delta}_i + W(\hat{\delta}_i)\} \hat{\delta}_i, \end{aligned}$$

em que $\hat{\boldsymbol{\epsilon}}_i = y_i - \hat{\mu}_i$, com $\hat{\mu}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \mathbf{n}_i^T \hat{\mathbf{f}}$ ($i = 1, \dots, n$).

Perturbação na variável explicativa

Neste esquema o interesse principal é perturbar uma variável explicativa contínua específica na forma $\mathbf{x}_{i\omega} = \mathbf{x}_{i0} + \omega_i \mathbf{z}$ ($i = 1, \dots, n$), em que $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ é o vetor de perturbação e \mathbf{z}_d um vetor ($p \times 1$) com um 1 na d -ésima posição e zeros nas outras posições. Nesse caso, o vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^T \in \mathbb{R}^n$. O logaritmo da função de verossimilhança penalizada para o modelo perturbado é dado por

$$L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}) = L(\boldsymbol{\theta} | \boldsymbol{\omega}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}, \quad (5.24)$$

em que $L(\cdot)$ é dada por (5.4) e avaliada em $\delta_{i\omega} = \phi^{-1}(y_i - \mu_{i\omega})^2$, com $\mu_{i\omega} = \mathbf{x}_{i\omega}^T \boldsymbol{\beta} + \mathbf{n}_i^T \mathbf{f}$. Usando resultados de diferenciação sobre $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$ obtemos

$$\begin{aligned} \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \frac{4}{\phi} W'(\hat{\delta}_i) \hat{\beta}_d \hat{\delta}_i \mathbf{x}_i + \frac{2}{\phi} W(\hat{\delta}_i) \{\hat{\beta}_d \mathbf{x}_i - \mathbf{z}_d \hat{\boldsymbol{\epsilon}}_i\}, \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \frac{2}{\phi} \hat{\beta}_d \{2 W'(\hat{\delta}_i) \hat{\delta}_i + W(\hat{\delta}_i)\} \mathbf{n}_i \quad \text{e} \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \phi \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \frac{2}{\phi^2} \hat{\beta}_d \{2 W'(\hat{\delta}_i) \hat{\delta}_i + 2 W(\hat{\delta}_i)\} \hat{\boldsymbol{\epsilon}}_i, \end{aligned}$$

em que $\hat{\epsilon}_i = y_i - \hat{\mu}_i$, com $\hat{\mu}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \mathbf{n}_i^T \hat{\mathbf{f}}$ ($i = 1, \dots, n$). Aqui $\hat{\beta}_d$ é o d -ésimo elemento de $\hat{\boldsymbol{\beta}}$.

5.7 Conclusões do capítulo

Neste capítulo foi apresentado o modelo linear t-Student univariado. A partir do logaritmo da função de verossimilhança penalizada do modelo calculamos as funções score e as matrizes de informação de Fisher penalizada associadas aos parâmetros do modelo. Baseados nesses resultados, construímos o processo iterativo score de Fisher para estimar o coeficiente de regressão e a função não paramétrica. Os erros padrão das estimativas dos parâmetros foram estimados usando a inversa da matriz de informação de Fisher penalizada. Definimos o vetor de resíduos padronizados e calculamos sua correspondente matriz de variância-covariância. O problema de estimação do parâmetro de suavização foi abordado e sugerimos o critério de validação cruzada generalizada como um procedimento de estimação. Finalmente, particularizamos os resultados de influência local apresentados no Capítulo 4 para o modelo linear parcial t-Student univariado.

Capítulo 6

Modelo misto semiparamétrico t-Student multivariado

Neste capítulo estudamos o modelo misto semiparamétrico assumindo que a distribuição conjunta do vetor de respostas observadas e os efeitos aleatórios segue uma distribuição t-Student multivariada. As principais contribuições expostas neste capítulo são as expressões analíticas para a função escore (penalizada) e a matriz de informação de Fisher (penalizada), o processo iterativo para estimar os parâmetros do modelo, e o desenvolvimento do método de influência local.

6.1 Introdução

O modelo misto semiparamétrico (MMS) foi proposto por Zhang et al. (1998) e tem sido aplicado em diversas áreas de pesquisa devido à sua flexibilidade para modelar a estrutura de covariância intraunidades experimentais através de efeitos aleatórios de processos estocásticos, à sua capacidade de tratar com dados balanceados como desbalanceados, e à sua flexibilidade para modelar os efeitos das covariáveis que contribuem em forma paramétrica e não paramétrica sobre a variável resposta. Alguns trabalhos relacionados a esta classe de modelos são os seguintes. Zeger e Diggle (1994) estudam o viés e a variância do estimador da função não paramétrica obtido através de suavização de *kernel*. Zhang et al. (1998) propõem uma forma fechada para os estimadores de máxima verossimilhança pe-

nalizada do coeficiente de regressão e da função não paramétrica. Além disso, eles estudam e comparam algumas propriedades teóricas desses estimadores desde o ponto de vista frequentista como bayesiano. Tais resultados mostram, por exemplo, que os intervalos de confiança bayesianos para a função não paramétrica apresentam um melhor comportamento em relação aos intervalos de confiança frequentistas; e que o parâmetro de suavização e os componentes de variância podem ser estimados por máxima verossimilhança restrita em forma simultânea, incorporando o inverso do parâmetro de suavização como um componente de variância extra. Zhang et al. (2000) desenvolvem um teste qui-quadrado para provar a igualdade de duas funções não paramétricas no modelo MMS gaussiano para dados longitudinais periódicos. O teste foi aplicado a um conjunto de dados reais e seu comportamento foi avaliado através de simulações. Lin e Zhang (1999) introduzem os modelos mistos aditivos generalizados (MMAGs) e mostram, através de um estudo de simulação, que o estimador de quase-verossimilhança duplamente penalizada das funções não paramétricas apresentam em geral um bom comportamento. Entretanto, o estimador de quase-verossimilhança duplamente penalizada dos componentes de variância apresentam problemas de viés. Nesse contexto, eles introduzem algumas modificações no procedimento de correção do viés de Lin e Breslow (1996) para obter melhores estimadores. O método foi aplicado a um conjunto de dados reais e seu comportamento avaliado através de simulações. Fung et al. (2002) apresentam uma importante contribuição, em que estendem as medidas de diagnóstico baseadas nos resíduos e nas distâncias de Cook para detectar observações influentes nas estimativas de máxima verossimilhança penalizada derivadas de um modelo MMS. E, recentemente, Zhu et al. (2003) estendem a técnica de influência local para esses modelos.

A distribuição t -Student multivariada tem sido utilizada por vários autores como alternativa à distribuição normal, principalmente pelo fato de acomodar observações aberrantes. Diversos trabalhos têm mostrado que os estimadores de máxima verossimilhança derivados dessa distribuição são menos sensíveis a observações aberrantes em relação ao caso normal. Alguns trabalhos relacionados

nessa direção são os seguintes. Maronna (1976) discute o problema de estimação de máxima verossimilhança da média e da matriz de variância-covariância sob erros com distribuição t -Student. Rubin (1983) considera o critério de máxima verossimilhança para estimar os parâmetros associados ao modelo t -Student multivariado e Little (1988) faz uma extensão desse trabalho. Lange et al. (1989) estudam alguns aspectos inferenciais do modelo t -Student multivariado com graus de liberdade desconhecidos, e discutem o uso dessa distribuição em regressão. Richardson (1997) estuda o modelo misto (MM) em que a distribuição marginal dos vetores de respostas é t -Student multivariada. Kowalski et al. (1999) comparam alguns aspectos inferenciais da teoria clássica e bayesiana no modelo linear t -Student multivariado. Recentemente, Pinheiro et al. (2001) introduziram os modelos mistos hierárquicos robustos nos quais os erros e efeitos aleatórios têm uma distribuição t -Student multivariada.

6.2 Especificação do modelo

Seja y_{ij} a j -ésima medição associada ao i -ésimo grupo. O modelo misto semi-paramétrico (MMS) assume a forma

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i + f(t_{ij}) + \epsilon_{ij}, \quad (6.1)$$

em que \mathbf{x}_{ij} e \mathbf{z}_{ij} são, respectivamente, vetores $(p \times 1)$ e $(q \times 1)$ que contêm valores de variáveis explicativas, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos $(p \times 1)$, \mathbf{b}_i é o vetor de efeitos aleatórios $(q \times 1)$, $f(\cdot)$ é uma função suave univariada, duas vezes diferenciável, que depende da covariável t_{ij} , e ϵ_{ij} é um erro aleatório $(i = 1, \dots, n; j = 1, \dots, m_i)$. O modelo (6.1) pode ser expresso matricialmente na forma

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{N}_i \mathbf{f} + \boldsymbol{\epsilon}_i, \quad (6.2)$$

em que $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^T$ é um vetor aleatório $(m_i \times 1)$ de respostas observadas para o i -ésimo grupo, \mathbf{X}_i é uma matriz de planejamento $(m_i \times p)$ com

linhas \mathbf{x}_{ij}^T , \mathbf{Z}_i é matriz de planejamento ($m_i \times q$) de efeitos aleatórios com linhas \mathbf{z}_{ij}^T , $\mathbf{f} = (f(t_1^0), \dots, f(t_r^0))^T$ com t_1^0, \dots, t_r^0 sendo os valores ordenados e distintos dos tempos t_{ij} , \mathbf{N}_i é uma matriz de incidência ($m_i \times r$) cujo (j, ℓ) -ésimo elemento é igual à função indicadora $I(t_{ij} = t_\ell^0)$ ($j = 1, \dots, m_i; \ell = 1, \dots, r$), e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im_i})^T$ é um vetor de erros aleatórios não observáveis ($m_i \times 1$). Fazendo $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$, $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$, $\mathbf{N} = (\mathbf{N}_1^T, \dots, \mathbf{N}_n^T)^T$, $\mathbf{Z} = \text{diag}\{\mathbf{Z}_1, \dots, \mathbf{Z}_n\}$, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)^T$ e $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \dots, \boldsymbol{\epsilon}_n^T)^T$, podemos escrever o modelo (6.2) compactamente como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{N}\mathbf{f} + \boldsymbol{\epsilon}. \quad (6.3)$$

Assumindo que,

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \sim \mathbf{t}_{(m_i+q+m_i)} \left\{ \begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \phi \mathbf{I}_{m_i} & \mathbf{Z}_i\mathbf{D} & \mathbf{V}_i \\ \mathbf{D}\mathbf{Z}_i^T & \mathbf{D} & \mathbf{0} \\ \mathbf{V}_i & \mathbf{0} & \mathbf{V}_i \end{pmatrix} \right\},$$

temos que os vetores aleatórios de respostas observadas \mathbf{y}_i 's seguem uma distribuição marginal na forma

$$\mathbf{y}_i \sim \mathbf{t}_{m_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu_i), \quad (6.4)$$

cuja função densidade é dada por

$$f_{\mathbf{y}}(\mathbf{y}_i) = \frac{|\boldsymbol{\Sigma}_i|^{-1/2} \Gamma(\frac{m_i+\nu_i}{2})}{(\pi\nu_i)^{m_i/2} \Gamma(\frac{\nu_i}{2})} \left(1 + \frac{\delta_i}{\nu_i}\right)^{-\left(\frac{m_i+\nu_i}{2}\right)},$$

em que $\delta_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$ é a distância de Mahalanobis, $\Gamma(\cdot)$ denota a função gama, ν_i denota os graus de liberdade, $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f}$ é um vetor ($m_i \times 1$), e $\boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \phi \mathbf{I}_{m_i}$ é uma matriz ($m_i \times m_i$), sendo $\mathbf{D} = \mathbf{D}(\boldsymbol{\lambda})$ uma matriz parametrizada pelo vetor de parâmetros $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{d_\lambda})$. De acordo com Lange et al. (1989), temos que

$$\begin{aligned} E(\mathbf{y}_i) &= \boldsymbol{\mu}_i \quad (\nu_i > 0) \quad \text{e} \\ \text{Var}(\mathbf{y}_i) &= \frac{\nu_i}{\nu_i - 2} \boldsymbol{\Sigma}_i \quad (\nu_i > 2). \end{aligned}$$

Vamos assumir que os graus de liberdade $\nu_i = \nu$ ($i = 1, \dots, n$) são fixos. Dessa forma, o vetor de parâmetros a ser estimado será

$$\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \boldsymbol{\lambda}^T, \phi)^T,$$

cujos espaço paramétrico associado é dado por

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{p^*} \mid \boldsymbol{\beta} \in \Theta_\beta, \mathbf{f} \in \Theta_f, \boldsymbol{\lambda} \in \Theta_\lambda, \phi \in \Theta_\phi \right\},$$

em que $\Theta_\beta \subseteq \mathbb{R}^p$, $\Theta_f \subseteq \mathbb{R}^r$, $\Theta_\lambda \subseteq \mathbb{R}^{d_\lambda}$ e $\Theta_\phi = \mathbb{R}^+$ denotam, respectivamente, os espaços paramétricos associados com os efeitos fixos do modelo (paramétricos e não paramétricos) e os componentes de efeitos aleatórios; e $p^* = p + q + d_\lambda + 1$. Logo, o logaritmo da função de verossimilhança de $\boldsymbol{\theta}$ pode ser expresso na forma

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n L_i(\boldsymbol{\theta}), \quad (6.5)$$

em que

$$L_i(\boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log \left\{ \frac{\Gamma\left(\frac{m_i + \nu}{2}\right)}{(\pi\nu)^{1/2} \Gamma\left(\frac{\nu}{2}\right)} \right\} + \log \left\{ 1 + \nu^{-1} \delta_i \right\}^{-\left(\frac{m_i + \nu}{2}\right)}. \quad (6.6)$$

Como a função f é um parâmetro infinito dimensional, propusemos no Capítulo 3 estimar o vetor de parâmetros $\boldsymbol{\theta}$ baseados no critério de máxima verossimilhança penalizada, o qual conduz a um estimador *spline* cúbico natural para f com nós nos pontos t_g^0 ($g = 1, \dots, r$). No caso específico do modelo misto semiparamétrico t-Student multivariado, o estimador de máxima verossimilhança penalizada de $\boldsymbol{\theta}$ maximiza o logaritmo da função de verossimilhança penalizada definido por

$$L_p(\boldsymbol{\theta}, \alpha) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \alpha), \quad (6.7)$$

em que

$$\begin{aligned} L_{p_i}(\boldsymbol{\theta}, \alpha) &= L_i(\boldsymbol{\theta}) - \frac{\alpha}{2n} \int_a^b [f^{(2)}(t)]^2 dt \\ &= L_i(\boldsymbol{\theta}) - \frac{\alpha}{2n} \mathbf{f}^T \mathbf{K} \mathbf{f}, \end{aligned} \quad (6.8)$$

$f_k^{(2)}(t) = \frac{d^2}{dt^2} f(t)$ com $t \in [a, b]$, f é uma função que pertence ao espaço de funções de Sobolev definido em (3.2.1), α é o parâmetro de suavização que regula a qualidade do ajuste e a ondulação da estimativa de f , e \mathbf{K} é a matriz de suavização positiva definida dada pela equação (2.3) de Green e Silverman (1994). Dessa forma, o valor de $\boldsymbol{\theta}$ que maximiza $L_p(\boldsymbol{\theta}, \alpha)$, em todo o espaço paramétrico Θ , digamos $\hat{\boldsymbol{\theta}}$, é chamado estimador de máxima verossimilhança penalizada (EMVP) de $\boldsymbol{\theta}$, e satisfaz

$$L_p(\hat{\boldsymbol{\theta}}, \alpha) \geq \sup_{\boldsymbol{\theta} \in \Theta} L_p(\boldsymbol{\theta}, \alpha).$$

Na seguinte seção, calculamos a função score e a matriz de informação de Fisher penalizada para o modelo misto semiparamétrico \mathbf{t} -Student multivariado, e apresentamos um processo iterativo para obter a estimativa de $\boldsymbol{\theta}$.

6.3 Processo de estimação

A investigação neste campo tem gerado várias estratégias para abordar o problema de estimação na classe de modelos mistos semiparamétricos. Alguns autores têm sugerido utilizar os métodos de suavização de *kernel* e suavização *spline*. Por exemplo, utiliza-se suavização de *kernel* para estimar a função não paramétrica. Zhang et al. (1998), por sua vez, obtém o estimador do coeficiente de regressão e da função não paramétrica baseados na verossimilhança penalizada. Aliás, eles mostram que o estimador da função não paramétrica é uma *spline* cúbica natural e que os componentes de variância e o parâmetro de suavização podem ser

estimados em forma simultânea usando um modelo misto unificado. No contexto dos modelos mistos aditivos generalizados, Lin e Zhang (1999) propõem estimar a função não paramétrica baseados na quase-verossimilhança penalizada, e os componentes de variância e o parâmetro de suavização através da quase-verossimilhança marginal. E, no mesmo contexto, Fahrmeir e Lang (2001) propõem o uso de técnicas bayesianas.

6.3.1 Função escore

Seja $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots, \tau_{d_\lambda})^T$, com $\tau_0 = \phi$, $\tau_\ell = \lambda_\ell$ ($\ell = 1, \dots, d_\lambda$), e $v_i(\boldsymbol{\theta}) = (\nu + m_i)/(\nu + \delta_i)$. Particularizando o resultado apresentado na equação (3.10), temos que as funções escore de $\boldsymbol{\beta}$, \mathbf{f} e $\boldsymbol{\tau}$, respectivamente, são dada por

$$\begin{aligned} \mathbf{U}_p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta}} \\ &= \sum_{i=1}^n \mathbf{U}_{p_i}(\boldsymbol{\theta}), \end{aligned}$$

em que

$$\mathbf{U}_{p_i}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_{p_i}^\beta(\boldsymbol{\theta}) \\ \mathbf{U}_{p_i}^f(\boldsymbol{\theta}) \\ \mathbf{U}_{p_i}^\tau(\boldsymbol{\theta}) \end{pmatrix}, \quad (6.9)$$

sendo as funções escore parciais associadas aos efeitos fixos do modelo dadas por

$$\begin{aligned} \mathbf{U}_p^\beta(\boldsymbol{\theta}) &= v_i(\boldsymbol{\theta}) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \quad \text{e} \\ \mathbf{U}_p^f(\boldsymbol{\theta}) &= v_i(\boldsymbol{\theta}) \mathbf{N}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) - \frac{\alpha}{n} \mathbf{Kf}, \end{aligned} \quad (6.10)$$

e as funções escore parciais associadas aos componentes de variância dadas por

$$\mathbf{U}_p^\tau(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_p^{\tau_0}(\boldsymbol{\theta}) \\ \mathbf{U}_p^{\tau_1}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{U}_p^{\tau_{d_\lambda}}(\boldsymbol{\theta}) \end{pmatrix}, \quad (6.11)$$

em que

$$U_{p_i}^{\tau_\ell}(\boldsymbol{\theta}) = -\frac{1}{2} \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \} - v_i(\boldsymbol{\theta}) (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right],$$

com $\dot{\boldsymbol{\Sigma}}_i(\ell) = \partial \boldsymbol{\Sigma}_i / \partial \tau_\ell$ ($\ell = 1, \dots, d_\lambda$). Note que os pesos $v_i(\boldsymbol{\theta})$ que aparecem nas equações acima são proporcionais à distância de Mahalanobis e portanto as observações aberrantes tendem a ter pesos pequenos no processo de estimação. Assim, é de esperar que o estimador de máxima verossimilhança penalizada gerado pelo modelo misto semiparamétrico \mathbf{t} -multivariado seja menos sensível a observações aberrantes em relação ao modelo misto semiparamétrico normal multivariado.

6.3.2 Matriz de informação de Fisher

Baseados nos resultados apresentados na equação (3.17), temos que a matriz de informação de Fisher ($p^* \times p^*$) para o modelo misto semiparamétrico \mathbf{t} -Student multivariado assume a seguinte forma bloco diagonal:

$$\begin{aligned} \mathcal{I}_p(\boldsymbol{\theta}) &= -\mathbb{E} \left\{ \frac{\partial^2 L_p(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} \\ &= \begin{pmatrix} \mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix}, \end{aligned} \quad (6.12)$$

em que

$$\mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{I}_{p_i}^{\beta\beta}(\boldsymbol{\theta}),$$

sendo

$$\mathcal{I}_{p_i}^{\beta\beta}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{p_i}^{\beta\beta}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{\beta\tau}(\boldsymbol{\theta}) \\ \mathcal{I}_{p_i}^{\tau\beta}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix}, \quad (6.13)$$

com

$$\begin{aligned}
\mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) &= \left(\frac{m_i + \nu}{m_i + \nu + 2} \right) \mathbf{X}_i^T \boldsymbol{\Sigma}_i \mathbf{X}_i, \\
\mathcal{I}_p^{\beta\mathbf{f}}(\boldsymbol{\theta}) &= \left(\frac{m_i + \nu}{m_i + \nu + 2} \right) \mathbf{X}_i^T \boldsymbol{\Sigma}_i \mathbf{N}_i, \\
\mathcal{I}_p^{\mathbf{f}\mathbf{f}}(\boldsymbol{\theta}) &= \left(\frac{m_i + \nu}{m_i + \nu + 2} \right) \mathbf{N}_i^T \boldsymbol{\Sigma}_i \mathbf{N}_i + \alpha \mathbf{K},
\end{aligned}$$

e $\mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta})$ é uma matriz $[(d^* + 1) \times (d^* + 1)]$ em que o (j^*, ℓ^*) -ésimo elemento pode ser expresso na forma

$$\begin{aligned}
\mathcal{I}_p^{\tau_{j^*} \tau_{\ell^*}}(\boldsymbol{\theta}) &= \frac{1}{2} \sum_{i=1}^n \left[\left(\frac{\nu}{m_i + \nu + 2} \right) \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_{j^*}} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_{\ell^*}} \right\} - \right. \\
&\quad \left. \left(\frac{1}{m_i + \nu + 2} \right) \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_{j^*}} \right\} \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_{\ell^*}} \right\} \right], \tag{6.14}
\end{aligned}$$

para $j^* = 0, 1, \dots, d^*$ e $\ell^* = 0, 1, \dots, d^*$. Os vetores de parâmetros associados aos efeitos fixos do modelo, $\boldsymbol{\beta}$ e \mathbf{f} , são, respectivamente, ortogonais em relação ao vetor de parâmetros associado aos componentes de variância $\boldsymbol{\tau}$.

6.3.3 Efeitos fixos e componentes de variância

Suponhamos que os parâmetros α e $\boldsymbol{\tau}$ são fixos. Baseados nos resultados apresentados na Subseção 3.5.1, as equações de estimação

$$\mathbf{U}_p^\beta(\boldsymbol{\theta}) = \mathbf{0} \quad \text{e} \quad \mathbf{U}_p^{\mathbf{f}}(\boldsymbol{\theta}) = \mathbf{0},$$

podem ser resolvidas através do seguinte algoritmo score de Fisher:

$$\begin{pmatrix} \mathbf{X}^T \mathbf{W}^* \mathbf{X} & \mathbf{X}^T \mathbf{W}^* \mathbf{N} \\ \mathbf{N}^T \mathbf{W}^* \mathbf{X} & \mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \phi \mathbf{K} \end{pmatrix}^{(u)} \begin{pmatrix} \boldsymbol{\beta}^{(u+1)} - \boldsymbol{\beta}^{(u)} \\ \mathbf{f}^{(u+1)} - \mathbf{f}^{(u)} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_p^\beta(\boldsymbol{\theta}) \\ \mathbf{U}_p^{\mathbf{f}}(\boldsymbol{\theta}) \end{pmatrix}^{(u)}.$$

Logo, após algumas manipulações algébricas, obtemos

$$\begin{pmatrix} \boldsymbol{\beta}^{(u+1)} \\ \mathbf{f}^{(u+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0 \{ \mathbf{r}_{\beta-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \\ \mathbf{S}_1 \{ \mathbf{r}_{f-}^{(u,u+1)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \} \end{pmatrix}, \quad (6.15)$$

em que

$$\mathbf{S}_k = \begin{cases} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^* & k = 0 \\ (\mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}^* & k = 1, \end{cases}$$

são matrizes de suavização que geram as estimativas do coeficiente de regressão, $\boldsymbol{\beta}$, e da função não paramétrica, f ,

$$\mathbf{r}_{\vartheta-}^{(u,u+1)} = \begin{cases} (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{N} \mathbf{f}^{(u+1)} & \vartheta = \beta \\ (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(u+1)} & \vartheta = f, \end{cases}$$

são os resíduos parciais que permitem ajustar o componente paramétrico e o componente não paramétrico do modelo. Note que,

$$\begin{aligned} \boldsymbol{\mu}^{(u)} &= \mathbf{X} \boldsymbol{\beta}^{(u)} + \mathbf{N} \mathbf{f}^{(u)}, \\ \widetilde{\mathbf{W}}^{(u)} &= \mathbf{I}_{n^*} - \mathbf{W}^{*-1} \mathbf{W}^{(u)}, \\ \mathbf{W}^* &= \text{diag}\{\psi_1 \mathbf{W}_1, \dots, \psi_n \mathbf{W}_n\} \quad \text{e} \\ \mathbf{W}^{(u)} &= \text{diag}\{v_1^{(u)} \mathbf{W}_1, \dots, v_n^{(u)} \mathbf{W}_n\}, \end{aligned}$$

em que $\psi_i = (m_i + \nu)/(m_i + \nu + 2)$, $v_i^{(u)} = (m_i + \nu)/(\nu + \delta_i^{(u)})$, $\delta_i^{(u)} = (\mathbf{y}_i - \boldsymbol{\mu}_i^{(u)})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{(u)})$, $\mathbf{W}_i = \boldsymbol{\Sigma}_i^{-1}$, e $n^* = \sum_{i=1}^n m_i$. Para resolver (6.15) podemos utilizar o algoritmo *backfitting* descrito na Tabela 6.1.

Para resolver a equação de estimação $\mathbf{U}_p^\tau(\boldsymbol{\theta}) = \mathbf{0}$ podemos utilizar o seguinte processo iterativo:

$$\boldsymbol{\tau}^{(u+1)} = \arg \max_{\boldsymbol{\tau}} \left\{ L_p^c(\boldsymbol{\beta}_\tau^{(u+1)}, \mathbf{f}_\tau^{(u+1)}, \boldsymbol{\tau}, \alpha) \right\}, \quad (6.16)$$

em que $L_p^c(\boldsymbol{\beta}_\tau^{(u+1)}, \mathbf{f}_\tau^{(u+1)}, \boldsymbol{\tau}, \alpha)$ denota o logaritmo da função de verossimilhança

penalizada concentrada de $\boldsymbol{\tau}$ definida por

$$L_p^c(\boldsymbol{\beta}_\tau^{(u+1)}, \mathbf{f}_\tau^{(u+1)}, \boldsymbol{\tau}, \alpha) \propto \sum_{i=1}^n \left\{ -\frac{1}{2} \log |\boldsymbol{\Sigma}_{\tau_i}| + \log \{1 + \nu^{-1} \tilde{\delta}_i\}^{-\left(\frac{m_i + \nu_i}{2}\right)} \right\} - \frac{\alpha}{2} \widehat{\mathbf{f}}_\tau^{(u+1)T} \mathbf{K} \widehat{\mathbf{f}}_\tau^{(u+1)},$$

com $\tilde{\delta}_i = \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_{\tau_i}^{-1} \boldsymbol{\varepsilon}_i$, e $\boldsymbol{\varepsilon}_i = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_\tau^{(u+1)} - \mathbf{N}_i \mathbf{f}_\tau^{(u+1)})$. Para maximizar (6.16) podemos usar o algoritmo escore de Fisher definido por

$$\boldsymbol{\tau}^{(u+1)} = \boldsymbol{\tau}^{(u)} + \left\{ \mathcal{I}_p^{\tau\tau}(\boldsymbol{\theta})^{(u)} \right\}^{-1} \mathbf{U}_p^\tau(\boldsymbol{\theta})^{(u)}. \quad (6.17)$$

O processo iterativo para estimar $\boldsymbol{\theta}$, definido pelas etapas (6.15)-(6.17), deve ser iniciado com valores $\boldsymbol{\beta}^{(0)}$, $\mathbf{f}^{(0)}$ e $\boldsymbol{\tau}^{(0)}$, e deve ser repetido até a convergência. Para iniciar o processo podemos considerar as estimativas do modelo normal.

Tabela 6.1: $(u^* + 1)$ -ésima etapa do algoritmo *backfitting* na $(u + 1)$ -ésima etapa do processo iterativo escore de Fisher sob o modelo misto semiparamétrico **t**-Student multivariado.

-
-
- (i) Inicie o processo iterativo com $\boldsymbol{\beta}^{(u)} = \boldsymbol{\beta}^{(u,0)}$ e $\mathbf{f}^{(u)} = \mathbf{f}^{(0,0)}$,
 - (ii) Para $u, u^* = 0, 1, 2, \dots$ calcular
 - (ii')

$$\mathbf{r}_{\beta^-}^{(u,u^*)} = (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{N} \mathbf{f}^{(u,u^*)} \quad \text{e}$$

$$\boldsymbol{\beta}^{(u+1,u^*+1)} = \mathbf{S}_0 \left\{ \mathbf{r}_{\beta^-}^{(u,u^*)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \right\} .$$
 - (ii'')

$$\mathbf{r}_{\mathbf{f}^-}^{(u,u^*)} = (\mathbf{I}_n - \widetilde{\mathbf{W}}^{(u)}) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(u+1,u^*+1)} \quad \text{e}$$

$$\mathbf{f}^{(u+1,u^*+1)} = \mathbf{S}_1 \left\{ \mathbf{r}_{\mathbf{f}^-}^{(u,u^*)} + \widetilde{\mathbf{W}}^{(u)} \boldsymbol{\mu}^{(u)} \right\} .$$
 - (iii) Repita (i) e (ii) até atingir a convergência desejada .
-
-

6.3.4 Erro padrão

O erro padrão estimado de $\widehat{\boldsymbol{\theta}}$ pode ser calculado desde a matriz de variância-covariância assintótica,

$$\widehat{\text{Cov}}(\widehat{\boldsymbol{\theta}}) \approx \mathcal{I}_p^{-1}(\widehat{\boldsymbol{\theta}}).$$

Com efeito, a partir de (6.12) podemos mostrar, após algumas manipulações algébricas, que a inversa da matriz de informação de Fisher é dada por

$$\mathcal{I}_p^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_p^{\beta f^{-1}}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_p^{\tau \tau^{-1}}(\boldsymbol{\theta}) \end{pmatrix},$$

em que

$$\mathcal{I}_p^{\beta f^{-1}}(\boldsymbol{\theta}) = \begin{pmatrix} (\mathbf{X}^T \mathcal{W}_x \mathbf{X})^{-1} & -\mathbf{E} \\ -\mathbf{E}^T & (\mathbf{N}^T \mathcal{W}_f \mathbf{N} + \alpha \mathbf{K})^{-1} \end{pmatrix}, \quad (6.18)$$

com

$$\begin{aligned} \mathbf{E} &= (\mathbf{X}^T \mathcal{W}_x \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^* \mathbf{N}) (\mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \mathbf{K})^{-1}, \\ \mathcal{W}_x &= \mathbf{W}^* - \mathbf{W}^* \mathbf{N} (\mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}^* \quad \text{e} \\ \mathcal{W}_f &= \mathbf{W}^* - \mathbf{W}^* \mathbf{X} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*. \end{aligned}$$

De (6.18) segue que as matrizes de variância-covariância assintótica de $\widehat{\boldsymbol{\beta}}$ e $\widehat{\mathbf{f}}$ assumem a forma

$$\begin{aligned} \text{Cov}_A(\widehat{\boldsymbol{\beta}}) &\approx (\mathbf{X}^T \mathcal{W}_x \mathbf{X})^{-1} \quad \text{e} \\ \text{Cov}_A(\widehat{\mathbf{f}}) &\approx (\mathbf{N}^T \mathcal{W}_f \mathbf{N} + \alpha \mathbf{K})^{-1}. \end{aligned}$$

Sob o modelo normal (assumindo $\boldsymbol{\tau}$ fixo), em que $v_i = 1$ e $d_{g_i} = \frac{m_i}{4}$, os erros padrão das estimativas $\widehat{\boldsymbol{\beta}}$ e $\widehat{\mathbf{f}}$ podem ser calculados usando um procedimento frequentista ou bayesiano. No caso frequentista, esses erros são calculados fixando a função f , entretanto no caso bayesiano os erros padrão são calculados assumindo distribuições a priori para $\boldsymbol{\beta}$ e f . Para o caso frequentista temos que

$$\begin{aligned}\text{Cov}_F(\widehat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W}_x \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_x \text{Cov}(\mathbf{y}) \mathbf{W}_x \mathbf{X} (\mathbf{X}^T \mathbf{W}_x \mathbf{X})^{-1} \quad \text{e} \\ \text{Cov}_F(\widehat{\mathbf{f}}) &= (\mathbf{N}^T \mathbf{W}_f \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}_f \text{Cov}(\mathbf{y}) \mathbf{W}_f \mathbf{N} (\mathbf{N}^T \mathbf{W}_f \mathbf{N} + \alpha \mathbf{K})^{-1},\end{aligned}$$

em que

$$\begin{aligned}\mathbf{W}_x &= \mathbf{W}^* - \mathbf{W}^* \mathbf{N} (\mathbf{N}^T \mathbf{W}^* \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}^* \quad \text{e} \\ \mathbf{W}_f &= \mathbf{W}^* - \mathbf{W}^* \mathbf{X} (\mathbf{X}^T \mathbf{W}^* \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^*,\end{aligned}$$

com $\text{Cov}(\mathbf{y}) = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n)$. Além disso, é possível mostrar que o viés de tais estimadores é, respectivamente,

$$\begin{aligned}\text{Viés}(\widehat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W}_x \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_x \mathbf{N} \mathbf{f} \quad \text{e} \\ \text{Viés}(\widehat{\mathbf{f}}) &= -\alpha (\mathbf{N}^T \mathbf{W}_f \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{K} \mathbf{f}.\end{aligned}$$

Note que quando $\alpha \downarrow 0$, $\text{Viés}(\widehat{\boldsymbol{\beta}}) \downarrow 0$ e $\text{Viés}(\widehat{\mathbf{f}}) \downarrow 0$. No contexto bayesiano (veja Zhang et al., 1998), as matrizes de variância-covariância de $\widehat{\boldsymbol{\beta}}$ e $\widehat{\mathbf{f}}$ têm uma expressão mais simples em relação à obtida pelo procedimento frequentista, isto é,

$$\begin{aligned}\text{Cov}_B(\widehat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W}_x \mathbf{X})^{-1} \quad \text{e} \\ \text{Cov}_B(\widehat{\mathbf{f}}) &= (\mathbf{N}^T \mathbf{W}_f \mathbf{N} + \alpha \mathbf{K})^{-1}.\end{aligned}$$

De acordo com Zhang et al. (1998), a diferença entre as matrizes de variância-covariância bayesiana e frequentista é semi-positiva definida, e portanto o erro padrão bayesiano de $\widehat{\boldsymbol{\beta}}$ e $\widehat{\mathbf{f}}$ é geralmente maior do que o erro padrão frequentista. Nessa direção, eles apresentam um interessante estudo de simulação e observam, por exemplo, que há diferenças no cálculo do erro padrão da estimativa do coeficiente de regressão entre o aproveitamento bayesiano e frequentista. É importante mencionar também que os erros padrão frequentista e bayesiano, podem ser usados na construção de intervalos de confiança (Wahba, 1983).

6.3.5 Shrinkage e valor ajustado

A propriedade conhecida na literatura como *shrinkage* no caso dos modelos mistos com erros normais, pode ser também estendida para os modelos mistos semiparamétricos com erros *t*-Student. Com efeito, estimando os efeitos aleatórios através do estimador empírico de Bayes,

$$\begin{aligned}\widehat{\mathbf{b}}_i &= \text{E}\{\widehat{\mathbf{b}}_i \mid \mathbf{y}_i\} \\ &= \widehat{\mathbf{D}} \mathbf{Z}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} - \mathbf{N}_i \widehat{\mathbf{f}}),\end{aligned}\tag{6.19}$$

em que $\widehat{\boldsymbol{\beta}}$ e $\widehat{\mathbf{f}}$ correspondem às estimativas de máxima verossimilhança penalizada, temos que o valor ajustado para o i -ésimo grupo pode ser expresso na forma

$$\begin{aligned}\widehat{\mathbf{y}}_i &= \mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{Z}_i \widehat{\mathbf{b}}_i + \mathbf{N}_i \widehat{\mathbf{f}} \\ &= \widehat{\phi} \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{N}_i \widehat{\mathbf{f}}) + (\mathbf{I}_{m_i} - \widehat{\phi} \widehat{\boldsymbol{\Sigma}}_i^{-1}) \mathbf{y}_i.\end{aligned}$$

Salientamos que $\widehat{\mathbf{y}}_i$ pode ser interpretado como uma média ponderada entre o perfil da população $(\mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{N}_i \widehat{\mathbf{f}})$ e os dados observados \mathbf{y}_i , com pesos $\widehat{\phi} \widehat{\boldsymbol{\Sigma}}_i^{-1}$ e $(\mathbf{I}_{m_i} - \widehat{\phi} \widehat{\boldsymbol{\Sigma}}_i^{-1})$, respectivamente.

6.4 Análise de resíduos

A análise de resíduos tem sido o primeiro procedimento de diagnóstico sugerido para detectar observações aberrantes e avaliar a adequação do ajuste do modelo proposto. Nesta seção, definimos dois tipos de resíduos padronizados no contexto dos modelos mistos semiparamétricos *t*-Student multivariado; especificamente, o resíduo marginal, associado aos efeitos fixos do modelo, e o resíduo condicional, associado aos efeitos fixos e aleatórios do modelo. Maiores detalhes da definição desses resíduos no contexto dos modelos mistos paramétricos, podem ser encontrados, por exemplo, em Verbeke e Molenberghs (2001). Para definir os resíduos marginal e condicional, é conveniente derivar uma solução em forma fechada para

as estimativas do coeficiente de regressão e da função não paramétrica. Suponhamos que α , Σ_i e $v_i(\boldsymbol{\theta}) = (\nu + m_i)/(\nu + \delta_i)$ são fixos ($i = 1, \dots, n$). A partir de (6.10), é possível obter as seguintes equações:

$$\begin{aligned} (\mathbf{X}^T \mathbf{W} \mathbf{X}) \boldsymbol{\beta} &= \mathbf{X}^T \mathbf{W} (\mathbf{y} - \mathbf{N} \mathbf{f}) \\ (\mathbf{N}^T \mathbf{W} \mathbf{N} + \alpha \mathbf{K}) \mathbf{f} &= \mathbf{N}^T \mathbf{W} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}). \end{aligned}$$

Logo, após alguma álgebra, temos que as estimativas de máxima verossimilhança penalizada de $\boldsymbol{\beta}$ e \mathbf{f} assumem a forma

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W}_x \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_x \mathbf{y} \quad \text{e} \quad (6.20)$$

$$\widehat{\mathbf{f}} = (\mathbf{N}^T \mathbf{W}_f \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}_f \mathbf{y}, \quad (6.21)$$

em que

$$\mathbf{W}_x = \mathbf{W} - \mathbf{W} \mathbf{N} (\mathbf{N}^T \mathbf{W} \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W} \quad \text{e}$$

$$\mathbf{W}_f = \mathbf{W} - \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}.$$

Maiores detalhes com respeito ao problema de estimação do coeficiente de regressão e da função não paramétrica no modelo misto semiparamétrico normal podem ser encontrados em Zhang et al. (1998) e Fung et al. (2002).

6.4.1 Resíduo marginal

Substituindo (6.20) e (6.21) em (6.3), e após algumas manipulações algébricas, podemos definir o vetor de resíduos marginal, associado aos efeitos fixos do modelo, a partir de

$$\begin{aligned} \widehat{\mathbf{y}} &= \widehat{\mathbb{E}(\mathbf{y})} \\ &= \mathbf{X} \widehat{\boldsymbol{\beta}} + \mathbf{N} \widehat{\mathbf{f}} \\ &= \overline{\mathbf{H}} \mathbf{y}, \end{aligned} \quad (6.22)$$

em que

$$\bar{\mathbf{H}} = (\mathbf{X} \ \mathbf{N}) \mathbf{C}^{-1} \begin{pmatrix} \mathbf{X}^T \\ \mathbf{N}^T \end{pmatrix} \mathbf{W}, \quad (6.23)$$

com

$$\mathbf{C} = \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{N} \\ \mathbf{N}^T \mathbf{W} \mathbf{X} & \mathbf{N}^T \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \end{pmatrix}.$$

Assim, o vetor de resíduos parcial pode ser expresso na forma

$$\begin{aligned} \hat{\mathbf{e}}_m &= \mathbf{y} - \widehat{\mathbf{E}(\mathbf{y})} \\ &= \mathbf{y} - \hat{\mathbf{y}} \\ &= (\mathbf{I}_{n^*} - \bar{\mathbf{H}}) \mathbf{y}. \end{aligned} \quad (6.24)$$

A matriz $\bar{\mathbf{H}}$ é equivalente à matriz de alavancas definida para os modelos lineares. Em analogia com esses modelos, os elementos diagonais de $\bar{\mathbf{H}}$ são chamados pontos de alavanca e desempenham um papel importante na construção de técnicas de diagnóstico para os modelos não paramétricos e semiparamétricos. Note que todos os pontos de alavanca associados aos efeitos fixos do modelo estão relacionados diretamente com a matriz $\bar{\mathbf{H}}$. As propriedades dessa matriz para regressão não paramétrica são discutidas por Eubank (1984) e sua aplicação na construção de técnicas de diagnóstico para os modelos MMS é estudada por Fung et al. (2002).

6.4.2 Resíduo condicional

A partir de (6.19) temos que o vetor de efeitos aleatórios estimados é dado por $\hat{\mathbf{b}} = (\hat{\mathbf{b}}_1^T, \dots, \hat{\mathbf{b}}_n^T)^T$. Baseados nesse resultado, podemos definir o vetor de resíduos condicional, associado aos efeitos fixos e aleatórios do modelo, a partir de

$$\begin{aligned} \hat{\mathbf{y}} &= \widehat{\mathbf{E}(\mathbf{y} | \mathbf{b})} \\ &= \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{N} \hat{\mathbf{f}} + \mathbf{Z} \hat{\mathbf{b}} \\ &= \mathbf{H} \mathbf{y}, \end{aligned} \quad (6.25)$$

em que $\mathbf{H} = \mathbf{I}_{n^*} - \phi \mathbf{W} + \phi \mathbf{W} \bar{\mathbf{H}}$, com $n^* = \sum_{i=1}^n m_i$ e $\bar{\mathbf{H}}$ definida na equação (6.23). Assim, o vetor de resíduos condicional pode ser expresso na forma

$$\begin{aligned} \hat{\mathbf{e}}_c &= \mathbf{y} - \widehat{\mathbf{E}(\mathbf{y} | \mathbf{b})} \\ &= \mathbf{y} - \hat{\mathbf{y}} \\ &= (\mathbf{I}_{n^*} - \mathbf{H}) \mathbf{y}. \end{aligned} \quad (6.26)$$

É importante lembrar que os resíduos apresentados acima estão condicionados no sentido de que α , Σ_i e $v_i(\boldsymbol{\theta})$ são fixos ($i = 1, \dots, n$). Na prática, esses parâmetros são substituídos pelas suas respectivas estimativas.

6.4.3 Resíduos padronizados

Cálculos diretos mostram que a matriz de variância-covariância de $\hat{\mathbf{e}}_m$ e $\hat{\mathbf{e}}_c$, respectivamente, é dada por

$$\begin{aligned} \text{Cov}(\hat{\mathbf{e}}_m) &= (\mathbf{I}_{n^*} - \bar{\mathbf{H}}) \text{Cov}(\mathbf{y}) (\mathbf{I}_{n^*} - \bar{\mathbf{H}})^T \quad \text{e} \\ \text{Cov}(\hat{\mathbf{e}}_c) &= (\mathbf{I}_{n^*} - \mathbf{H}) \text{Cov}(\mathbf{y}) (\mathbf{I}_{n^*} - \mathbf{H})^T, \end{aligned}$$

em que $\text{Cov}(\mathbf{y}) = \kappa \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_n)$, com $\kappa = \nu/(\nu - 2)$. Se substituirmos Σ_i e $v_i(\boldsymbol{\theta})$ pelas suas estimativas, isto é $\hat{\Sigma}_i$ e $v_i(\hat{\boldsymbol{\theta}})$, obtemos uma aproximação da matriz de variância-covariância de $\hat{\mathbf{e}}_m$ e $\hat{\mathbf{e}}_c$. Assim, podemos definir uma versão padronizada dos resíduos marginais e condicionais da seguinte maneira:

$$\hat{\mathbf{e}}_{m_l}^* = \frac{\hat{\mathbf{e}}_{m_l}}{\sqrt{\widehat{\text{Var}}(\hat{\mathbf{e}}_{m_l})}} \quad \text{e} \quad (6.27)$$

$$\hat{\mathbf{e}}_{c_l}^* = \frac{\hat{\mathbf{e}}_{c_l}}{\sqrt{\widehat{\text{Var}}(\hat{\mathbf{e}}_{c_l})}}, \quad (6.28)$$

em que $\widehat{\text{Var}}(\widehat{\mathbf{e}}_{m_l})$ ($\widehat{\text{Var}}(\widehat{\mathbf{e}}_{c_l})$) é o l -ésimo elemento da diagonal principal da matriz $\widehat{\text{Cov}}(\widehat{\mathbf{e}}_m)$ ($\widehat{\text{Cov}}(\widehat{\mathbf{e}}_c)$) e denota a variância da l -ésima observação, para $l = 1, \dots, n^*$ e $n^* = \sum_{i=1}^n m_i$. Os primeiros m_1 elementos do vetor $\widehat{\mathbf{e}}_c^*$ correspondem às observações do primeiro grupo, e assim por diante. A análise de diagnóstico baseada nos resíduos marginal e condicional deve considerar os pesos $v_i(\boldsymbol{\theta})$'s pois, eventualmente, podem existir observações aberrantes que apresentem um peso pequeno no processo inferencial, não representando, portanto, um problema relevante na análise. Outros detalhes relacionados a análise de resíduos são discutidos por Verbeke e Molenberghs (2001) e Nobre (2003) para os modelos mistos normal, Savalli (2006) para os modelos mistos elípticos, Fung et al. (2002) para os modelos mistos semiparamétricos normal, e Kim et al. (2002) e Zhu et al. (2003) para os modelos lineares parciais com dados independentes.

6.5 Validação cruzada generalizada

Como foi dito na Subseção 3.2.1, o parâmetro de suavização regula a relação entre a fidelidade dos dados e a suavidade da curva estimada, e portanto a escolha de um valor apropriado para esse parâmetro é um problema essencial que deve ser considerado. Baseados em Green e Silverman (1994), Seção 4.4, podemos definir o escore de validação cruzada generalizada para os modelos mistos semiparamétricos t-Student multivariados na forma

$$\text{VCG}(\alpha) = \frac{\|\mathbf{y} - \widehat{\mathbf{y}}\|^2}{(1 - n^{-1} \text{tr}\{\mathbf{H}(\alpha)\})^2}, \quad (6.29)$$

em que o vetor de valores ajustados $\widehat{\mathbf{y}}$ é definido na equação (6.22) e a matriz $\mathbf{H}(\alpha) = \mathbf{H}$ na Subseção 6.4.2. Conseqüentemente, escolhemos o valor de α que minimize o escore $\text{VCG}(\alpha)$. Note que o critério de validação cruzada generalizada está condicionado ao fato de que a matriz de escala $\boldsymbol{\Sigma}_i$ e os pesos $v_i(\boldsymbol{\theta})$ são fixos. Na prática, a matriz $\boldsymbol{\Sigma}_i$ e os pesos $v_i(\boldsymbol{\theta})$ devem ser substituídos pelas correspondentes estimativas.

6.6 Qualidade do ajuste: distâncias transformadas

Como uma forma de avaliar os ajustes dos modelos normal e **t**-Student multivariados construiremos os gráficos das distâncias transformadas sugeridos por Little (1988) e Lange et al. (1989). Para o caso normal multivariado, temos que a distância de Mahalanobis,

$$\delta_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \quad (6.30)$$

com $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \phi \mathbf{I}_{m_i}$, segue uma distribuição $\chi_{m_i}^2$, e que $\widehat{\delta}_i = (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i)^T \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i)$ também segue, assintoticamente, a mesma distribuição de δ_i ($i = 1, \dots, n$). Logo, usando a aproximação de Wilson-Hilferty (veja Galea, 1995), temos que as distâncias transformadas

$$d_i^{[N]} = \frac{\{(\widehat{\delta}_i/m_i)^{\frac{1}{3}} - (1 - \frac{2}{9m_i})\}}{(\frac{2}{9})^{\frac{1}{2}}}, \quad (6.31)$$

seguem aproximadamente uma distribuição normal padrão ($i = 1, \dots, n$). Para o modelo **t**-Student multivariado, temos que $F_i = \delta_i/m_i$ segue uma distribuição $\mathcal{F}_{(m_i, \nu)}$. Além disso, $\widehat{F}_i = \widehat{\delta}_i/m_i$ tem assintoticamente a mesma distribuição de F_i ($i = 1, \dots, n$). Analogamente ao caso normal, usando a aproximação de Wilson-Hilferty temos que as distâncias transformadas

$$d_i^{[t]} = \frac{(1 - \frac{2}{9m_i})\widehat{F}_i^{1/3} - (1 - \frac{2}{9m_i})}{\left\{(\frac{2}{9m_i})\widehat{F}_i^{2/3} + (\frac{2}{9m_i})\right\}^{1/2}}, \quad (6.32)$$

seguem aproximadamente uma distribuição normal padrão ($i = 1, \dots, n$). Gráficos normais de probabilidade das distâncias transformadas $d_i^{[N]}$ e $d_i^{[t]}$ podem ser utilizados para avaliar os ajustes dos modelos normal e **t**-Student multivariado. Note que m_i corresponde ao número de medições associadas à i -ésima unidade amostral.

6.7 Diagnóstico de influência local

Nesta seção apresentamos a matriz de informação observada penalizada $-\ddot{\mathbf{L}}_p(\boldsymbol{\theta})$ e a matriz de perturbações penalizada $\Delta_p(\boldsymbol{\theta})$. Os esquemas de perturbação aqui considerados são ponderação de casos, perturbação na matriz de escala, perturbação aditiva na covariável associada ao componente paramétrico do modelo, e perturbação na variável resposta.

6.7.1 Matriz de Informação observada

Desde (6.5) temos que o logaritmo da função de verossimilhança penalizada é dado por

$$L_p(\boldsymbol{\theta}, \alpha) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \alpha),$$

em que a função $L_{p_i}(\boldsymbol{\theta}, \alpha)$ é definida na equação (6.8). A matriz de informação observada penalizada é dada por

$$-\ddot{\mathbf{L}}_p(\boldsymbol{\theta}) = -\sum_{i=1}^n \ddot{\mathbf{L}}_{p_i}(\boldsymbol{\theta}), \quad (6.33)$$

com

$$\ddot{\mathbf{L}}_{p_i}(\boldsymbol{\theta}) = \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \ddot{\mathbf{L}}_{p_{11},i} & \ddot{\mathbf{L}}_{p_{12},i} & \ddot{\mathbf{L}}_{p_{13},i} \\ \ddot{\mathbf{L}}_{p_{12},i}^T & \ddot{\mathbf{L}}_{p_{22},i} & \ddot{\mathbf{L}}_{p_{23},i} \\ \ddot{\mathbf{L}}_{p_{13},i}^T & \ddot{\mathbf{L}}_{p_{23},i}^T & \ddot{\mathbf{L}}_{p_{33},i} \end{pmatrix},$$

em que

$$\begin{aligned} \ddot{\mathbf{L}}_{p_{11},i} &= \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 2 \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[2 W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i, \\ \ddot{\mathbf{L}}_{p_{22},i} &= \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \mathbf{f} \partial \mathbf{f}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 2 \mathbf{N}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[2 W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{N}_i - \frac{\alpha}{n} \mathbf{K}, \\ \ddot{\mathbf{L}}_{p_{12},i} &= \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\beta} \partial \mathbf{f}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 2 \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \left[2 W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \right] \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{N}_i, \end{aligned}$$

$$\begin{aligned}\ddot{\mathbf{L}}_{\mathbf{P}_{33},i} &= \left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \\ \ddot{\mathbf{L}}_{\mathbf{P}_{13},i} &= \left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\tau}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{e} \\ \ddot{\mathbf{L}}_{\mathbf{P}_{23},i} &= \left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \alpha)}{\partial \mathbf{f} \partial \boldsymbol{\tau}^T} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},\end{aligned}$$

cujos elementos são, respectivamente, da forma

$$\begin{aligned}\left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_j \partial \tau_\ell} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{2} \text{tr} \left\{ \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[\dot{\boldsymbol{\Sigma}}(j) \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(\ell) - \ddot{\boldsymbol{\Sigma}}(j, \ell) \right] \right\} \\ &+ \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[W'(\widehat{\delta}_i) \dot{\boldsymbol{\Sigma}}(j) \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(\ell) - W(\widehat{\delta}_i) \ddot{\boldsymbol{\Sigma}}(j, \ell) \right. \\ &\left. + W(\widehat{\delta}_i) \dot{\boldsymbol{\Sigma}}(j) \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(\ell) + W(\widehat{\delta}_i) \dot{\boldsymbol{\Sigma}}(\ell) \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i,\end{aligned}$$

$$\left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\beta} \partial \tau_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 2 \mathbf{X}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[W'(\widehat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T + W(\widehat{\delta}_i) \widehat{\boldsymbol{\Sigma}}_i \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i$$

e

$$\left. \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \alpha)}{\partial \mathbf{f} \partial \tau_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 2 \mathbf{N}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \left[W'(\widehat{\delta}_i) \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^T + W(\widehat{\delta}_i) \widehat{\boldsymbol{\Sigma}}_i \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{\varepsilon}}_i,$$

com

$$W(\delta_i) = -\frac{1}{2} \left\{ \frac{\nu + m_i}{\nu + \delta_i} \right\} \quad \text{e} \quad W'(\delta_i) = \frac{1}{2} \left\{ \frac{\nu + m_i}{(\nu + \delta_i)^2} \right\},$$

$\widehat{\delta}_i = \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{\boldsymbol{\Sigma}}_i \widehat{\boldsymbol{\varepsilon}}_i$, $\widehat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} - \mathbf{N}_i \widehat{\mathbf{f}}$, e $\widehat{\boldsymbol{\Sigma}}_i = \mathbf{Z}_i \mathbf{D}(\widehat{\boldsymbol{\lambda}}) \mathbf{Z}_i^T + \widehat{\phi} \mathbf{I}_{n_i}$ ($i = 1, \dots, n$). Note que as matrizes $\dot{\boldsymbol{\Sigma}}(j) = \partial \boldsymbol{\Sigma}_i / \partial \tau_j$, $\dot{\boldsymbol{\Sigma}}(\ell) = \partial \boldsymbol{\Sigma}_i / \partial \tau_\ell$ e $\ddot{\boldsymbol{\Sigma}}(j, \ell) = \partial \boldsymbol{\Sigma}_i / \partial \tau_j \partial \tau_\ell$, são avaliadas em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ ($j, \ell = 0, \dots, d_\lambda$).

6.7.2 Esquemas de perturbação

A matriz $\Delta_p(\boldsymbol{\theta})$ para cada esquema de perturbação é definida na forma

$$\Delta_p(\boldsymbol{\theta}) = \frac{\partial^2 L_p(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \quad (6.34)$$

em que $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança penalizada e $\boldsymbol{\omega}_0$ o vetor de não perturbação. A seguir são apresentadas as expressões da matriz $\Delta_p(\boldsymbol{\theta})$ para os três esquemas de perturbação abordados neste capítulo.

Ponderação de casos

Se nosso objetivo é avaliar a contribuição individual das observações, podemos atribuir diferentes ponderações a cada uma delas no logaritmo da função de verossimilhança penalizada. Neste caso, o logaritmo da função de verossimilhança penalizada para o modelo perturbado é dado por

$$L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i L_i(\boldsymbol{\theta}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}^T, \quad (6.35)$$

em que $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ é o vetor de pesos, com $0 \leq \omega_i \leq 1$ ($i = 1, \dots, n$). Nesse caso, o vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^T \in \mathbb{R}^n$. Usando resultados de diferenciação sobre $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$ obtemos

$$\begin{aligned} \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2W(\hat{\delta}_i) \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\epsilon}}_i, \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -2W(\hat{\delta}_i) \mathbf{N}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\epsilon}}_i \quad \text{e} \\ \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \tau_j \partial \omega_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\frac{1}{2} \text{tr} \{ \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \} - W(\hat{\delta}_i) \hat{\boldsymbol{\epsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\epsilon}}_i, \end{aligned}$$

com $\dot{\boldsymbol{\Sigma}}(j)$ avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ ($i = 1, \dots, n$ e $j = 0, \dots, d_\lambda$).

Perturbação na matriz de escala

A perturbação na matriz de escala é introduzida ao considerar

$$\mathbf{y}_i \sim \mathbf{t}_{m_i}(\boldsymbol{\mu}_i, \omega_i^{-1} \boldsymbol{\Sigma}_i, \nu), \quad (6.36)$$

em que $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$, com $\omega_i > 0$ ($i = 1, \dots, n$). Nesse caso, o vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^T \in \mathbb{R}^n$ tal que $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}) = L_p(\boldsymbol{\theta}, \alpha)$. Após algumas manipulações algébricas obtemos

$$\begin{aligned} \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \omega=\omega_0} &= -2 \{W'(\hat{\delta}_i) \hat{\delta}_i + W(\hat{\delta}_i)\} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i, \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \omega=\omega_0} &= -2 \{W'(\hat{\delta}_i) \hat{\delta}_i + W(\hat{\delta}_i)\} \mathbf{N}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i \quad \text{e} \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \tau_j \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \omega=\omega_0} &= -\{W'(\hat{\delta}_i) \hat{\delta}_i + W(\hat{\delta}_i)\} \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i, \end{aligned}$$

com $\dot{\boldsymbol{\Sigma}}(j)$ avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ ($i = 1, \dots, n$ e $j = 0, \dots, d_\lambda$).

Perturbação na variável explicativa

Aqui o interesse principal é perturbar uma variável explicativa contínua específica na forma $\mathbf{x}_{i\omega} = \mathbf{x}_{il} + \boldsymbol{\omega}_i$, em que \mathbf{x}_{il} corresponde à l -ésima coluna da matriz \mathbf{X}_i e $\boldsymbol{\omega}_i$ é um vetor ($m_i \times 1$) de perturbações. Nesse caso, o vetor de não perturbação é $\boldsymbol{\omega}_0 = (0, \dots, 0)^T \in \mathbb{R}^{n^*}$, com $n^* = \sum_{i=1}^n m_i$, e o logaritmo da função de verossimilhança penalizada para o modelo perturbado é dado por

$$L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}) = L(\boldsymbol{\theta} | \boldsymbol{\omega}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}^T,$$

em que

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

sendo $L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) \propto \log \{1 + \nu^{-1} \delta_{i\omega}\}^{-\left(\frac{m_i + \nu}{2}\right)}$, $\delta_{i\omega} = \boldsymbol{\varepsilon}_{i\omega}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_{i\omega}$, $\boldsymbol{\varepsilon}_{i\omega} = \mathbf{y}_i - \mathbf{X}_{i\omega} \boldsymbol{\beta} - \mathbf{N}_i \mathbf{f}$ e $\mathbf{X}_{i\omega} = \mathbf{X}_i + \boldsymbol{\omega}_i \mathbf{z}_i^T$. Usando resultados de diferenciação sobre $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$ obtemos

$$\begin{aligned}
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= 4W'(\hat{\delta}_i) \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\beta}}_l + 2W(\hat{\delta}_i) \{ \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_l - \mathbf{z}_l \hat{\boldsymbol{\varepsilon}}_i^T \} \hat{\boldsymbol{\Sigma}}_i^{-1}, \\
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{N}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{ 4W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \} \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\beta}}_l \quad \text{e} \\
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \tau_j \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= 2\hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \{ 2W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \} \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\beta}}_l,
\end{aligned}$$

com $\dot{\boldsymbol{\Sigma}}(j)$ avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ ($i = 1, \dots, n$ e $j = 0, \dots, d_\lambda$). Aqui \mathbf{z}_l denota um vetor $(p \times 1)$ com um 1 na l -ésima posição e zeros nas outras posições. $\hat{\boldsymbol{\beta}}_l$ é o l -ésimo elemento $\hat{\boldsymbol{\beta}}$.

Perturbação na variável resposta

Através deste esquema de perturbação desejamos avaliar a sensibilidade das estimativas quando são introduzidas pequenas perturbações nos componentes de cada vetor de respostas. Seja $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{im_i})^T \in \mathbb{R}^{m_i}$ o vetor de perturbação, e $\boldsymbol{\omega}_0 = \mathbf{0}$ ($\in \mathbb{R}^{n^*}$) o vetor de não perturbação. Consideremos a perturbação do vetor de respostas na forma $\mathbf{y}_{i\omega} = \mathbf{y}_i + \boldsymbol{\omega}_i$. Então, o logaritmo da função de verossimilhança penalizada do modelo (2.12) perturbado é dado por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) = \sum_{i=1}^n L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f},$$

em que $L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log g(\delta_{i\omega})$, com $\delta_{i\omega} = \boldsymbol{\varepsilon}_{i\omega}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_{i\omega}$, $\boldsymbol{\varepsilon}_{i\omega} = \mathbf{y}_{i\omega} - \boldsymbol{\mu}_i$. Diferenciando $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})$ em relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}_i$, obtemos que

$$\begin{aligned}
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{ 4W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \} \hat{\boldsymbol{\Sigma}}_i^{-1}, \\
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\mathbf{N}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \{ 4W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \} \hat{\boldsymbol{\Sigma}}_i^{-1} \quad \text{e} \\
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \tau_j \partial \boldsymbol{\omega}_i^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\hat{\boldsymbol{\varepsilon}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}(j) \hat{\boldsymbol{\Sigma}}_i^{-1} \{ 2W'(\hat{\delta}_i) \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^T + 2W(\hat{\delta}_i) \hat{\boldsymbol{\Sigma}}_i \} \hat{\boldsymbol{\Sigma}}_i^{-1},
\end{aligned}$$

com $\tilde{\Sigma}(j)$ avaliada em $\theta = \hat{\theta}$ ($i = 1, \dots, n$ e $j = 0, \dots, d_\lambda$).

6.8 Conclusões do capítulo

Neste capítulo foi apresentado o modelo misto semiparamétrico **t**-Student multivariado. Assumindo que o vetor de respostas e o vetor dos efeitos aleatórios de cada grupo segue uma distribuição conjunta **t**-Student multivariado derivamos a distribuição marginal do vetor de respostas observadas. A partir do logaritmo da função de verossimilhança penalizada do modelo marginal calculamos as funções escore e as matrizes de informação de Fisher penalizada associadas aos parâmetros do modelo. Baseados nesses resultados, construímos o processo iterativo escore de Fisher para estimar o coeficiente de regressão e a função não paramétrica. Os componentes de variância e os efeitos aleatórios foram estimados através do algoritmo escore de Fisher e a estimativa empírica de Bayes, respectivamente, e os erros padrão das estimativas dos parâmetros mediante a inversa da matriz de informação de Fisher. Dois tipos de resíduos padronizados foram definidos e conseqüentemente estudamos algumas de suas propriedades. Também consideramos o problema de estimação do parâmetro de suavização e sugerimos o critério de validação cruzada generalizada como um procedimento de estimação razoável. Finalmente, particularizamos os resultados de influência local apresentados no Capítulo 4 para o modelo misto semiparamétrico **t**-Student multivariado.

Capítulo 7

Aplicações

Neste capítulo apresentamos algumas aplicações dos modelos discutidos nos capítulos anteriores. Três conjuntos de dados reais são utilizados para ilustrar a metodologia de influência local nos modelos mistos aditivos semiparamétricos elípticos.

7.1 Dados AFP

Nesta seção discutimos uma aplicação que envolve um conjunto de dados reais fornecidos pela Superintendência de AFPs do Chile, em que as variáveis resposta correspondem às rentabilidades mensais da administradora de fundos de pensões *Cuprum*, registradas durante o período de janeiro de 1990 a dezembro de 2003. A variável explicativa corresponde à rentabilidade do indicador macroeconômico IPSA. Inicialmente, poderíamos ajustar um modelo linear simples para examinar o efeito das mudanças no valor do IPSA sobre as rentabilidades. A Figura 7.1a mostra o diagrama de dispersão entre essas duas variáveis e como resultado, podemos observar fortes indícios de uma tendência linear, com algumas observações mostrando um comportamento atípico. Como as rentabilidades foram registradas através do tempo, uma alternativa atrativa seria considerar um modelo que permita modelar a dependência da rentabilidade com relação ao tempo. A Figura 7.1b mostra o diagrama de dispersão das rentabilidades contra o tempo (meses) indicando que tal dependência pode ser representada em uma forma não paramétrica.

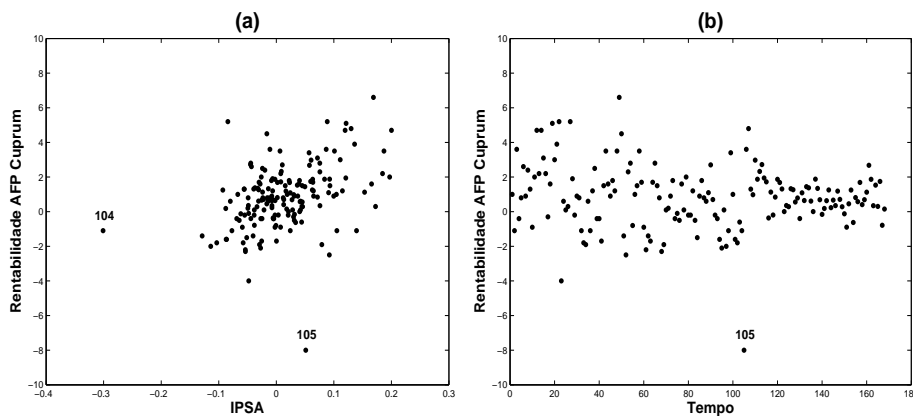


Figura 7.1: Diagrama de dispersão: rentabilidades da AFP *Cuprum* contra as rentabilidades do IPSA (a), rentabilidades contra o tempo (b).

7.1.1 Modelo proposto

Consideraremos o seguinte modelo semiparamétrico:

$$y_i = x_i \beta + f(t_i) + \epsilon_i, \quad (7.1)$$

em que y_i denota o valor observado da rentabilidade no tempo t_i , x_i denota a rentabilidade do IPSA no mesmo período, β é o parâmetro associado ao coeficiente de regressão que determina o incremento no valor da rentabilidade quando o IPSA aumenta uma unidade, f é uma função arbitrária que depende do tempo e que usualmente é estimada através de um processo de suavização, e ϵ_i é um erro aleatório ($i = 1, \dots, 168$). O modelo (7.1) pode ser expresso na forma

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{N}f + \boldsymbol{\epsilon}, \quad (7.2)$$

em que \mathbf{y} é um vetor (168×1) que contém as respostas observadas, \mathbf{X} é um vetor (168×1) cujos elementos correspondem aos valores da variável explicativa x , \mathbf{N} é uma matriz de incidência (168×168) que, neste caso, é igual a uma matriz

identidade, \mathbf{f} é um vetor (168×1) cujos componentes correspondem à avaliação funcional de $f(\cdot)$ nos valores da variável explicativa t que pertencem ao conjunto $\{t_1 = 1, t_2 = 2, \dots, t_{168} = 168\}$, e ϵ representa um vetor de erros aleatórios (168×1). Usualmente, assume-se que os erros aleatórios são não correlacionados através do tempo e que seguem uma distribuição normal. Porém, é sabido que as estimativas de máxima verossimilhança derivados do modelo normal são sensíveis a observações aberrantes. Nesse caso, uma alternativa é assumir um modelo de caudas mais pesadas para acomodar tais observações. Em virtude disso, vamos supor que o valor observado da rentabilidade no tempo t_i ($i = 1, \dots, 168$) segue uma distribuição na forma,

$$y_i \sim \mathcal{E}l_1(\mathbf{x}_i \beta + f(t_i), \phi).$$

Dessa forma, o vetor de parâmetros a ser estimado é $\boldsymbol{\theta} = (\beta, \mathbf{f}^T, \phi)^T$. A fim de avaliar e comparar a sensibilidade das estimativas de máxima verossimilhança penalizada a observações aberrantes, vamos considerar as distribuições normal e t-Student.

7.1.2 Ajustando os modelos normal e t-Student univariado

Os modelos foram ajustados usando a verossimilhança penalizada sob erros normal e t-Student. O parâmetro de suavização α foi estimado através do método de validação cruzada generalizada que, sob o modelo linear parcial com erros normais e ϕ conhecido, é $\hat{\alpha} = 100$. Utilizando o critério de informação de Schwarz descrito na Seção 3.8 para escolher os graus de liberdade da distribuição t-Student, obtivemos $\nu = 4$; veja a Tabela 7.1. Os erros padrão dos estimadores do coeficiente de regressão, da função não paramétrica e do parâmetro de escala, foram calculados a partir da matriz de informação de Fisher penalizada. Os resultados dos ajustes são apresentados na Tabela 7.2. Devido à alta dimensão do vetor \mathbf{f} , as estimativas e os erros padrão sob o modelo normal e t-Student não são apresentadas. Os ajustes não paramétricos para ambos os modelos são exibidos na Figura 7.2.

Tabela 7.1: Valores estimados do logaritmo da função de verossimilhança penalizada e do critério de informação de Schwarz sob o modelo t-Student para diferentes graus de liberdade ajustado aos dados de AFP.

ν	$-2L_p(\hat{\boldsymbol{\theta}}, \hat{\alpha})$	$SIC(\hat{\boldsymbol{\theta}})$
1	617,34	1488,4
2	615,26	1486,3
3	601,82	1472,9
4	600,34	1471,4
5	600,98	1472,1

Tabela 7.2: Estimativas de máxima verossimilhança penalizada para os modelos normal e t-Student com $\nu = 4$ graus de liberdade ajustados aos dados de AFP.

Parâmetro	Normal			t-Student		
	Estimativa	E.P.	$L_p(\hat{\boldsymbol{\theta}}, \hat{\alpha})$	Estimativa	E.P.	$L_p(\hat{\boldsymbol{\theta}}, \hat{\alpha})$
β	7,924	1,961	-315,32	7,752	1,876	-300,17
ϕ	2,433	0,045		1,193	0,121	

Da Tabela 7.2 podemos notar que as estimativas do coeficiente de regressão são similares entre os dois modelos ajustados. Porém, o erro padrão de $\hat{\beta}$ sob o modelo t-Student é menor do que o erro padrão obtido sob o modelo normal, indicando que o modelo com caudas mais pesadas parece fornecer estimativas mais precisas para o coeficiente de regressão. Com relação ao parâmetro de escala dos modelos, temos que as estimativas são diferentes entre os modelos ajustados e que o erro padrão de $\hat{\phi}$ sob o modelo normal é menor do que o erro padrão sob para o modelo t-Student. Porém, tais resultados não são comparáveis.

Como uma forma de avaliar os ajustes dos modelos normal e t-Student construiremos os gráficos das distâncias transformadas sugeridos por Little (1988) e Lange et al. (1989) no contexto multivariado e que podem ser particularizados diretamente para o caso univariado. Baseados nos valores de $L_p(\hat{\boldsymbol{\theta}}, \hat{\alpha})$ e nos gráficos normais de probabilidade das distâncias transformadas apresentados na Figura

7.3, temos que o modelo t-Student com 4 graus de liberdade parece apresentar um ajuste mais adequado em relação ao modelo normal. Salientamos que esses gráficos são comparáveis se a distância transformada for normal para ambos os modelos.

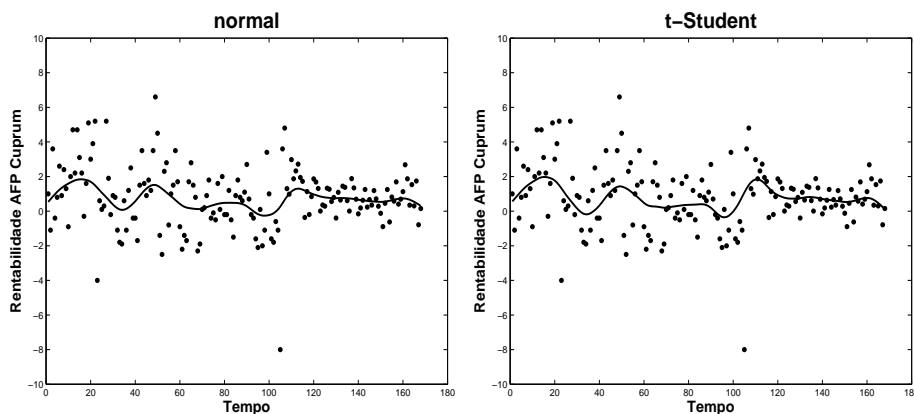


Figura 7.2: Gráficos dos ajustes não paramétricos (linha contínua) aos dados de AFP sob o modelo normal e t-Student com $\nu = 4$ graus de liberdade.

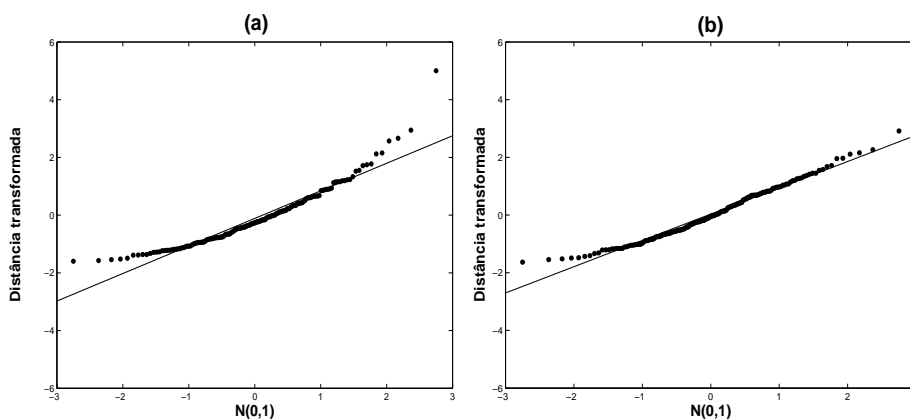


Figura 7.3: Gráficos normais de probabilidade para a distância transformada sob os modelos normal (a) e t-Student (b) ajustados aos dados de AFP.

7.1.3 Detectando observações aberrantes

Para detectar observações aberrantes observamos os gráficos de índices das distâncias entre o valor observado y_i e o seu valor ajustado \hat{y}_i , digamos $\hat{\delta}_i = \frac{(y_i - \hat{\mu}_i)^2}{\hat{\phi}}$.

Como resultado, a Figura 7.4 revela que sob os modelos normal e t-Student as observações 22, 23, 52 e 105 aparecem como possíveis observações aberrantes, com um forte destaque para a observação 105. A Figura 7.4c mostra o gráfico dos pesos estimados contra as distâncias $\hat{\delta}_i$ sob o modelo t-Student. Como é possível observar, as observações 22, 23, 52 e 105 receberam pesos menores no processo de estimação. Isso indica que as estimativas de máxima verossimilhança penalizada apresentam algumas sinais de robustez, no sentido da distância $\hat{\delta}_i$, contra observações aberrantes.

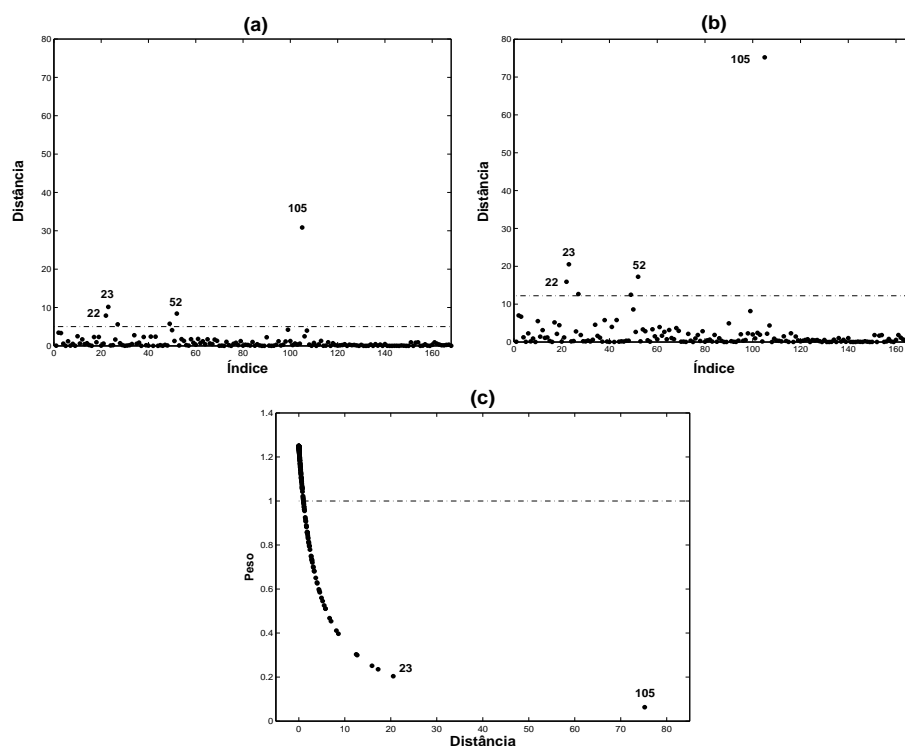


Figura 7.4: Gráficos de índices para a distância $\hat{\delta}_i$ sob os modelos normal (a) e t-Student (b) ajustados aos dados de AFP, e entre os pesos estimados e a distância $\hat{\delta}_i$ sob o modelo t-Student (c).

7.1.4 Diagnóstico de influência

Para detectar observações influentes e avaliar a sensibilidade das estimativas de máxima verossimilhança penalizada de β , f e ϕ sob alguns esquemas de perturbação, apresentamos os gráficos de índices da medida de influência local total para ponderação de casos, perturbação de escala, e perturbação na variável explicativa IPSA. A representação gráfica de $C_i(\theta)$ foi omitida, pelo fato de apresentar um comportamento similar aos gráficos obtidos da análise de influência local para f . Por simplicidade, condicionamos nossa análise de influência local assumindo que o parâmetro de suavização é fixo. Porém, é possível que exista algum grau de sensibilidade nos resultados a mudanças nesse parâmetro; vide Zhu et al. (2003).

Ponderação de casos

Nas Figuras 7.5, 7.6 e 7.7 são apresentados os gráficos de índices de $C_i(\beta)$, $C_i(f)$ e $C_i(\phi)$, para os modelos normal e t-Student. Tais Figuras indicam que sob o modelo normal a observação 22 é a mais influente em $\hat{\beta}$ seguida das observações 2, 23, 49, 105 e 107 que exercem uma influência moderada. Agora, sob o modelo t-Student, observamos que a observação 51 é a mais influente seguida das observações 2, 26, 98 e 107. É possível observar também que, sob o modelo normal, a observação 105 é altamente influente em \hat{f} e $\hat{\phi}$, mas, sob o modelo t-Student, não aparecem observações exercendo influência relevante.

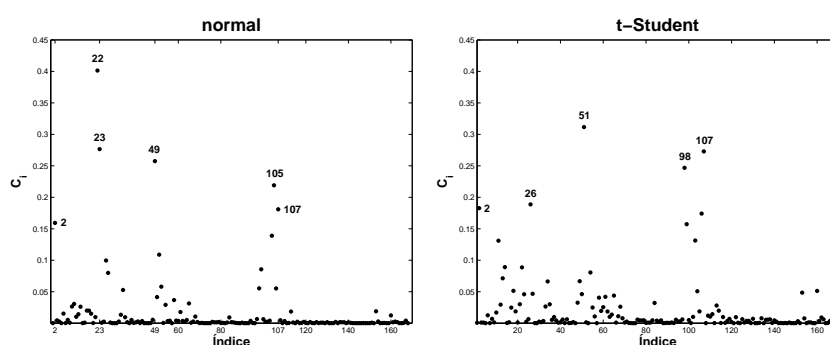


Figura 7.5: Gráficos de índices de C_i para $\hat{\beta}$ sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de AFP.

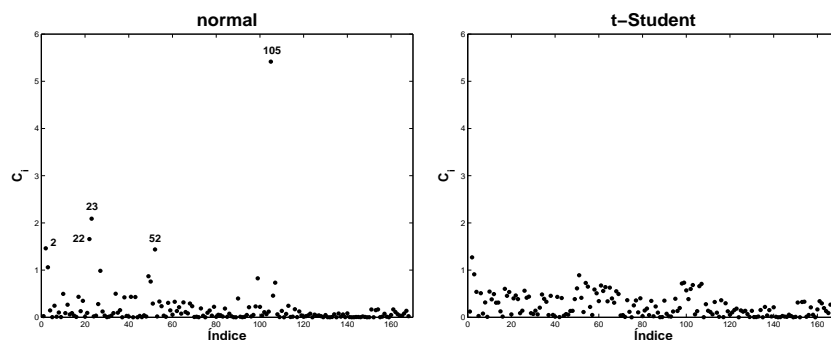


Figura 7.6: Gráficos de índices de C_i para \hat{f} sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de AFP.

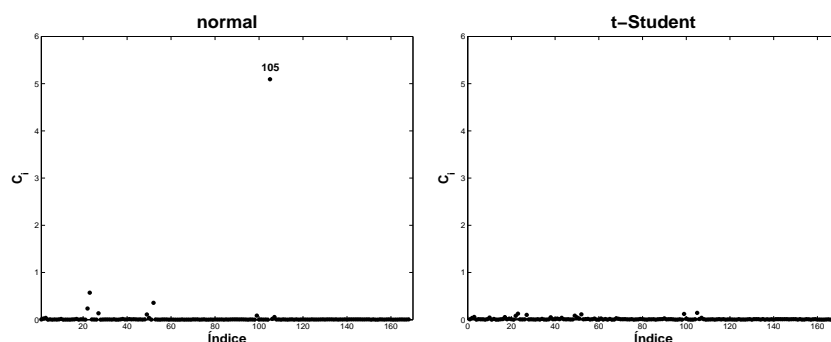


Figura 7.7: Gráficos de índices de C_i para $\hat{\phi}$ sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de AFP.

Perturbação de escala

Os gráficos de índices de $C_i(\beta)$, $C_i(f)$ e $C_i(\phi)$ para os modelos normal e t-Student, assumindo um esquema de perturbação de escala, são apresentados nas Figuras 7.8, 7.9 e 7.10. A partir dessas figuras podemos observar que os gráficos de influência local total sob o modelo normal e t-Student têm um comportamento similar aos obtidos para ponderação de casos, exceto que sob o modelo t-Student não há observações exercendo influência relevante. Portanto, tais comentários são válidos para este esquema de perturbação.

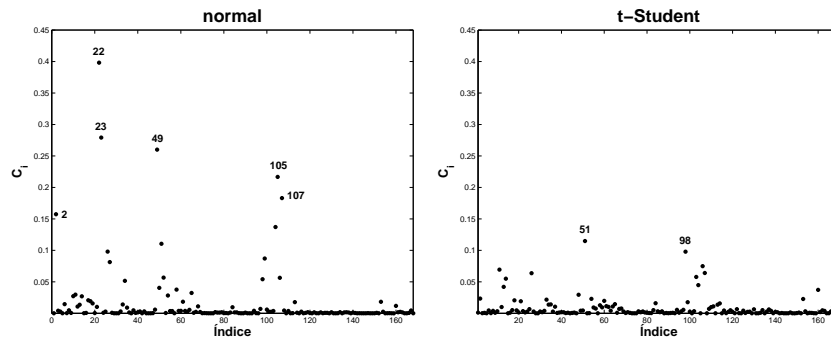


Figura 7.8: Gráficos de índices de C_i para $\hat{\beta}$ sob perturbação de escala para os modelos normal e t-Student ajustados aos dados de AFP.

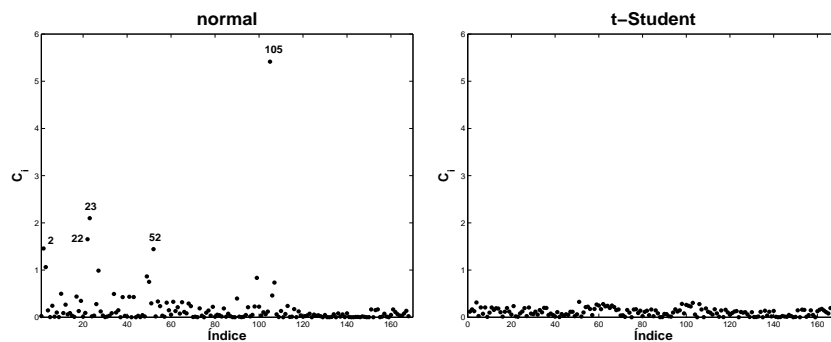


Figura 7.9: Gráficos de índices de C_i para \hat{f} sob perturbação de escala para os modelos normal e t-Student ajustados aos dados de AFP.

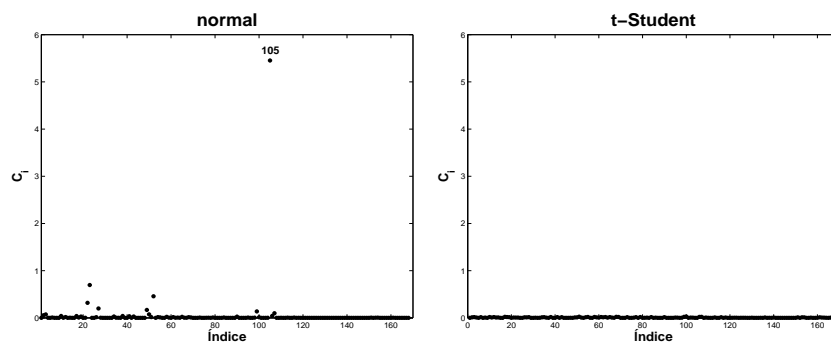


Figura 7.10: Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação de escala para os modelos normal e t-Student ajustados aos dados de AFP.

Perturbação no IPSA

Nas Figuras 7.11, 7.12 e 7.13 são apresentados os gráficos de índices de $C_i(\beta)$, $C_i(f)$ e $C_i(\phi)$, para os modelos normal e t-Student, no caso em que a variável explicativa é perturbada em forma aditiva. Tais figuras revelam que sob o modelo normal a observação 105 é a mais influente em $\hat{\beta}$ e $\hat{\phi}$, enquanto que sob o modelo t-Student não aparecem observações influentes. Além disso, é possível observar que sob o modelo normal não há observações influentes em \hat{f} , mas, sob o modelo t-Student, as observações 1 e 168 têm uma influência mais acentuada.

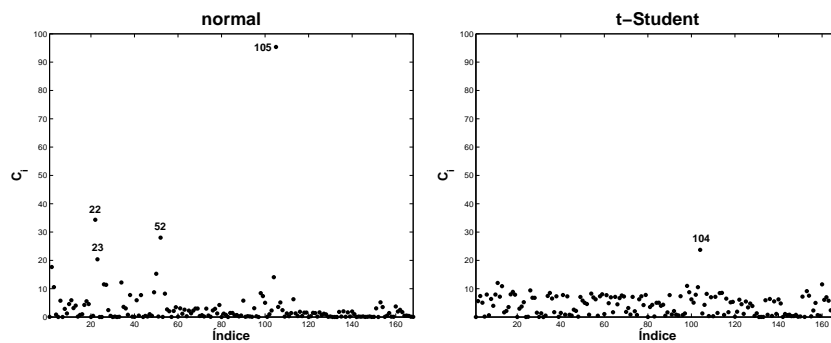


Figura 7.11: Gráficos de índices de C_i para $\hat{\beta}$ sob perturbação do IPSA para os modelos normal e t-Student ajustados aos dados de AFP.

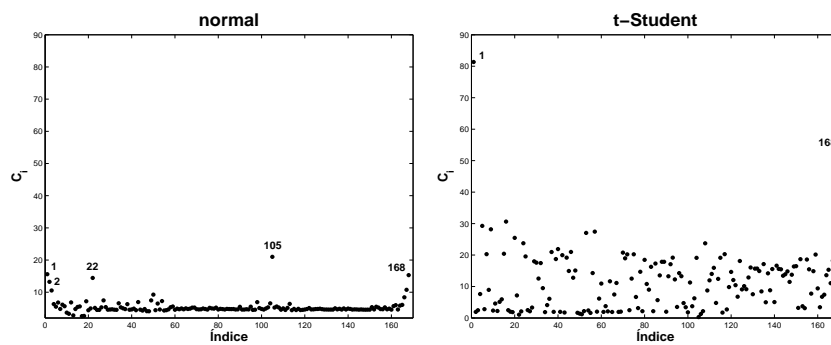


Figura 7.12: Gráficos de índices de C_i para \hat{f} sob perturbação do IPSA para os modelos normal e t-Student ajustados aos dados de AFP.

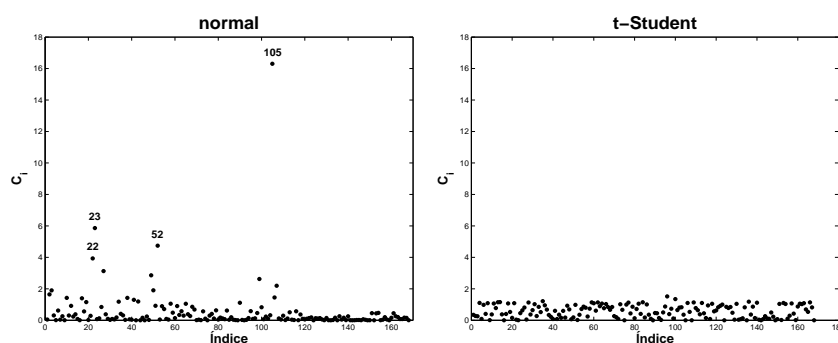


Figura 7.13: Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação do IPISA para os modelos normal e t-Student ajustados aos dados de AFP.

Com o objetivo de comparar os resultados de influência local, na Figura 7.14 apresentamos os gráficos de índices dos resíduos padronizados, dos pontos de alavanca, das distâncias de Cook para $\hat{\beta}$, e da medida DFIT para \hat{f} .

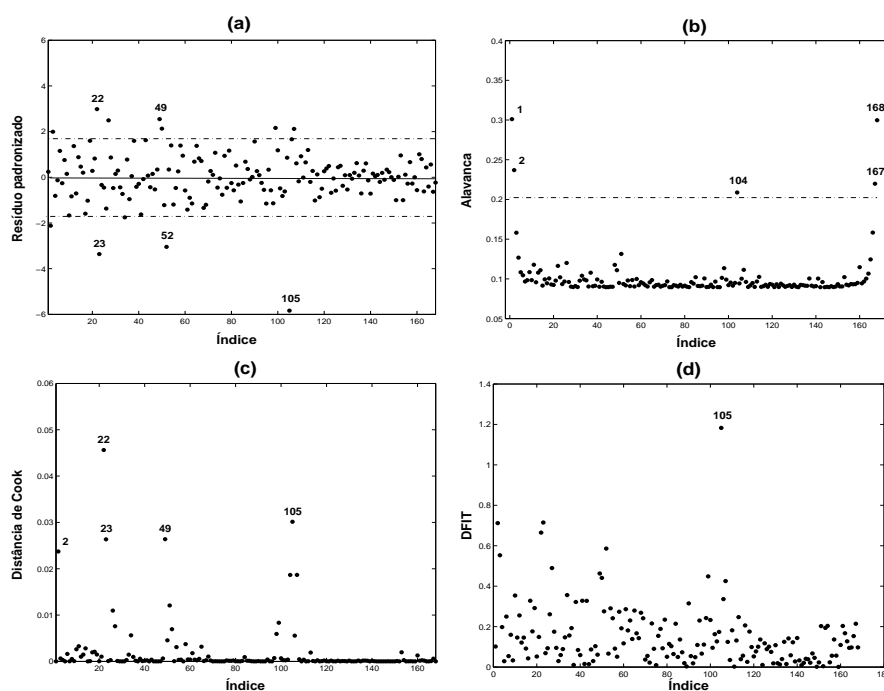


Figura 7.14: Gráficos de índices: resíduos padronizados (a), pontos de alavanca (b), distâncias de Cook para $\hat{\beta}$ (c) e medida DFITS para \hat{f} (d), referentes ao ajuste do modelo normal aos dados de AFP.

Como resultado, podemos observar a análise de diagnóstico de eliminação de casos apresentados na Figura 7.14 concordam com os obtidos por influência local. As expressões das medidas de eliminação de casos utilizadas acima são dadas em Fung et al. (2002) no contexto dos modelos mistos semiparamétricos gaussianos, e podem ser facilmente particularizadas para os modelos lineares parciais; veja também Eubank (1985) e Kim (1996). De acordo com Fung et al. (2002) é conveniente fazer o diagnóstico condicionando a estimação do parâmetro de escala e do parâmetro de suavização a todo o conjunto de dados. Na prática é possível que mudanças em $\hat{\phi}$ e $\hat{\alpha}$ devido à exclusão de uma observação tenham um efeito adicional nas estimativas de β e f .

7.1.5 Análise confirmatória

A seguir, é apresentada a análise confirmatória obtida reajustando o modelo (7.1) sem as observações 22, 23, 52 e 105 para avaliar o impacto nas estimativas de β e ϕ . Os valores apresentados entre parênteses correspondem às mudanças relativas (impacto percentual) na estimativa do parâmetro, quando eliminamos a correspondente observação. Aqui, $\hat{\beta}_{(I)}$ e $\hat{\phi}_{(I)}$ denotam respectivamente, as estimativas de máxima verossimilhança penalizada de β e ϕ obtidas após a eliminação da observação I . Na Tabela 7.3 observamos que a maior variação percentual na estimativa de máxima verossimilhança penalizada de β é obtida sob o modelo normal. Essa variação atinge um máximo quando eliminamos a observação 22. Porém, sob o modelo t-Student, a variação é menor quando eliminamos a mesma observação. Observamos ainda, que a maior variação percentual na estimativa de máxima verossimilhança de ϕ , sob o modelo normal, atinge um máximo quando eliminamos a observação 105. Já para o modelo t-Student, essa variação é bem menor quando eliminamos a mesma observação. Ao observar o conjunto de dados notamos que as observações 23, 52 e 105 correspondem aos menores valores das rentabilidades (negativas), enquanto as observações 22 e 49 correspondem aos maiores valores das rentabilidades (positivas) recebidas pela AFP. É evidente que tais observações têm um comportamento atípico.

Tabela 7.3: Mudanças relativas (em parênteses) nas estimativas de máxima verossimilhança penalizada para os modelos normal e t-Student com $\nu = 4$ graus de liberdade ajustados aos dados de AFP.

	normal		t-Student	
I	$\hat{\beta}_{(I)}$	$\hat{\phi}_{(I)}$	$\hat{\beta}_{(I)}$	$\hat{\phi}_{(I)}$
–	7,92	2,43	7,75	1,19
22	8,87 (11,94)	2,32 (4,55)	7,58 (2,23)	1,12 (5,67)
23	7,14 (9,84)	2,26 (6,91)	7,58 (2,25)	1,12 (6,38)
52	8,27 (4,41)	2,30 (5,40)	7,97 (2,84)	1,12 (6,43)
105	8,61 (8,63)	1,94 (20,20)	7,71 (0,54)	1,22 (2,61)

Finalmente, a partir dos resultados de influência local descritos acima, podemos concluir que as observações que exercem maior influência na estimativa de β não necessariamente têm o mesmo grau de influência na função estimada de f . Por exemplo, para ponderação de casos, as observações 22 e 105 são altamente influentes em $\hat{\beta}$ e \hat{f} , respectivamente. Além disso, notamos que a estimativa de f , derivada do modelo com erros t-Student, apresenta uma importante sensibilidade quando perturbamos a covariável IPSA. Por sua parte, a estimativa do coeficiente de regressão sob o modelo t-Student também mostra-se sensível quando atribuímos diferentes ponderações às observações.

7.2 Dados de glucose

Em um estudo médico foram aplicados diferentes testes padrão de tolerância à glucose a 20 pacientes obesos. Os dados correspondem às medições de fosfato inorgânico no plasma, obtidas a partir de amostras de sangue para cada paciente, registradas às 0; 0,5; 1; 1,5; 2; 3; 4 e 5 horas após a aplicação oral de uma dose de glucose padrão; veja Apêndice D. A variável resposta foi medida em mg/dl. Os

dados foram reportados por Zerbe (1979) e analisados por diferentes autores. Por exemplo, Reinsel (1984) ajusta esses dados a um modelo de curva de crescimento com efeitos aleatórios e sugere modelar a resposta média de cada paciente através de uma função linear por partes com ponto de mudança às duas horas (veja Figura 7.15a). Chi e Reinsel (1989) fazem uma análise desses dados através do ajuste de um modelo com efeitos aleatórios e erros AR(1). Mais recentemente, Pang e Fang (2002) realizam estudos de sensibilidade utilizando os procedimentos de eliminação de casos e influência local em modelos de curva de crescimento com erros normais. Para ilustrar a aplicabilidade de tais resultados eles propõem modelar os dados descritos acima através de um modelo de curva de crescimento, assumindo um polinômio de segundo grau. Nessa análise, eles confirmam a presença de observações aberrantes e influentes; ver Keramidias e Lee (1995).

7.2.1 Modelo proposto

A partir da Figura 7.15 onde são apresentados os perfis individuais de cada paciente, temos indícios de que os níveis de fosfato inorgânico no plasma, após da aplicação oral da dose de glucose, muda através do tempo em uma forma que torna difícil modelar sua tendência usando uma função paramétrica simples. Nesse caso, sugerimos analisar este conjunto de dados usando o seguinte modelo misto não paramétrico:

$$y_{ij} = f(t_{ij}) + b_i + \epsilon_{ij}, \quad (7.3)$$

em que y_{ij} representa a j -ésima medição de fosfato inorgânico no plasma do i -ésimo paciente no tempo t_{ij} ($i = 1, \dots, 20; j = 1, \dots, 8$), $f(\cdot)$ é uma função arbitrária que depende do tempo, b_i denota o efeito aleatório do i -ésimo paciente, e ϵ_{ij} é um erro aleatório dentro do paciente. O modelo (7.3) pode ser escrito na forma

$$\mathbf{y}_i = \mathbf{N}_i \mathbf{f} + \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \quad (7.4)$$

em que \mathbf{y}_i é um vetor (8×1) para as respostas do i -ésimo paciente, $\mathbf{N}_i = \mathbf{I}_8$ é uma matriz de incidência (8×8), onde \mathbf{I}_8 denota uma matriz identidade (8×8), \mathbf{f} é um

vetor (8×1) cujos componentes correspondem à avaliação funcional de $f(\cdot)$ nos valores dos tempos que pertencem ao conjunto $\mathbf{t}^0 = \{ t_1^0 = 0, t_2^0 = 0,5, \dots, t_8^0 = 5 \}$, \mathbf{Z}_i é a matriz de planejamento (8×1) para os efeitos aleatórios dada por

$$\mathbf{Z}_i = \left(1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \right)^T,$$

e $\boldsymbol{\epsilon}_i$ representa o vetor de erros aleatórios (8×1) . Assumindo que,

$$\begin{pmatrix} \mathbf{y}_i \\ b_i \end{pmatrix} \sim \mathcal{E}l_8 \left\{ \begin{pmatrix} \mathbf{f} \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda \mathbf{Z}_i \mathbf{Z}_i^T + \phi \mathbf{I}_8 & \lambda \mathbf{Z}_i \\ \lambda \mathbf{Z}_i^T & \lambda \end{pmatrix} \right\},$$

temos que o vetor de respostas associado ao i -ésimo paciente $(i = 1, \dots, 20)$ segue uma distribuição marginal na forma,

$$\mathbf{y}_i \sim \mathcal{E}l_8(\mathbf{f}, \lambda \mathbf{Z}_i \mathbf{Z}_i^T + \phi \mathbf{I}_8).$$

Sob essa suposição, o vetor de parâmetros a ser estimado é $\boldsymbol{\theta} = (\mathbf{f}^T, \lambda, \phi)^T$. Com o propósito de comparar o ajuste do modelo (7.3) para alguns membros da família das distribuições elípticas, vamos considerar especificamente as distribuições normal e t-Student multivariadas.

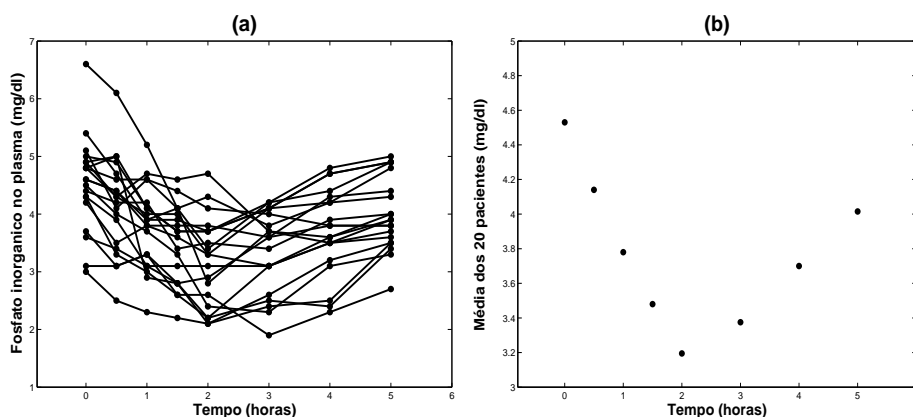


Figura 7.15: Gráficos de perfis (a) e das médias (b) de fosfato inorgânico no plasma.

7.2.2 Ajustando os modelos normal e t-Student multivariado

O modelo (7.3) assumindo uma distribuição normal e t-Student multivariada, respectivamente, foi ajustado usando a verossimilhança penalizada correspondente. O parâmetro de suavização $\hat{\alpha} = 0,105$ foi estimado através do método de validação cruzada generalizada assumindo normalidade. Os graus de liberdade da distribuição t-Student foram escolhidos mediante o critério de informação de Schwarz obtendo-se $\nu = 7$; veja Tabela 7.4. Os erros padrão do estimador do componente não paramétrico e do componente de variância foram calculados a partir da matriz de informação de Fisher penalizada. O resumo dos ajustes é apresentado na Tabela 7.5.

Tabela 7.4: Valores estimados do logaritmo da função de verossimilhança penalizada e do critério de informação de Schwarz sob o modelo t-Student multivariado para diferentes graus de liberdade ajustado aos dados de glucose.

ν	$-2L_p(\hat{\theta}, \hat{\alpha})$	$SIC(\hat{\theta})$
1	232,56	262,52
2	222,26	252,22
3	218,50	248,45
4	216,80	246,76
5	215,98	245,95
6	215,62	245,58
7	215,48	245,45
8	215,50	245,46

Da Tabela 7.5 podemos notar que as estimativas do componente não paramétrico são, em termos gerais, bastante similares entre os modelos ajustados. Porém, como os erros padrão de $\hat{\mathbf{f}}$ sob o modelo t-Student multivariado são menores que os erros padrão obtidos sob o modelo normal, temos indícios de que o modelo com caudas mais pesadas gera estimativas mais precisas para o componente não paramétrico. Com relação aos demais parâmetros, temos que as inferências para os componentes da escala são similares entre os modelos, porém não são comparáveis.

Tabela 7.5: Estimativas de máxima verossimilhança penalizada para os modelos normal e **t**-Student multivariado com $\nu = 7$ graus de liberdade ajustados aos dados de glicose.

Parâmetro	Normal			t-Student		
	Estimativa	E.P.	$L_p(\hat{\theta}, \hat{\alpha})$	Estimativa	E.P.	$L_p(\hat{\theta}, \hat{\alpha})$
λ	0,355	0,120	-110,92	0,297	0,112	-107,75
ϕ	0,163	0,020		0,120	0,022	
$f(t_1)$	4,530	0,161		4,496	0,145	
$f(t_2)$	4,140	0,163		4,110	0,146	
$f(t_3)$	3,782	0,164		3,780	0,146	
$f(t_4)$	3,476	0,164		3,490	0,146	
$f(t_5)$	3,200	0,162		3,216	0,145	
$f(t_6)$	3,374	0,160		3,403	0,144	
$f(t_7)$	3,700	0,159		3,736	0,143	
$f(t_8)$	4,015	0,160		4,047	0,144	

Para avaliar os ajustes dos modelos construiremos os gráficos das distâncias transformadas sugeridos na Seção 6.6. Baseados nos valores de $L_p(\hat{\theta}, \hat{\alpha})$ e nos gráficos normais de probabilidade das distâncias transformadas apresentados na Figura 7.16, temos que o modelo **t**-Student multivariado com 7 graus de liberdade parece apresentar um ajuste mais adequado com relação ao modelo normal.

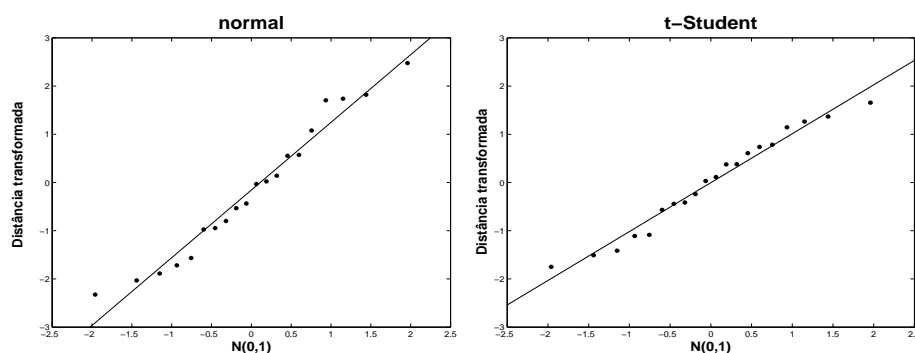


Figura 7.16: Gráficos normais de probabilidade das distâncias transformadas sob os modelos normal e **t**-Student com 7 graus de liberdade ajustados aos dados de glicose.

A Figura 7.17 mostra os valores ajustados obtidos a partir do modelo com efeito aleatório não paramétrico (7.3) para os 20 pacientes ($i = 1, \dots, 20$), $\hat{\mathbf{y}}_i = \hat{\mathbf{f}} + \mathbf{Z}_i \hat{\mathbf{b}}_i$, em que $\hat{\mathbf{b}}_i$ é o estimador empírico de Bayes com todos os parâmetros substituídos pelas suas estimativas de máxima verossimilhança penalizada derivadas do modelo t-Student. Em termos gerais, essa figura indica que as predições ajustadas para cada perfil do paciente são razoáveis quando consideramos um modelo que incorpora um efeito aleatório para modelar a estrutura de covariância do paciente e um componente não paramétrico para modelar o efeito do tempo. Chi e Reinsel (1989) obtêm resultados similares aos nossos, em que consideram um componente de efeitos aleatórios e uma função linear por parte, com ponto de mudança às duas horas, para modelar a tendência do tempo. Além disso, eles incorporam um processo autorregressivo AR(1) na modelagem do tempo. Nesse contexto, as contribuições de Zeger e Diggle (1994) e Zhang et al. (1998) podem ser consideradas.

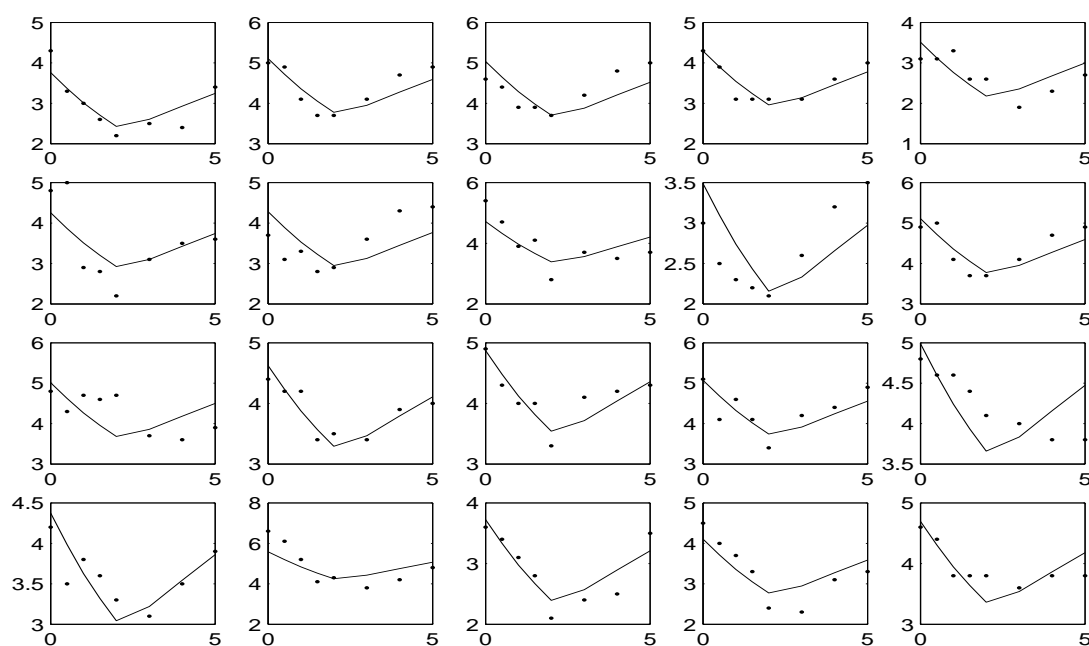


Figura 7.17: Gráfico de perfis ajustados (—) sob o modelo t-Student com 7 graus de liberdade para os 20 pacientes do estudo de dados de glucose (da esquerda para a direita).

7.2.3 Detectando observações aberrantes

Pang e Fang (2002) constataam que sob erros normais há observações aberrantes e influentes no conjunto de dados de glucose. No intuito de detectar observações aberrantes, apresentamos na Figura 7.18(a-b) os gráficos de índices das distâncias de Mahalanobis ajustadas $\hat{\delta}_i$ e das distâncias modificadas $\hat{F}_i = \hat{\delta}_i/m_i$, respectivamente, para os modelos normal e t-Student multivariados.

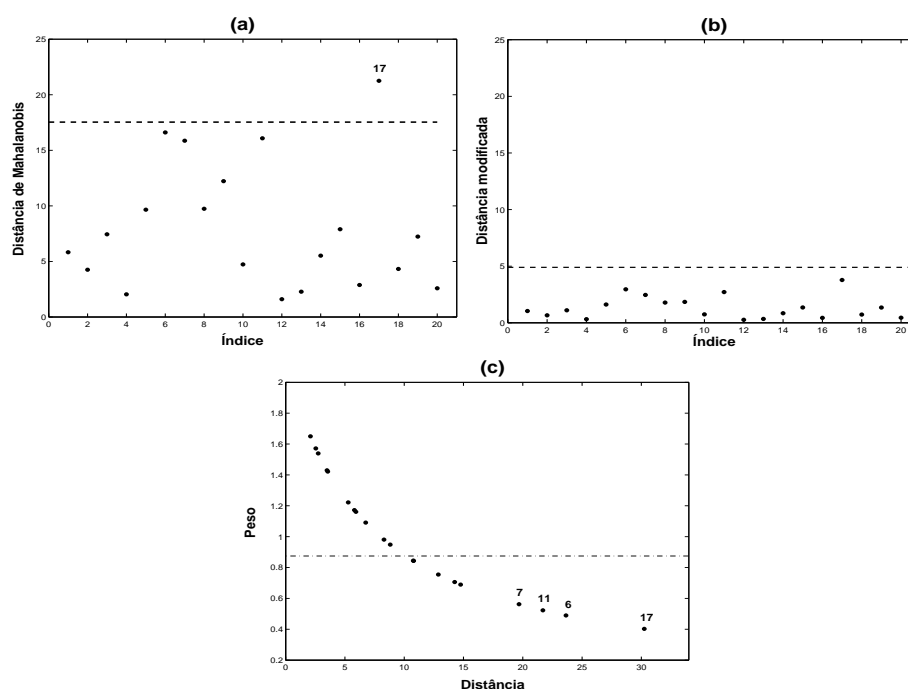


Figura 7.18: Gráficos de índices para a distância $\hat{\delta}_i$ sob os modelos normal (a) e t-Student (b) ajustados aos dados de glucose, e entre os pesos estimados e a distância $\hat{\delta}_i$ sob o modelo t-Student (c).

Como resultado identificamos o paciente 17 como uma possível observação aberrante quando ajustamos um modelo normal. A Figura 7.17c mostra o gráfico dos pesos estimados contra $\hat{\delta}_i$ sob o modelo t-Student. Como é possível observar, a observação 17 recebe um peso menor no processo de estimação. Isso indica que as estimativas de máxima verossimilhança penalizada apresentam algumas sinais de robustez, no sentido da distância de Mahalanobis $\hat{\delta}_i$, contra observações aberrantes.

7.2.4 Diagnóstico de influência

A fim de identificar observações influentes e de avaliar a sensibilidade da estimativa de máxima verossimilhança penalizada de f , λ e ϕ a tais observações, apresentamos os resultados obtidos da análise de influência local. Utilizamos a medida de influência local total, C_i , e dois esquemas de perturbação: ponderação de casos e perturbação de escala. A representação gráfica de $C_i(\boldsymbol{\theta})$ foi omitida, devido à similaridade dos resultados com aqueles obtidos na análise de influência local parcial.

Ponderação de casos

Nas Figuras 7.19, 7.20 e 7.21 são apresentados os gráficos de índices de $C_i(f)$, $C_i(\lambda)$ e $C_i(\phi)$ para os modelos normal e t-Student com 7 g.l., para o caso em que atribuímos diferentes ponderações às observações. Como resultado é possível notar que as observações 6, 7, 11 e 17 aparecem como observações influentes nas estimativas de máxima verossimilhança penalizada da função não paramétrica e dos componentes de variância sob o modelo normal. Observamos também que a observação 17 é mais influente em relação às outras observações. Já para o modelo t-Student não observa-se nenhuma observação exercendo influência de forma relevante.

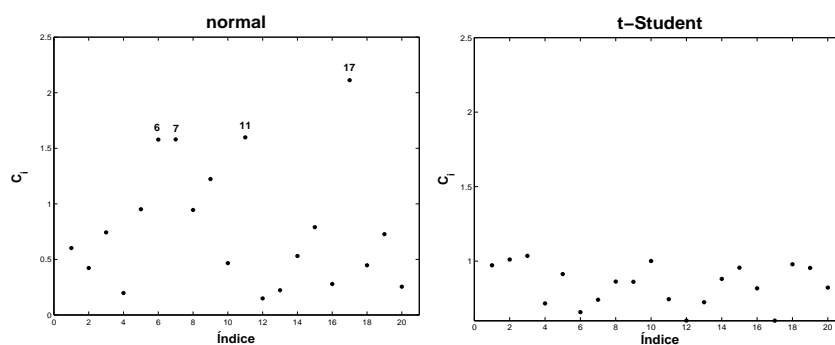


Figura 7.19: Gráficos de índices de C_i para \hat{f} sob ponderação de casos para os modelos normal e t-Student ajustados aos dados de glucose.

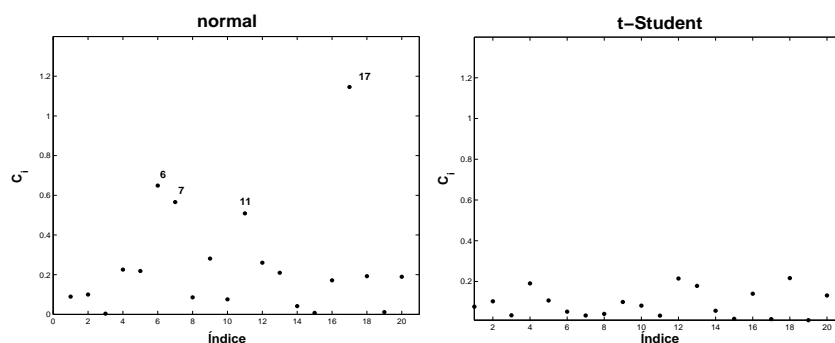


Figura 7.20: Gráficos de índices de C_i para $\hat{\lambda}$ sob ponderação de casos para os modelos normal e **t-Student** ajustados aos dados de glucose.

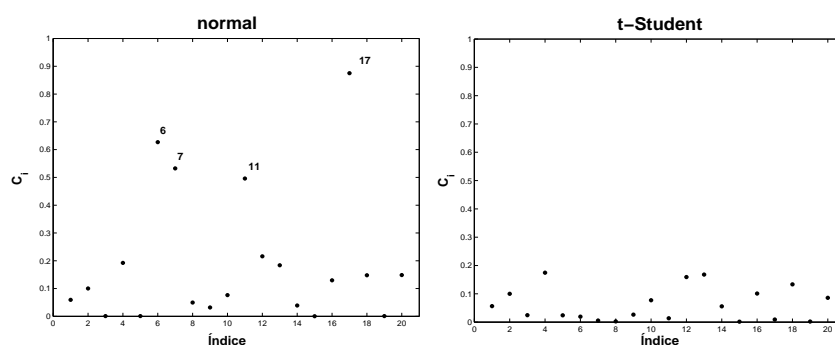


Figura 7.21: Gráficos de índices de C_i para $\hat{\phi}$ sob ponderação de casos para os modelos normal e **t-Student** ajustados aos dados de glucose.

Perturbação de escala

Nas Figuras 7.22, 7.23 e 7.24 apresentamos os gráficos de índices $C_i(f)$, $C_i(\lambda)$ e $C_i(\phi)$ para os modelos normal e **t-Student** com 7 g.l.. Podemos notar que as observações 6, 7, 11 e 17 aparecem como observações influentes nas estimativas de máxima verossimilhança penalizada da função não paramétrica e dos componentes de variância obtidos sob o modelo normal. Note que uma vez mais a observação 17 apresenta a maior influência perante as demais observações. Para o modelo **t-Student** podemos observar que os gráficos de influência local total não revelam a presença de observações influentes nas estimativas.

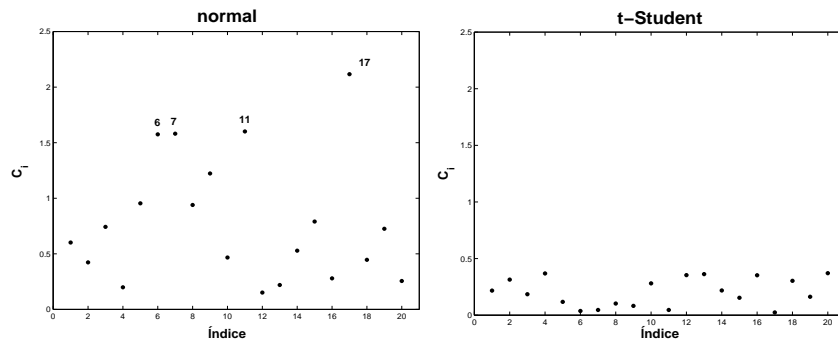


Figura 7.22: Gráficos de índices de C_i para \hat{f} sob perturbação de escala para os modelos normal e **t-Student** ajustados aos dados de glucose.

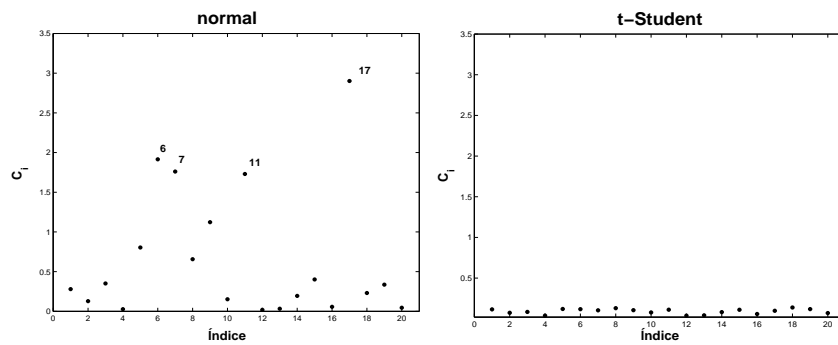


Figura 7.23: Gráficos de índices de C_i para $\hat{\lambda}$ sob perturbação de escala para os modelos normal e **t-Student** ajustados aos dados de glucose.

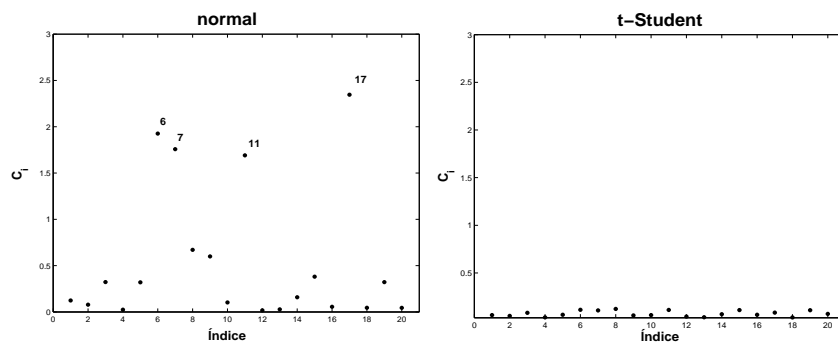


Figura 7.24: Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação de escala para os modelos normal e **t-Student** ajustados aos dados de glucose.

Como já foi mencionado no início, Pan e Fang (2002) ajustaram a esses dados um modelo de curva de crescimento assumindo normalidade e conseqüentemente discutiram alguns procedimentos para detectar observações discordantes e influentes. Nessa análise encontraram a observação 30 (para nós 17) como fortemente discordante e influente, coincidindo, desta forma, com os nossos resultados de influência local. Uma revisão da base de dados mostra que tal paciente tem o primeiro e segundo valor de entrada relativamente grande, mas os restantes diminuem rapidamente e voltam a crescer após duas horas da aplicação oral da dose de glucose. É evidente que a tendência no tempo dessas medições são diferentes em relação à maioria dos dados, como é observado a partir da Figura 7.15.

A partir dos resultados de diagnóstico de influência local total apresentados acima, temos que as estimativas da função não paramétrica e do componente de variância para o modelo normal apresentam uma importante sensibilidade quando atribuímos ponderações diferentes às observações e quando perturbamos a matriz de escala. Além disso, podemos observar que tais estimativas derivados do modelo *t*-Student com 7 g.l. são robustos, no sentido da distância de Mahalanobis, contra observações aberrantes, uma vez que a observação 17, identificada com uma distância grande sob o modelo normal multivariado, recebeu um peso pequeno no processo de estimação sob o modelo *t*-Student multivariado.

7.3 Dados de pressão ocular

Estes dados foram reportados em um estudo médico realizado com 30 pacientes com o objetivo de descrever o comportamento da pressão ocular do olho direito e do olho esquerdo em um dia específico. Esse conjunto de dados foi analisado por alunos da carreira de engenharia em Estatística da Univesidade de Valpararíso, Chile. As variáveis resposta correspondem às medições da pressão ocular dos pacientes registradas em intervalos de tempo de três horas; isto é, às 6, 9, 12, 15, 18, 21 e 24 horas; veja Apêndice E. O sexo e idade foram incluídas no estudo como covariáveis. Em alguns pacientes, apenas foi possível medir a pressão ocular em um dos olhos. Nesta aplicação consideraremos todos os pacientes aos quais foi possível

medir a pressão ocular do olho esquerdo, obtendo-se um total de 29 pacientes. As variáveis sexo e idade não serão consideradas em nossa análise. A Figura 7.25 mostra os perfis individuais dos pacientes. A forma dessas curvas sugere que as medições da pressão ocular poderiam ser modeladas de maneira razoável usando um modelo com efeitos aleatórios e um componente não paramétrico.

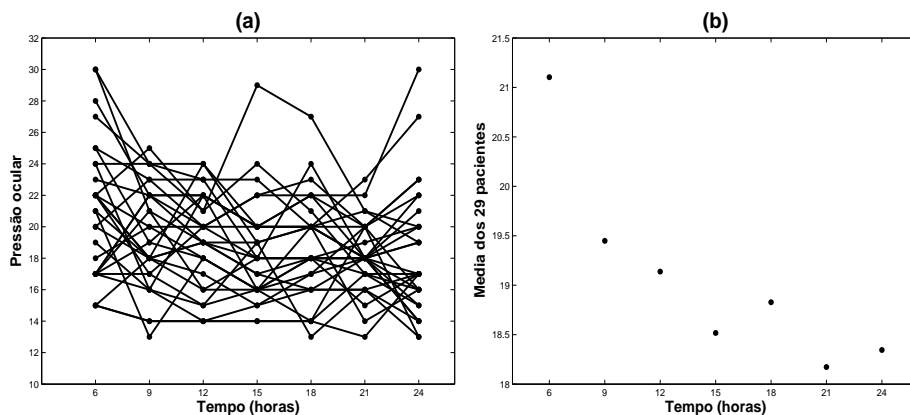


Figura 7.25: Gráficos de perfis (a) e das médias (b) da pressão ocular do olho esquerdo.

7.3.1 Modelo proposto

Em virtude do exposto acima, sugerimos analisar este conjunto de dados usando o seguinte modelo misto não paramétrico:

$$y_{ij} = f(t_{ij}) + b_i + \epsilon_{ij}, \quad (7.5)$$

em que y_{ij} representa a j -ésima medição da pressão ocular do i -ésimo paciente no tempo t_{ij} ($i = 1, \dots, 29$; $j = 1, \dots, 7$), f é uma função arbitrária que depende do tempo, b_i denota o efeito aleatório do i -ésimo paciente incorporado para modelar a estrutura de covariância, e ϵ_{ij} é um erro aleatório. Em termos matriciais temos

$$\mathbf{y}_i = \mathbf{N}_i \mathbf{f} + \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \quad (7.6)$$

em que \mathbf{y}_i é um vetor (7×1) para as medições do i -ésimo paciente, $\mathbf{N}_i = \mathbf{I}_7$ é uma matriz de incidência (7×7), onde \mathbf{I}_7 denota uma matriz identidade (7×7), \mathbf{f} é um vetor (7×1) cujos componentes correspondem à avaliação funcional de $f(\cdot)$ nos valores dos tempos que pertencem ao conjunto $\mathbf{t}^0 = \{ t_1^0 = 6, \dots, t_7^0 = 24 \}$, \mathbf{Z}_i é a matriz de planejamento (7×1) para os efeitos aleatórios dada por

$$\mathbf{Z}_i = \left(1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \right)^T,$$

e $\boldsymbol{\epsilon}_i$ representa o vetor de erros aleatórios (7×1). Assumindo que,

$$\begin{pmatrix} \mathbf{y}_i \\ b_i \end{pmatrix} \sim \mathcal{E}l_7 \left\{ \begin{pmatrix} \mathbf{f} \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda \mathbf{Z}_i \mathbf{Z}_i^T + \phi \mathbf{I}_7 & \lambda \mathbf{Z}_i \\ \lambda \mathbf{Z}_i^T & \lambda \end{pmatrix} \right\},$$

temos que o vetor de respostas associado ao i -ésimo paciente ($i = 1, \dots, 29$) segue uma distribuição marginal na forma,

$$\mathbf{y}_i \sim \mathcal{E}l_7(\mathbf{f}, \lambda \mathbf{Z}_i \mathbf{Z}_i^T + \phi \mathbf{I}_7).$$

Neste caso, o vetor de parâmetros a ser estimado é dado por $\boldsymbol{\theta} = (\mathbf{f}^T, \lambda, \phi)^T$. As distribuições normal e \mathbf{t} -Student serão usadas para comparar a sensibilidade das estimativas de máxima verossimilhança penalizada para este conjunto de dados.

7.3.2 Ajustando os modelos normal e \mathbf{t} -Student multivariado

O modelo (7.5) sob as distribuições normal e \mathbf{t} -Student multivariadas, respectivamente, foi ajustado usando o critério da verossimilhança penalizada. O grau de suavização $\hat{\alpha} = 4,3$ foi estimado através do critério de validação cruzada generalizada sob normalidade. Os graus de liberdade da distribuição \mathbf{t} -Student foram calculados através do critério de informação de Schwarz obtendo-se $\nu = 9$; veja Tabela 7.6. Os erros padrão do estimador do componente não paramétrica e do componente de variância foram estimados a partir da matriz de informação de Fisher penalizada. Como resultado, temos os ajustes descritos na Tabela 7.7.

Tabela 7.6: Valores estimados do logaritmo da função de verossimilhança penalizada e do critério de informação de Schwarz sob o modelo t -Student para diferentes graus de liberdade ajustado aos dados sobre pressão ocular.

ν	$-2L_p(\hat{\theta}, \hat{\alpha})$	$SIC(\hat{\theta})$
1	992,86	1023,20
2	977,84	1008,10
3	972,20	1002,50
4	969,44	999,75
5	968,02	998,33
6	967,28	997,58
7	966,88	997,19
8	966,70	997,01
9	966,66	996,96
10	966,70	996,99

Da Tabela 7.7 podemos notar que as estimativas do componente não paramétrico são, em termos gerais, similares entre os dois modelos ajustados. Porém, como os erros padrão de \hat{f} sob o modelo t -Student são menores do que os erros padrão obtidos para o modelo normal, temos indícios de que o modelo com caudas mais pesadas gera estimativas mais precisas para o componente não paramétrico. Em relação aos parâmetros associados ao componente de variância, temos que as inferências são diferentes entre ambos os modelos, particularmente os valores das estimativas. Porém não são comparáveis. Concluimos, também, baseados nos valores de $L_p(\hat{\theta}, \hat{\alpha})$, que o modelo t -Student com 9 graus de liberdade apresenta um ajuste adequado, como também parece indicar a Figura 7.26 onde são apresentados os gráficos normais de probabilidade.

7.3.3 Detectando observações aberrantes

A fim de explorar se há observações discrepantes no conjunto de dados (usando o mesmo critério adotado na Subseção 7.2.3), apresentamos na Figura 7.27(a-b) os gráficos de índices das distâncias de Mahalanobis ajustadas $\hat{\delta}_i$ e das distâncias modificadas $\hat{F}_i = \hat{\delta}_i/m_i$, respectivamente, para os modelos normal e t -Student.

Como resultado dessa análise, notamos que os pacientes 3, 26 e 28 mostram-se como observações aberrantes quando ajustamos um modelo normal. Para o modelo com caudas mais pesadas não aparece nenhuma observação discordante.

Tabela 7.7: Estimativas de máxima verossimilhança penalizada para os modelos normal e **t**-Student com $\nu = 9$ graus de liberdade ajustados aos dados sobre pressão ocular.

Parâmetro	Normal			t-Student		
	Estimativa	E.P.	$L_P(\hat{\theta}, \hat{\alpha})$	Estimativa	E.P.	$L_P(\hat{\theta}, \hat{\alpha})$
λ	6,265	1,839	-486,53	5,602	1,797	-483,33
ϕ	5,116	0,548		3,976	0,584	
$f(t_1)$	21,058	0,645		20,574	0,582	
$f(t_2)$	19,552	0,780		19,162	0,630	
$f(t_3)$	19,037	0,897		18,751	0,667	
$f(t_4)$	18,636	0,831		18,273	0,646	
$f(t_5)$	18,565	0,623		18,343	0,561	
$f(t_6)$	18,510	0,573		18,218	0,521	
$f(t_7)$	18,198	0,623		17,793	0,567	

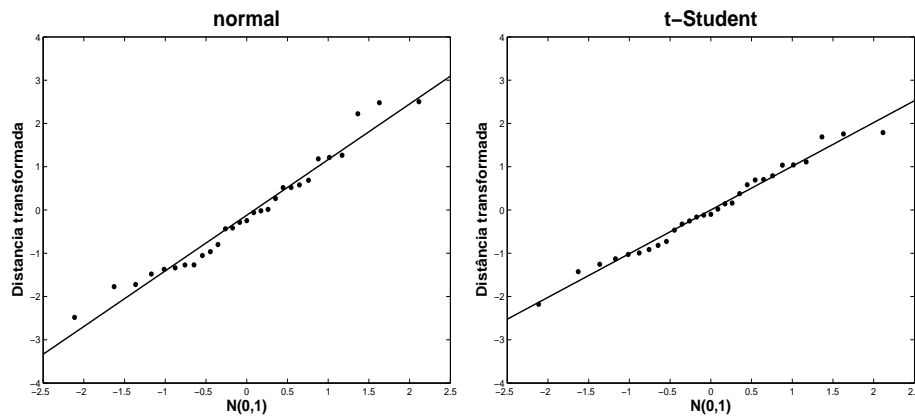


Figura 7.26: Gráficos normais de Probabilidades das distâncias transformadas sob os modelos normal e **t**-Student com 9 graus de liberdade ajustados aos dados sobre pressão ocular.

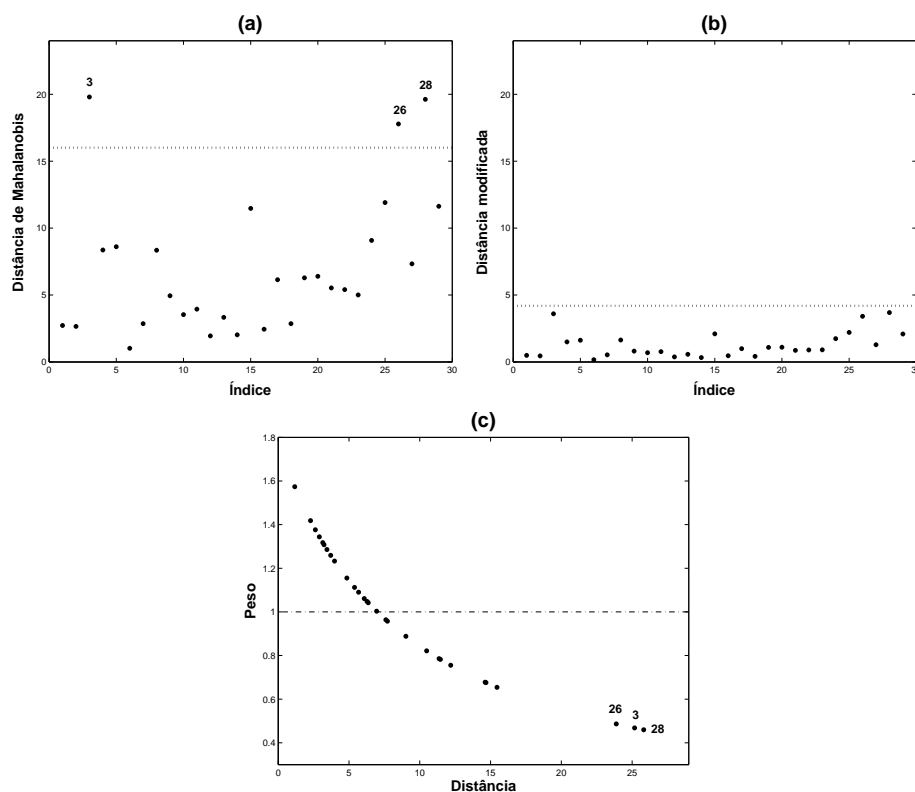


Figura 7.27: Gráficos de índices para a distância $\hat{\delta}_i$ sob os modelos normal (a) e t -Student (b) ajustados aos dados sobre pressão ocular, e entre os pesos estimados e a distância $\hat{\delta}_i$ sob o modelo t -Student (c).

7.3.4 Diagnóstico de influência

A seguir apresentamos os resultados da análise de influência local para identificar observações influentes e avaliar a sensibilidade das estimativas de máxima verossimilhança penalizada de f , λ e ϕ contra tais observações. A medida de influência local total C_i e os esquemas de ponderação de casos e perturbação de escala são considerados. Os gráficos de índices de $C_i(\theta)$ são omitidos devido à semelhança de tais resultados com os obtidos na análise de influência local parcial. Vamos supor que o parâmetro de suavização é fixo.

Ponderação de casos

Nas Figuras 7.28, 7.29 e 7.30 são apresentados os gráficos de índices de $C_i(f)$, $C_i(\lambda)$ e $C_i(\phi)$ para os modelos normal e t-Student com 7 g.l., para o caso em que atribuímos diferentes ponderações às observações. Os gráficos de influência revelam que as observações 3, 26 e 28 mostram-se influentes nas estimativas de máxima verossimilhança penalizada da função não paramétrica e do componente de variância, sob o modelo normal. Para o modelo t-Student não observa-se nenhuma observação exercendo influência de forma relevante nas estimativas de f , λ e ϕ .

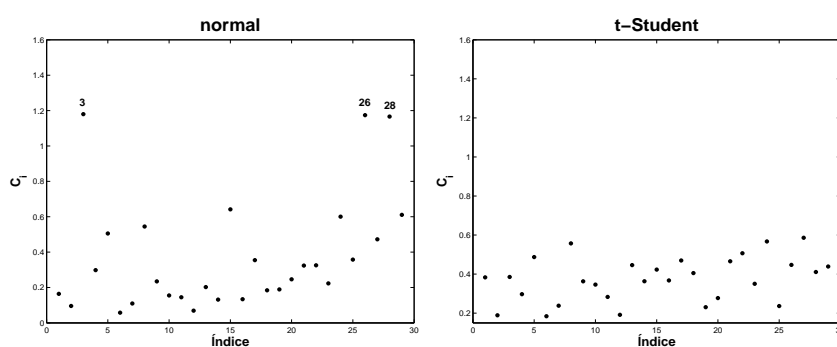


Figura 7.28: Gráficos de índices de C_i para \hat{f} sob ponderação de casos para os modelos normal e t-Student ajustados aos dados sobre pressão ocular.

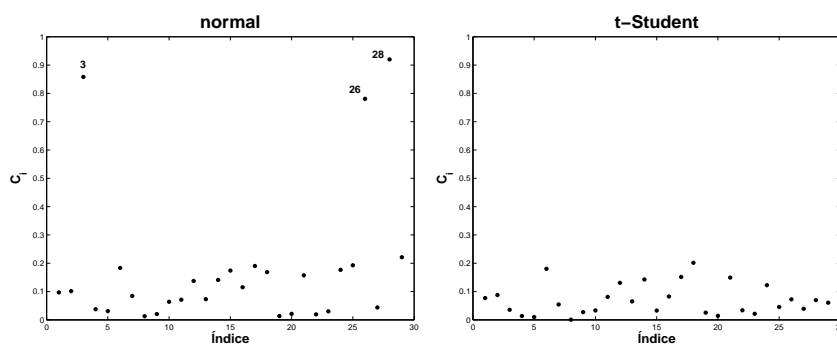


Figura 7.29: Gráficos de índices de C_i para $\hat{\lambda}$ sob ponderação de casos para os modelos normal e t-Student ajustados aos dados sobre pressão ocular.

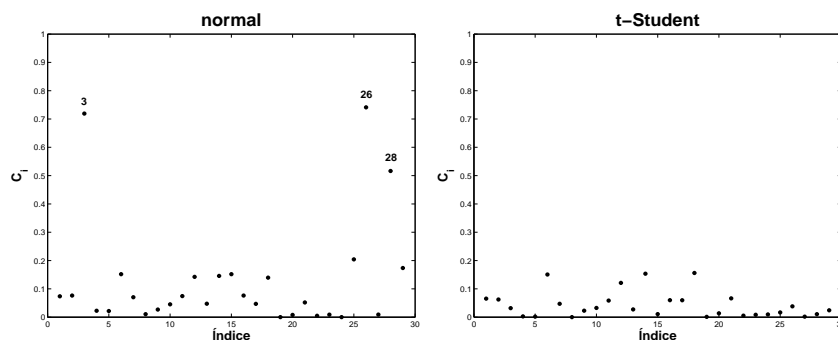


Figura 7.30: Gráficos de índices de C_i para $\hat{\phi}$ sob ponderação de casos para os modelos normal e **t-Student** ajustados aos dados sobre pressão ocular.

Perturbação na matriz de escala

Nas Figuras 7.31, 7.31 e 7.31 apresentamos os gráficos de índices $C_i(f)$, $C_i(\lambda)$ e $C_i(\phi)$ para os modelos normal e **t-Student** com 7 g.l.. Podemos notar que as observações 3, 26 e 28 emergem como observações influentes nas estimativas de máxima verossimilhança penalizada de f , λ e ϕ . Para o modelo **t-Student**, os gráficos de influência não revelam a presença de observações influentes sobre as estimativas.

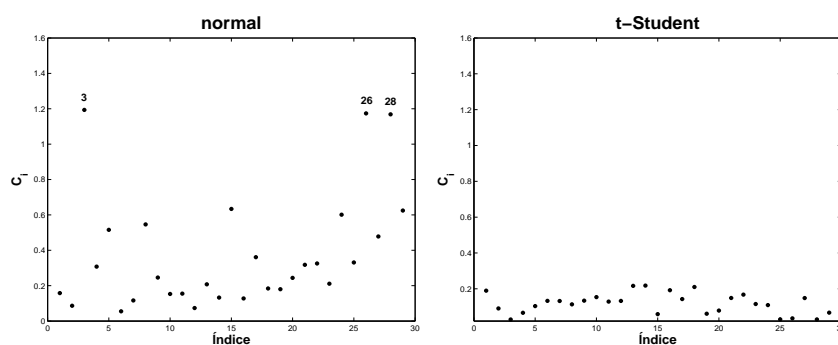


Figura 7.31: Gráficos de índices de C_i para \hat{f} sob perturbação de escala para os modelos normal e **t-Student** ajustados aos dados sobre pressão ocular.

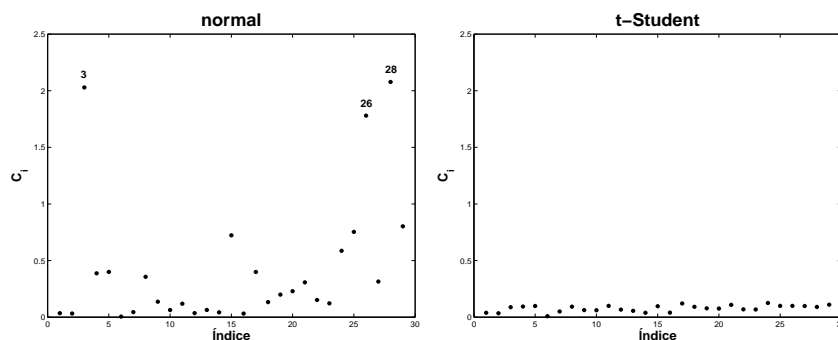


Figura 7.32: Gráficos de índices de C_i para $\hat{\lambda}$ sob perturbação de escala para os modelos normal e **t-Student** ajustados aos dados sobre pressão ocular.

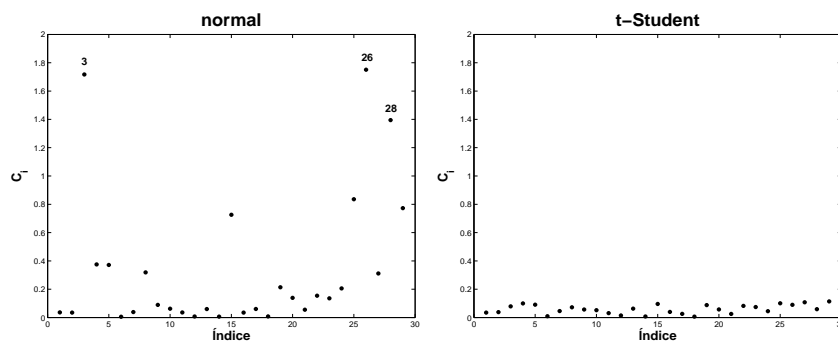


Figura 7.33: Gráficos de índices de C_i para $\hat{\phi}$ sob perturbação de escala para os modelos normal e **t-Student** ajustados aos dados sobre pressão ocular.

Ao observar a base de dados notamos que o paciente 26, por exemplo, tem um valor de entrada alto, mas os valores restantes da pressão ocular diminuem abruptamente. Além disso, observamos que as medições inicial e final da pressão ocular do paciente 28 têm um valor relativamente grande quando comparadas com as medições dos demais pacientes. Isto indica, portanto, um comportamento atípico de tais observações. A partir dos resultados de diagnóstico de influência local total obtidos nesta aplicação, temos que a estimativa de MVP de f , λ e ϕ sob o modelo normal mostra uma sensibilidade maior quando atribuem-se ponderações diferentes às observações e se perturba a matriz de escala. Temos ainda, que as estimativas derivadas do modelo **t-Student** com 9 g.l. são robustas, no sentido

da distância de Mahalanobis, contra observações aberrantes, uma vez que as observações 3, 26 e 28, que tinham uma distância relativamente grande sob o modelo normal, recebem pesos pequenos no processo de estimação.

7.4 Conclusões do capítulo

Neste capítulo apresentamos três aplicações com dados reais para ilustrar a aplicabilidade do modelo misto aditivo semiparamétrico elíptico juntamente com método de influência local. A medida de influência local total foi utilizada para identificar observações influentes e avaliar a sensibilidade das estimativas de MVP do coeficiente de regressão, da função não paramétrica e dos componentes de variância. Dos resultados obtidos na primeira aplicação observamos que sob o modelo normal as observações que têm uma grande influência na estimativa do coeficiente de regressão, não necessariamente têm o mesmo grau de influência na estimativa da função não paramétrica e vice versa. Também observamos que a estimativa de MVP do parâmetro de escala sob o modelo t-Student univariado é menos sensível a observações aberrantes. Os resultados obtidos na segunda e terceira aplicações mostraram que a estimativa de MVP da função não paramétrica e dos componentes de variância sob o modelo normal multivariado são altamente sensíveis sob a presença de observações aberrantes. Para o modelo t-Student multivariado, observa-se que as estimativas são robustas no sentido da distância de Mahalanobis. Essas aplicações confirmam que o método de influência local logra identificar observações que têm um comportamento atípico quando ajustamos um modelo misto semiparamétrico de contornos elípticos. Finalmente, é importante lembrar que em tais aplicações condicionamos nossa análise de influência assumindo que o parâmetro de suavização é fixo e, portanto, é possível que ocorra algum grau de sensibilidade nos resultados a mudanças no parâmetro de suavização.

Capítulo 8

Considerações finais

8.1 Discussão

Nesta tese estudamos uma nova classe de modelos denominada modelos mistos aditivos semiparamétricos elípticos. Essa classe torna possível a análise de dados de experimentos relacionados a fenômenos em que precisa-se modelar a dependência das medidas intraunidades amostrais, modelar os efeitos das covariáveis que contribuem de maneira paramétrica e não paramétrica sobre a variável resposta, e estender a modelagem estatística a outras distribuições elípticas além da distribuição normal, o que é muito importante dado o caráter dinâmico da modelagem de fenômenos cada vez mais complexos. Do ponto de vista teórico, nossas principais contribuições são a proposta de um novo modelo estatístico que generaliza o modelo misto semiparamétrico gaussiano, a obtenção de um procedimento para estimar os parâmetros do modelo, e o desenvolvimento do método de influência local para detectar observações influentes e avaliar a sensibilidade das estimativas sob esta nova classe de modelos. Do ponto de vista prático, as principais contribuições são a aplicabilidade dos modelos MMASE a conjuntos de dados reais, a obtenção de evidências empíricas da robustez, no sentido da distância de Mahalanobis, das estimativas de MVP geradas desde distribuições com caudas mais pesadas do que as caudas da normal e, aliás, a obtenção de evidências empíricas que revelam a capacidade do método de influência local para detectar observações influentes.

8.2 Implementação computacional

Nesta tese o processo iterativo baseado no algoritmo de escore de Fisher e *backfitting* para estimar o coeficiente de regressão, as funções não paramétricas, e os componentes de variância sob o modelo misto aditivo semiparamétrico elíptico, foi implementado no software MATLAB. A estrutura bloco-diagonal da matriz de informação de Fisher facilita o desenvolvimento desse método iterativo, ao permitir estimar os parâmetros associados aos efeitos fixos do modelo (paramétricos e não paramétricos) e os parâmetros associados aos componentes de variância usando processos iterativos independentes. A análise de diagnóstico de influência local também foi implementado computacionalmente usando o software MATLAB.

8.3 Perspectivas de trabalhos futuros

A primeira perspectiva de trabalho futuro esta relacionada com os esquemas de amostragem longitudinal. Tipicamente, em estudos longitudinais as observações de uma mesma unidade experimental são realizadas de forma sistemática, isto é, sequencialmente ao longo do tempo, e portanto deve-se considerar a correlação serial das observações de uma mesma unidade experimental. Na prática, os dados relacionados a estudos longitudinais são frequentemente não balanceados ou incompletos, isto é, quando para algumas unidades experimentais não há observação da resposta em uma ou mais ocasiões ou as unidades experimentais não têm a mesma matriz de planejamento, e o número de observações para cada unidade é relativamente maior em algumas delas. Nesses casos, é conveniente considerar modelos que possuam a capacidade de acomodar a natureza não balanceada dos dados longitudinais e uma estrutura de variância-covariância concorde com o critério de parcimônia. Em virtude disso, podemos considerar modelos que além de incorporar um componente de efeitos aleatórios, incorporem também erros aleatórios que sigam um processo de série de tempo autoregressivo de ordem 1. Nesse contexto, Chi e Reinsel (1989-1991) derivam um procedimento de estimação de máxima verossimilhança baseados no método de escore de Fisher para o modelo com efeitos

aleatórios e erros autoregressivos de ordem 1. Além disso, desenvolvem um teste escore para avaliar a presença de autocorrelação nos erros aleatórios. No contexto dos modelos mistos elípticos paramétricos, Cao e Lin (2009) também consideram erros autoregressivos de ordem 1 e desenvolvem alguns métodos de diagnóstico e teste de hipótese para avaliar a presença de autocorrelação nos erros aleatórios. Na mesma direção, Zhang et al. (1998) propõem flexibilizar a estrutura de variância-covariância incorporando processos estocásticos estacionários e não estacionários para modelar a correlação serial intraunidades experimentais.

A segunda perspectiva de trabalho futuro tem relação com aquelas situações nas quais a relação linear não é adequada devido à existência de uma interpretação física que relaciona os dados e parâmetros. Nesse caso, é necessário incorporar relações mais elaboradas, apesar da complexidade resultante dessas suposições. Tais situações têm motivado o desenvolvimento de modelos mais complexos, como por exemplo, os modelos mistos não lineares (MMNLs) e os modelos mistos não lineares semiparamétricos (MMNLSs). Os modelos MMNLs são uma extensão dos modelos MMLs e têm sido explorados nos últimos anos devido à sua flexibilidade para lidar com dados relacionados a medidas repetidas ou dados longitudinais em situações nas quais o valor esperado da variável resposta é não linear tanto nos efeitos fixos quanto nos efeitos aleatórios. Nesse contexto, Lindstrom e Bates (1990) propõem estimadores para os parâmetros combinando os estimadores de mínimos quadrados para os modelos de efeitos fixos não lineares e os estimadores de máxima verossimilhança para os modelos mistos lineares. Alternativamente, Vonesh e Carter (1992) propõem um modelo MMNL em que os efeitos fixos são não lineares enquanto os efeitos aleatórios são lineares, e propõem estimar os parâmetros do modelo através de um procedimento de mínimos quadrados generalizados. Outros trabalhos relacionados com os modelos MMNLs são apresentados, por exemplo, em Pinheiro e Bates (1995), entre outros. Por sua parte, os modelos MMNLSs introduzidos por Ke e Wang (2001) são uma extensão dos modelos MMNLs e dos modelos mistos não paramétricos (MMNs) propostos por Wang (1998), e caracterizam-se principalmente devido ao fato de que o valor esperado da variável resposta depende

dos efeitos fixos e aleatórios, e da função não paramétrica, em forma não linear. Eles aproximam a verossimilhança marginal e a verossimilhança penalizada baseados no método de Laplace, e utilizam suavização *spline* para modelar as funções não paramétricas. Além disso, desenvolvem intervalos de confiança bayesianos aproximados para as funções não paramétricas baseados em uma formulação bayesiana do modelo MMNLS.

A terceira perspectiva de trabalho futuro que merece um destaque especial tem relação com a inclusão do parâmetro de suavização como parte da análise de diagnóstico de influência local (Thomas, 1991). Nesta tese, a inferência estatística e a análise de influência local sob os modelos mistos aditivos semiparamétricos elípticos foi condicionada assumindo que os parâmetros de suavização são fixos. Eventualmente, a análise de diagnóstico poderia apresentar algum grau de sensibilidade se são introduzidas algumas perturbações nos parâmetros de suavização. Portanto, a busca de um esquema de perturbação que permita perturbar o parâmetro de suavização e as covariáveis que contribuem de maneira não paramétrica sobre a resposta, também deve ser investigada (Zhu et al., 2007).

Finalmente, uma quarta perspectiva de trabalho futuro é considerar o desenvolvimento de outros aspectos inferenciais sob esta classe de modelos, tais como o desenvolvimento de testes de hipóteses e intervalos de confiança. Nesse contexto, podemos propor, por exemplo, testes para os componentes de variância e testes para avaliar a hipótese de linearidade das funções não paramétricas; veja, por exemplo, Hardle et al. (1998), Pitrun et al. (2006), Bianco et al. (2006), Liang (2006), e Lombardia e Sperlich (2008).

Apêndice A

Matrizes de incidência e suavização

A.1 Construção das matrizes

Apenas para ilustrar a forma de construir uma matriz de incidência a partir de um conjunto dados, consideremos o modelo de regressão não paramétrico clássico

$$y_i = f(t_i) + \epsilon_i ,$$

em que os valores da covariável t_i ($i = 1, \dots, 9$) são apresentados na Tabela [A.1](#).

Tabela A.1: Descrição dos valores da variável t_i para o modelo de regressão não paramétrico clássico.

i	t_i	g	t_g^0
1	0,63	1	0,63
2	0,63	2	0,65
3	0,63	3	0,66
4	0,65	4	0,69
5	0,65	5	0,71
6	0,66		
7	0,69		
8	0,71		
9	0,71		

Escrevendo o modelo em termos matriciais temos que

$$\mathbf{y} = \mathbf{N}\mathbf{f} + \boldsymbol{\epsilon},$$

em que $\mathbf{y} = (y_1, \dots, y_9)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_9)^T$, $\mathbf{f} = (f(t_1^0), \dots, f(t_5^0))^T$. Logo, os elementos da matriz de incidência, \mathbf{N} , são obtidos na forma

$$\begin{aligned} (1, 1) &= I(t_1 = t_1^0) = 1 \\ &\vdots \\ (1, 5) &= I(t_1 = t_5^0) = 0 \\ &\vdots \\ (9, 1) &= I(t_9 = t_1^0) = 0 \\ &\vdots \\ (9, 5) &= I(t_9 = t_5^0) = 1. \end{aligned}$$

Nessas condições, a matriz de incidência (9×5) assume a forma

$$\mathbf{N} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Além disso, temos que as matrizes \mathbf{Q} (5×3) e \mathbf{R} (3×3) são, respectivamente,

dadas por

$$\mathbf{Q} = \begin{pmatrix} 50 & 0 & 0 \\ -150 & 100 & 0 \\ 100 & -133,3333 & 33,3333 \\ 0 & 33,3333 & -83,3333 \\ 0 & 0 & 50 \end{pmatrix} \quad \text{e} \quad \mathbf{R} = \begin{pmatrix} 0,01 & 0,0017 & 0 \\ 0,0017 & 0,0133 & 0,005 \\ 0 & 0,005 & 0 \end{pmatrix}.$$

Dessa forma, a matriz de suavização \mathbf{K} é obtida calculando

$$\mathbf{K} = \mathbf{QR}^{-1}\mathbf{Q}^T.$$

A.2 Programas em linguagem MATLAB

Cálculo da matriz N

```

for i = 1 : n
    for j = 1 : r
        if t(i) == t0(j)
            N(i, j) = 1;
        elseif t(i) = t0(j)
            N(i, j)=0;
        end
    end
end
end

```

Cálculo da matriz Q

```

for i = 1 : (r - 1)
    h(i) = t0(i + 1)-t0(i);

```

```

end
for  $i = 1 : r$ 
  for  $j = 2 : (r - 1)$ 
    if  $\text{abs}(i - j) < 2$ 
       $q(j - 1, j) = h^{-1}(j - 1);$ 
       $q(j, j) = -h^{-1}(j - 1) - h^{-1}(j);$ 
       $q(j + 1, j) = h^{-1}(j);$ 
    elseif  $\text{abs}(i - j) \geq 2$ 
       $q(i, j) = 0;$ 
    end
  end
end
end
 $Q = Q(1 : r, 2 : (r - 1))$ 

```

Cálculo da matriz R

```

for  $i = 2 : (r - 2)$ 
  for  $j = 2 : (r - 2)$ 
    if  $\text{abs}(i - j) < 2$ 
       $r(i, i) = (1/3) * (h(i - 1) + h(i));$ 
       $r(i, i + 1) = (1/6) * h(i);$ 
       $r(i + 1, i) = (1/6) * h(i);$ 
    elseif  $\text{abs}(i - j) \geq 2$ 
       $r(i, j) = 0;$ 
    end
  end
end
end
 $R = R(2 : (r - 1), 2 : (r - 1));$ 
 $K = Q * \text{inv}(R) * Q';$ 

```


Apêndice B

Prova de alguns resultados

Neste apêndice apresentamos as provas de alguns dos principais resultados relacionados com os modelos mistos aditivos semiparamétricos elípticos. Especificamente, provamos que (a) \mathbf{y}_i possui uma distribuição marginal elíptica e que (b) $\mathbf{b}_i | \mathbf{y}_i$ tem uma distribuição condicional elíptica.

B.1 Prova de (2.12)

Consideremos a distribuição conjunta

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \sim \mathcal{E}l_{m_i^*} \left\{ \begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i & \mathbf{Z}_i \mathbf{D} & \mathbf{V}_i \\ \mathbf{D} \mathbf{Z}_i^T & \mathbf{D} & \mathbf{0} \\ \mathbf{V}_i & \mathbf{0} & \mathbf{V}_i \end{pmatrix} \right\},$$

em que $m_i^* = m_i + q + m_i$. Além disso, consideremos a seguinte partição:

$$\mathbf{v}_i = \begin{pmatrix} \mathbf{y}_i^{(1)} \\ \mathbf{y}_i^{(2)} \end{pmatrix}, \quad \mathbf{y}_i^{(1)} = \mathbf{y}_i, \quad \mathbf{y}_i^{(2)} = \begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix},$$
$$\boldsymbol{\eta}_i = \begin{pmatrix} \boldsymbol{\mu}_i^{(1)} \\ \boldsymbol{\mu}_i^{(2)} \end{pmatrix}, \quad \boldsymbol{\mu}_i^{(1)} = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k, \quad \boldsymbol{\mu}_i^{(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

e

$$\mathbf{\Gamma}_i = \begin{pmatrix} \mathbf{\Gamma}_{i11} & \mathbf{\Gamma}_{i12} \\ \mathbf{\Gamma}_{i12}^T & \mathbf{\Gamma}_{i22} \end{pmatrix},$$

em que

$$\mathbf{\Gamma}_{i11} = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i, \quad \mathbf{\Gamma}_{i12} = \begin{pmatrix} \mathbf{Z}_i \mathbf{D} & \mathbf{V}_i \end{pmatrix} \quad \text{e} \quad \mathbf{\Gamma}_{i22} = \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_i \end{pmatrix}.$$

De acordo com a Propriedade 1.4.3,

$$\boldsymbol{\delta} + \mathbf{B}^T \mathbf{v}_i \sim \mathcal{E}l_{m_i^*}(\boldsymbol{\delta} + \mathbf{B}^T \boldsymbol{\eta}_i, \mathbf{B}^T \mathbf{\Gamma}_i \mathbf{B}). \quad (\text{B.1})$$

Para $\mathbf{B}^T = (\mathbf{I}_{m_i} \mathbf{0}) \in \mathbb{R}^{m_i \times m_i^*}$ e $\boldsymbol{\delta} = \mathbf{0} \in \mathbb{R}^{m_i}$, em que \mathbf{I}_{m_i} é uma matriz identidade ($m_i \times m_i$) e $\mathbf{0}$ é uma matriz nula ($m_i \times q_i$), com $q_i = m_i + q$, temos que

$$\begin{aligned} \boldsymbol{\delta} + \mathbf{B}^T \mathbf{v}_i &= \mathbf{y}_i \\ \boldsymbol{\delta} + \mathbf{B}^T \boldsymbol{\eta}_i &= \boldsymbol{\mu}_i \\ \mathbf{B}^T \mathbf{\Gamma}_i \mathbf{B} &= \mathbf{\Gamma}_{i11}. \end{aligned}$$

Daqui segue que $\mathbf{y}_i \sim \mathcal{E}l_{m_i}(\boldsymbol{\mu}_i, \mathbf{\Gamma}_{i11})$. ◀

B.2 Prova de (3.30)

Consideremos, agora, a distribuição conjunta

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{pmatrix} \sim \mathcal{E}l_{(m_i+q)} \left\{ \begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i & \mathbf{Z}_i \mathbf{D} \\ \mathbf{D} \mathbf{Z}_i^T & \mathbf{D} \end{pmatrix} \right\}.$$

Além disso, consideremos a seguinte partição:

$$\mathbf{v}_i = \begin{pmatrix} \mathbf{y}_i^{(1)} \\ \mathbf{y}_i^{(2)} \end{pmatrix}, \quad \mathbf{y}_i^{(1)} = \mathbf{y}_i, \quad \mathbf{y}_i^{(2)} = \mathbf{b}_i,$$

$$\boldsymbol{\eta}_i = \begin{pmatrix} \boldsymbol{\mu}_i^{(1)} \\ \boldsymbol{\mu}_i^{(2)} \end{pmatrix}, \quad \boldsymbol{\mu}_i^{(1)} = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k, \quad \boldsymbol{\mu}_i^{(2)} = \mathbf{0},$$

e

$$\boldsymbol{\Gamma}_i = \begin{pmatrix} \boldsymbol{\Gamma}_{i11} & \boldsymbol{\Gamma}_{i12} \\ \boldsymbol{\Gamma}_{i12}^T & \boldsymbol{\Gamma}_{i22} \end{pmatrix},$$

em que

$$\boldsymbol{\Gamma}_{i11} = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i, \quad \boldsymbol{\Gamma}_{i12} = \mathbf{Z}_i \mathbf{D} \quad \text{e} \quad \boldsymbol{\Gamma}_{i22} = \mathbf{D}.$$

De acordo com a Propriedade 1.4.4,

$$(\mathbf{y}^{(2)} \mid \mathbf{y}_0^{(1)}) \sim \mathcal{E}l_q(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}), \quad (\text{B.2})$$

em que

$$\begin{aligned} \boldsymbol{\mu}_{2.1} &= \boldsymbol{\mu}_i^{(2)} + \boldsymbol{\Gamma}_{i21} \boldsymbol{\Gamma}_{i11}^{-1} (\mathbf{y}_0^{(1)} - \boldsymbol{\mu}_i^{(1)}) \\ &= \mathbf{0} + \mathbf{D} \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{y}_0^{(1)} - \mathbf{X}_i \boldsymbol{\beta} - \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k \right) \quad \text{e} \\ \boldsymbol{\Sigma}_{22.1} &= \boldsymbol{\Gamma}_{i22} - \boldsymbol{\Gamma}_{i21} \boldsymbol{\Gamma}_{i11}^{-1} \boldsymbol{\Gamma}_{i12} \\ &= \mathbf{D} - \mathbf{D} \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \mathbf{D}, \end{aligned}$$

com $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{V}_i$. Daqui segue que $\mathbf{b}_i \mid \mathbf{y}_i \sim \mathcal{E}l_q(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1})$. ◀

Apêndice C

Derivadas da log-verossimilhança penalizada

Neste apêndice apresentamos os cálculos das derivadas de primeira e segunda ordem do logaritmo da função de verossimilhança penalizada do modelo misto aditivo semiparamétrico elíptico. Esses cálculos envolvem alguns resultados sobre álgebra e diferenciação de matrizes que podem ser encontrados, por exemplo, em Magnus e Neudecker (1988).

C.1 Derivadas de primeiro ordem

No modelo misto aditivo semiparamétrico elíptico (2.12) o logaritmo da função de verossimilhança penalizada é dado por

$$L_P(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^n L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \quad (\text{C.1})$$

em que

$$L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = L_i(\boldsymbol{\theta}) - \frac{1}{2n} \sum_{k=1}^s \alpha_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k, \quad (\text{C.2})$$

com $L_i(\boldsymbol{\theta})$ definida pela equação (3.2) e $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}_1^T, \dots, \mathbf{f}_s^T, \boldsymbol{\lambda}^T, \boldsymbol{\gamma}^T)^T$. Usando resultados de diferenciação de matrizes temos que,

$$\frac{\partial L_P(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \frac{\partial L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}}. \quad (\text{C.3})$$

Com efeito, derivando (C.2) com relação a β e \mathbf{f}_k , temos que

$$\begin{aligned}
\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \beta} &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \beta} \\
&= \frac{\partial \log g(\delta_i)}{\partial \beta} \\
&= \frac{g'(\delta_i)}{g(\delta_i)} \frac{\partial \boldsymbol{\varepsilon}_i^T}{\partial \beta} \frac{\partial [\boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \boldsymbol{\varepsilon}_i} \\
&= v(\delta_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i
\end{aligned} \tag{C.4}$$

e

$$\begin{aligned}
\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k} &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \mathbf{f}_k} - \frac{\alpha_k}{2n} \frac{\partial [\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k]}{\partial \mathbf{f}_k} \\
&= \frac{\partial \log g(\delta_i)}{\partial \mathbf{f}_k} - \frac{\alpha_k}{n} \mathbf{K}_k \mathbf{f}_k \\
&= \frac{g'(\delta_i)}{g(\delta_i)} \frac{\partial \boldsymbol{\varepsilon}_i^T}{\partial \mathbf{f}_k} \frac{\partial [\boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \boldsymbol{\varepsilon}_i} - \frac{\alpha_k}{n} \mathbf{K}_k \mathbf{f}_k \\
&= v(\delta_i) \mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha_k}{n} \mathbf{K}_k \mathbf{f}_k, \quad k = 1, \dots, s,
\end{aligned} \tag{C.5}$$

em que $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \sum_{k=1}^s \mathbf{N}_{ki} \mathbf{f}_k$ e $v(\delta_i) = -2W_g(\delta_i)$, com $W_g(\delta_i) = \frac{g'(\delta_i)}{g(\delta_i)}$. Usando a notação $\dot{\boldsymbol{\Sigma}}_i(\ell) = \partial \boldsymbol{\Sigma}_i / \partial \lambda_\ell$ e $\dot{\boldsymbol{\Sigma}}_i(j) = \partial \boldsymbol{\Sigma}_i / \partial \gamma_j$, e derivando (C.1) com relação a $\boldsymbol{\lambda}$ e $\boldsymbol{\gamma}$ obtemos que

$$\begin{aligned}
\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \lambda_\ell} &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \lambda_\ell} \\
&= -\frac{1}{2} \frac{\partial \log |\boldsymbol{\Sigma}_i|}{\partial \lambda_\ell} + \frac{\partial \log g(\delta_i)}{\partial \lambda_\ell} \\
&= -\frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\} + W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \lambda_\ell} \boldsymbol{\varepsilon}_i \\
&= -\frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\} - W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \lambda_\ell} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \\
&= -\frac{1}{2} \left[\text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\} - v(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right]
\end{aligned} \tag{C.6}$$

e

$$\begin{aligned}
\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \gamma_j} &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \gamma_j} & (C.7) \\
&= -\frac{1}{2} \frac{\partial \log |\boldsymbol{\Sigma}_i|}{\partial \gamma_j} + \frac{\partial \log g(\delta_i)}{\partial \gamma_j} \\
&= -\frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\} + W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \gamma_j} \boldsymbol{\varepsilon}_i \\
&= -\frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\} - W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \gamma_j} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \\
&= -\frac{1}{2} \left[\text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\} - v(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right],
\end{aligned}$$

para $\ell = 1, \dots, d_\lambda$ e $j = 1, \dots, d_\gamma$.

C.2 Derivadas de segunda ordem

Usando resultados de diferenciação de matrizes temos que a matriz de segundas derivadas em relação a $\boldsymbol{\theta}$ é dada por

$$\frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \sum_{i=1}^n \frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \quad (C.8)$$

Usando a notação $\dot{\mathbf{D}}(\ell) = \partial \mathbf{D} / \partial \lambda_\ell$ e $\dot{\mathbf{V}}_i(j) = \partial \mathbf{V}_i / \partial \gamma_j$, e derivando (C.4) com relação a $\boldsymbol{\beta}$, \mathbf{f}_k , $\boldsymbol{\lambda}$ e $\boldsymbol{\gamma}$, respectivamente, temos que as matrizes de segundas derivadas parciais são dadas por

$$\begin{aligned}
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \frac{\partial [v(\delta_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \boldsymbol{\beta}^T} \\
&= \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[\boldsymbol{\varepsilon}_i \frac{\partial v(\delta_i)}{\partial \boldsymbol{\beta}^T} + v(\delta_i) \frac{\partial \boldsymbol{\varepsilon}_i}{\partial \boldsymbol{\beta}^T} \right] \\
&= \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[-2 v'(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - v(\delta_i) \mathbf{X}_i \right] \\
&= 2 \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[2 W_g'(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \mathbf{f}_k^T} &= \frac{\partial [v(\delta_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \mathbf{f}_k^T} \\
&= \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[\boldsymbol{\varepsilon}_i \frac{\partial v(\delta_i)}{\partial \mathbf{f}_k^T} + v(\delta_i) \frac{\partial \boldsymbol{\varepsilon}_i}{\partial \mathbf{f}_k^T} \right] \\
&= \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[-2 v'(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{ki} - v(\delta_i) \mathbf{N}_{ki} \right] \\
&= 2 \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[2 W_g'(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{ki}, \quad k = 1, \dots, s,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \lambda_\ell} &= \frac{\partial [v(\delta_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \lambda_\ell} \\
&= -2 \mathbf{X}_i^T \left[\boldsymbol{\Sigma}_i^{-1} \frac{\partial W_g(\delta_i)}{\partial \lambda_\ell} + W_g(\delta_i) \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \lambda_\ell} \right] \boldsymbol{\varepsilon}_i \\
&= -2 \mathbf{X}_i^T \left[-W_g'(\delta_i) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \right. \\
&\quad \left. - W_g(\delta_i) \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \right] \boldsymbol{\varepsilon}_i \\
&= 2 \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[W_g'(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\varepsilon}_i, \quad \ell = 1, \dots, d_\lambda,
\end{aligned}$$

e

$$\begin{aligned}
\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \gamma_j} &= \frac{\partial [v(\delta_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \gamma_j} \\
&= -2 \mathbf{X}_i^T \left[\boldsymbol{\Sigma}_i^{-1} \frac{\partial W_g(\delta_i)}{\partial \gamma_j} + W_g(\delta_i) \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \gamma_j} \right] \boldsymbol{\varepsilon}_i \\
&= -2 \mathbf{X}_i^T \left[-W_g'(\delta_i) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \right. \\
&\quad \left. - W_g(\delta_i) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \right] \boldsymbol{\varepsilon}_i \\
&= 2 \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \left[W_g'(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\varepsilon}_i, \quad j = 1, \dots, d_\gamma.
\end{aligned}$$

Da mesma maneira, derivando (C.5) com relação a \mathbf{f}_k , $\boldsymbol{\lambda}$ e $\boldsymbol{\gamma}$, respectivamente, temos que as matrizes de segundas derivadas parciais, para $k, k' = 1, \dots, s$, $\ell = 1, \dots, d_\lambda$ e $j = 1, \dots, d_\gamma$, são dadas por

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \mathbf{f}_{k'}^T} = \begin{cases} 2\mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1} \left[2W'_g(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{ki} - \frac{\alpha_k}{n} \mathbf{K}_k & k = k' \\ 2\mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1} \left[2W'_g(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{k'i} & k \neq k'. \end{cases}$$

$$\begin{aligned} \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \lambda_\ell} &= \frac{\partial [v(\delta_i) \mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \lambda_\ell} \\ &= -2\mathbf{N}_{ki}^T \left[\boldsymbol{\Sigma}_i^{-1} \frac{\partial W_g(\delta_i)}{\partial \lambda_\ell} + W_g(\delta_i) \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \lambda_\ell} \right] \boldsymbol{\varepsilon}_i \\ &= -2\mathbf{N}_{ki}^T \left[-W'_g(\delta_i) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \right. \\ &\quad \left. - W_g(\delta_i) \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \right] \boldsymbol{\varepsilon}_i \\ &= 2\mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1} \left[W'_g(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\varepsilon}_i, \end{aligned}$$

e

$$\begin{aligned} \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \gamma_j} &= \frac{\partial [v(\delta_i) \mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]}{\partial \gamma_j} \\ &= -2\mathbf{N}_{ki}^T \left[\boldsymbol{\Sigma}_i^{-1} \frac{\partial W_g(\delta_i)}{\partial \gamma_j} + W_g(\delta_i) \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \gamma_j} \right] \boldsymbol{\varepsilon}_i \\ &= -2\mathbf{N}_{ki}^T \left[-W'_g(\delta_i) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \right. \\ &\quad \left. - W_g(\delta_i) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \right] \boldsymbol{\varepsilon}_i \\ &= 2\mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1} \left[W'_g(\delta_i) \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T + W_g(\delta_i) \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\varepsilon}_i. \end{aligned}$$

Usando a notação $\ddot{\mathbf{D}}(\ell^*, \ell) = \partial^2 \mathbf{D} / \partial \lambda_{\ell^*} \partial \lambda_\ell$ e $\ddot{\mathbf{V}}_i(j^*, j) = \partial^2 \mathbf{V}_i / \partial \gamma_{j^*} \partial \gamma_j$, e derivando (C.6) e (C.7) com relação a $\boldsymbol{\lambda}$ e $\boldsymbol{\gamma}$, respectivamente, obtemos, para $\ell, \ell^* = 1, \dots, d_\lambda$ e $j, j^* = 1, \dots, d_\gamma$,

$$\begin{aligned}
\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \lambda_{\ell^*} \partial \lambda_{\ell}} &= -\frac{1}{2} \frac{\partial \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \} - v(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right]}{\partial \lambda_{\ell^*}} \\
&= -\frac{1}{2} \frac{\partial \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \} \right]}{\partial \lambda_{\ell^*}} - \frac{\partial \left[W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right]}{\partial \lambda_{\ell^*}} \\
&= \frac{1}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \{ \dot{\mathbf{D}}(\ell^*) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) - \ddot{\mathbf{D}}(\ell^*, \ell) \} \mathbf{Z}_i^T \\
&\quad + \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \left[W'_g(\delta_i) \dot{\mathbf{D}}(\ell^*) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) - W_g(\delta_i) \dot{\mathbf{D}}(\ell^*, \ell) \right. \\
&\quad \left. + W_g(\delta_i) \dot{\mathbf{D}}(\ell^*) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) + W_g(\delta_i) \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell^*) \right] \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \gamma_{j^*} \partial \gamma_j} &= -\frac{1}{2} \frac{\partial \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \} - v(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right]}{\partial \gamma_{j^*}} \\
&= -\frac{1}{2} \frac{\partial \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \} \right]}{\partial \gamma_{j^*}} - \frac{\partial \left[W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right]}{\partial \gamma_{j^*}} \\
&= \frac{1}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \{ \dot{\mathbf{V}}_i(j^*) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) - \ddot{\mathbf{V}}_i(j^*, j) \} \\
&\quad + \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \left[W'_g(\delta_i) \dot{\mathbf{V}}_i(j^*) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) - W_g(\delta_i) \dot{\mathbf{V}}_i(j^*, j) \right. \\
&\quad \left. + W_g(\delta_i) \dot{\mathbf{V}}_i(j^*) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) + W_g(\delta_i) \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j^*) \right] \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i .
\end{aligned}$$

Finalmente, derivando (C.6) com relação a $\boldsymbol{\gamma}$, obtemos, para $\ell = 1, \dots, d_\lambda$ e $j = 1, \dots, d_\gamma$,

$$\begin{aligned}
\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \gamma_j \partial \lambda_{\ell}} &= -\frac{1}{2} \frac{\partial \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \} - v(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right]}{\partial \gamma_j} \\
&= -\frac{1}{2} \frac{\partial \left[\text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \} \right]}{\partial \gamma_j} - \frac{\partial \left[W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right]}{\partial \gamma_j} \\
&= \frac{1}{2} \text{tr} \{ \widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \} \\
&\quad + \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \left[W'_g(\delta_i) \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \right. \\
&\quad \left. + W_g(\delta_i) \dot{\mathbf{V}}_i(j) \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T + W_g(\delta_i) \mathbf{Z}_i \dot{\mathbf{D}}(\ell) \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{V}}_i(j) \right] \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i .
\end{aligned}$$

C.3 Matriz de informação de Fisher penalizada

Nesta seção mostramos como obter a matriz de informação de Fisher penalizada sob o modelo misto aditivo semiparamétrico elíptico, apresentada na Seção 3.15. A seguir, apresentamos alguns resultados prévios necessários para obter essa matriz. Tais resultados podem ser encontrados, por exemplo, em Lange et al. (1989) e Fang et al. (1999). Fazendo

$$\begin{aligned}
 \delta_i &= \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \\
 &= \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\varepsilon}_i \\
 &= \mathbf{P}_i^T \mathbf{P}_i \\
 &= \|\mathbf{P}_i\|^2,
 \end{aligned} \tag{C.9}$$

em que $\mathbf{P}_i = \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\varepsilon}_i$, temos que

$$\begin{aligned}
 \mathbb{E}\left(W_g(\delta_i) \|\mathbf{P}_i\|^2\right) &= -\frac{m_i}{2}, \\
 \mathbb{E}\left(W_g^2(\delta_i) \|\mathbf{P}_i\|^2\right) &= d_{g_i}, \\
 \mathbb{E}\left(W_g^2(\delta_i) \|\mathbf{P}_i\|^4\right) &= f_{g_i}, \\
 \mathbb{E}\left(\frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \mathbf{C} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\|\right) &= \frac{1}{h} \text{tr}\{\mathbf{C}\} \quad \text{e} \\
 \mathbb{E}\left(\frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \mathbf{C} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \mathbf{E} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\|\right) &= \frac{1}{h(h+2)} \left(2\text{tr}\{\mathbf{C}\mathbf{E}\} + \right. \\
 &\quad \left. \text{tr}\{\mathbf{C}\}\text{tr}\{\mathbf{E}\}\right),
 \end{aligned}$$

em que \mathbf{C} e \mathbf{E} são matrizes $(h \times h)$.

C.3.1 Prova de (3.18)

De (C.4) e (C.9) temos que

$$\begin{aligned} \frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} &= \frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &= -2W_g(\delta_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\varepsilon}_i \\ &= -2W_g(\delta_i) \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i, \end{aligned}$$

e, consequentemente,

$$\begin{aligned} \frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \beta_j} &= \frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta})}{\partial \beta_j} \\ &= -2W_g(\delta_i) \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i, \end{aligned}$$

em que \mathbf{x}_{ij}^{*T} denota a j -ésima coluna da matriz de planejamento \mathbf{X}_i . A partição da matriz de informação de Fisher penalizada referente a β_j e β_ℓ ($j, \ell = 1, \dots, p$) para o i -ésimo grupo é dada por

$$\begin{aligned} \mathcal{I}_{\mathbf{p}_i}^{\beta_j \beta_\ell}(\boldsymbol{\theta}) &= \mathbb{E} \left(- \frac{\partial^2 L_{\mathbf{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \beta_j \partial \beta_\ell} \right) \\ &= \mathbb{E} \left(- \frac{\partial^2 L_{\mathbf{p}_i}(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_\ell} \right) \\ &= \mathbb{E} \left(\frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta})}{\partial \beta_\ell} \right) \\ &= \mathbb{E} \left\{ \mathbb{E} \left(\frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{\mathbf{p}_i}(\boldsymbol{\theta})}{\partial \beta_\ell} \middle| \|\mathbf{P}_i\| \right) \right\}, \end{aligned}$$

em que

$$\begin{aligned}
\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_\ell} &= \left(-2W_g(\delta_i) \mathbf{x}_{ij}^* \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \left(-2W_g(\delta_i) \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \\
&= 4W_g^2(\delta_i) \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \mathbf{x}_{ij}^* \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\
&= 4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{x}_{ij}^* \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|}
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{E} \left(\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_\ell} \middle| \|\mathbf{P}_i\| \right) &= \mathbb{E} \left\{ 4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \times \right. \\
&\quad \left. \mathbb{E} \left(\frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{x}_{ij}^* \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \right) \middle| \|\mathbf{P}_i\| \right\} \\
&= \mathbb{E} \left(4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \frac{1}{m_i} \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1/2} \mathbf{x}_{ij}^* \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \right\} \right) \\
&= \frac{1}{m_i} \text{tr} \left\{ \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{x}_{ij}^* \right\} \mathbb{E} \left(4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \right) \\
&= \frac{4}{m_i} \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{x}_{ij}^* \mathbb{E} \left(W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \right) \\
&= \frac{4d_{g_i}}{m_i} \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{x}_{ij}^* .
\end{aligned}$$

Assim,

$$\mathcal{I}_{\mathbf{P}_i}^{\beta_j \beta_\ell}(\boldsymbol{\theta}) = \frac{4d_{g_i}}{m_i} \mathbf{x}_{i\ell}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{x}_{ij}^* , \quad (\text{C.10})$$

e, portanto, a matriz de informação de Fisher penalizada referente a $\boldsymbol{\beta}$ será dada por

$$\mathcal{I}_{\mathbf{P}}^{\beta\beta}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{4d_{g_i}}{m_i} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i . \quad (\text{C.11})$$

C.3.2 Prova de (3.19)

De (C.4) temos que

$$\begin{aligned} \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \mathbf{f}_k^T} &= \frac{\partial}{\partial \mathbf{f}_k^T} \left[\frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \right] \\ &= \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \mathbf{f}_k^T}, \end{aligned}$$

e, conseqüentemente,

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \beta_j \partial \zeta_{k_\ell}} = \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \beta_j \partial \zeta_{k_\ell}}.$$

A partição da matriz de informação de Fisher penalizada referente a β_j e ζ_{k_ℓ} ($j = 1, \dots, p$, $\ell = 1, \dots, r_k$ e $k = 1, \dots, s$) para o i -ésimo grupo é dada por

$$\begin{aligned} \mathcal{I}_{P_i}^{\beta_j \zeta_{k_\ell}}(\boldsymbol{\theta}) &= \mathbb{E} \left(- \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \beta_j \partial \zeta_{k_\ell}} \right) \\ &= \mathbb{E} \left(- \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \beta_j \partial \zeta_{k_\ell}} \right) \\ &= \mathbb{E} \left(\frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} \right) \\ &= \mathbb{E} \left\{ \mathbb{E} \left(\frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} \middle| \|\mathbf{P}_i\| \right) \right\}, \end{aligned}$$

em que

$$\begin{aligned} \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} &= \left(-2W_g(\delta_i) \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \left(-2W_g(\delta_i) \mathbf{n}_{kil}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \\ &= \left(-2W_g(\delta_i) \mathbf{n}_{kil}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \left(-2W_g(\delta_i) \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \\ &= 4W_g^2(\delta_i) \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{kil}^* \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\ &= 4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{kil}^* \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \end{aligned}$$

e

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} \middle| \|\mathbf{P}_i\|\right) &= \mathbb{E}\left\{4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \times \right. \\
&\quad \left. \mathbb{E}\left(\frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{kil}^* \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \right) \middle| \|\mathbf{P}_i\|\right\} \\
&= \mathbb{E}\left(4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \frac{1}{m_i} \text{tr}\left\{\boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{kil}^* \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2}\right\}\right) \\
&= \frac{1}{m_i} \text{tr}\left\{\mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{kil}^*\right\} \mathbb{E}\left(4W_g^2(\delta_i) \|\mathbf{P}_i\|^2\right) \\
&= \frac{4}{m_i} \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{kil}^* \mathbb{E}\left(W_g^2(\delta_i) \|\mathbf{P}_i\|^2\right) \\
&= \frac{4d_{g_i}}{m_i} \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{kil}^*,
\end{aligned}$$

em que \mathbf{n}_{kil}^{*T} corresponde a ℓ -ésima coluna da matriz de incidência \mathbf{N}_{ki} . Assim,

$$\boldsymbol{\mathcal{I}}_{\mathbf{P}_i}^{\beta_j \zeta_{k_\ell}}(\boldsymbol{\theta}) = \frac{4d_{g_i}}{m_i} \mathbf{x}_{ij}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{kil}^*, \quad (\text{C.12})$$

e, portanto, a matriz de informação de Fisher penalizada referente a $\boldsymbol{\beta}$ e \mathbf{f}_k será dada por

$$\boldsymbol{\mathcal{I}}_{\mathbf{P}}^{\beta \mathbf{f}_k}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{4d_{g_i}}{m_i} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{ki}. \quad (\text{C.13})$$

C.3.3 Prova de (3.20)

Por simplicidade, consideremos a seguinte notação:

$$\mathbf{f}_k = \begin{pmatrix} f_k(t_{k_1}^0) \\ \vdots \\ f_k(t_{k_{r_k}}^0) \end{pmatrix} = \begin{pmatrix} \zeta_{k_1} \\ \vdots \\ \zeta_{k_{r_k}} \end{pmatrix}, \quad k, k' = 1, \dots, s.$$

Caso para $k = k'$

De (C.5) e (C.9) temos que

$$\begin{aligned} \frac{\partial L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k} &= \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \mathbf{f}_k} - \frac{\alpha_k}{n} \frac{\partial(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \mathbf{f}_k} \\ &= -2W_g(\delta_i) \mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\varepsilon}_i - \frac{\alpha_k}{n} \frac{\partial(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \mathbf{f}_k} \\ &= -2W_g(\delta_i) \mathbf{N}_{ki}^T \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i - \frac{\alpha_k}{n} \frac{\partial(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \mathbf{f}_k}, \end{aligned}$$

e, portanto,

$$\begin{aligned} \frac{\partial L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \zeta_{k_j}} &= \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j}} - \frac{\alpha_k}{n} \frac{\partial(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \zeta_{k_j}} \\ &= -2W_g(\delta_i) \mathbf{n}_{kij}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i - \frac{\alpha_k}{n} \frac{\partial(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \zeta_{k_j}}, \end{aligned}$$

em que \mathbf{n}_{kij}^{*T} corresponde a j -ésima coluna da matriz de incidência \mathbf{N}_{ki} . Daqui segue que

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}} = \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}} - \frac{\alpha_k}{n} \frac{\partial^2(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}}.$$

Logo, a partição da matriz de informação de Fisher penalizada referente a ζ_{k_j} e ζ_{k_ℓ} ($j, \ell = 1, \dots, r_k$ e $k = 1, \dots, s$) para o i -ésimo grupo é dada por

$$\begin{aligned} \mathcal{I}_{P_i}^{\zeta_{k_j} \zeta_{k_\ell}}(\boldsymbol{\theta}) &= \mathbb{E} \left(- \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}} \right) + \frac{\alpha_k}{n} \frac{\partial^2(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}} \\ &= \mathbb{E} \left(\frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j}} \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} \right) + \frac{\alpha_k}{n} \frac{\partial^2(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}} \\ &= \mathbb{E} \left\{ \mathbb{E} \left(\frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j}} \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} \middle| \|\mathbf{P}_i\| \right) \right\} + \frac{\alpha_k}{n} \frac{\partial^2(\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}}, \end{aligned}$$

em que

$$\begin{aligned}
\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j}} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} &= \left(-2W_g(\delta_i) \mathbf{n}_{k_{ij}}^* \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \left(-2W_g(\delta_i) \mathbf{n}_{k_{i\ell}}^* \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right) \\
&= 4W_g^2(\delta_i) \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{k_{ij}}^* \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\
&= 4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{k_{ij}}^* \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|}
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{E} \left(\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j}} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \zeta_{k_\ell}} \middle| \|\mathbf{P}_i\| \right) &= \mathbb{E} \left\{ 4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \times \right. \\
&\quad \left. \mathbb{E} \left(\frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{k_{ij}}^* \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \right) \middle| \|\mathbf{P}_i\| \right\} \\
&= \mathbb{E} \left(4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \frac{1}{m_i} \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{k_{ij}}^* \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \right\} \right) \\
&= \frac{1}{m_i} \text{tr} \left\{ \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{k_{ij}}^* \right\} \mathbb{E} \left(4W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \right) \\
&= \frac{4}{m_i} \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{k_{ij}}^* \mathbb{E} \left(W_g^2(\delta_i) \|\mathbf{P}_i\|^2 \right) \\
&= \frac{4d_{g_i}}{m_i} \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{k_{ij}}^*.
\end{aligned}$$

Assim,

$$\mathcal{I}_{\mathbf{P}_i}^{\zeta_{k_j} \zeta_{k_\ell}}(\boldsymbol{\theta}) = \frac{4d_{g_i}}{m_i} \mathbf{n}_{k_{i\ell}}^{*T} \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{k_{ij}}^* + \frac{\alpha_k}{n} \frac{\partial^2 (\mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k)}{\partial \zeta_{k_j} \partial \zeta_{k_\ell}}, \quad (\text{C.14})$$

e, portanto, a matriz de informação de Fisher penalizada referente a \mathbf{f}_k será dada por

$$\mathcal{I}_{\mathbf{P}}^{\mathbf{f}_k \mathbf{f}_k}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{4d_{g_i}}{m_i} \mathbf{N}_{k_i}^T \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_{k_i} + \alpha_k \mathbf{K}_k. \quad (\text{C.15})$$

Caso para $k \neq k'$

Neste caso, devemos notar que $(k, k' = 1, \dots, s)$

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_k \partial \mathbf{f}_{k'}^T} = \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \mathbf{f}_k \partial \mathbf{f}_{k'}^T}$$

e

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \zeta_{k_j} \partial \zeta_{k'_\ell}} = \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j} \partial \zeta_{k'_\ell}}.$$

De aqui segue que

$$\mathbb{E} \left(- \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \zeta_{k_j} \partial \zeta_{k'_\ell}} \right) = \mathbb{E} \left(- \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \zeta_{k_j} \partial \zeta_{k'_\ell}} \right).$$

Logo, o procedimento para obter a partição da matriz de informação de Fisher penalizada referente a ζ_{k_j} e $\zeta_{k'_\ell}$ ($j, \ell = 1, \dots, r_k$ e $k = 1, \dots, s$) para o i -ésimo grupo, isto é, $\mathcal{I}_{P_i}^{\zeta_{k_j} \zeta_{k'_\ell}}(\boldsymbol{\theta})$, é análogo ao procedimento descrito para o caso $k = k'$. A partir desse resultado, podemos obter diretamente $\mathcal{I}_P^{\mathbf{f}_k \mathbf{f}_{k'}}(\boldsymbol{\theta})$.

C.3.4 Prova de (3.22) e (3.23)

Seja $\boldsymbol{\tau} = (\boldsymbol{\lambda}^T, \boldsymbol{\gamma}^T)^T = (\tau_1, \tau_2, \dots, \tau_{d^*})^T$, com $d^* = d_\lambda + d_\gamma$. De (C.7) e (C.9)

$$\begin{aligned} \frac{\partial L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \tau_j} &= \frac{\partial L_{P_i}(\boldsymbol{\theta})}{\partial \tau_j} \\ &= -\frac{1}{2} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \} - W_g(\delta_i) \boldsymbol{\varepsilon}_i^T \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \\ &= -\frac{1}{2} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \} - W_g(\delta_i) \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i. \end{aligned}$$

Daqui segue que

$$\frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \tau_j \partial \tau_\ell} = \frac{\partial^2 L_{P_i}(\boldsymbol{\theta})}{\partial \tau_j \partial \tau_\ell}.$$

Logo, a partição da matriz de informação de Fisher penalizada referente a τ_j e τ_ℓ ($j, \ell = 1, \dots, d^*$) para o i -ésimo grupo é dada por

$$\begin{aligned}
\mathcal{I}_{\mathbf{P}_i}^{\tau_j \tau_\ell}(\boldsymbol{\theta}) &= \mathbb{E} \left(- \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \tau_j \partial \tau_\ell} \right) \\
&= \mathbb{E} \left(- \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_j \partial \tau_\ell} \right) \\
&= \mathbb{E} \left(\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_j} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_\ell} \right) \\
&= \mathbb{E} \left\{ \mathbb{E} \left(\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_j} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_\ell} \middle| \|\mathbf{P}_i\| \right) \right\} \\
&= \mathbb{E} \left\{ \mathbb{E} (A_1 + A_2 + A_3 + A_4 | \|\mathbf{P}_i\|) \right\} \\
&= \mathbb{E} \left\{ \mathbb{E} (A_1 | \|\mathbf{P}_i\|) + \mathbb{E} (A_2 | \|\mathbf{P}_i\|) + \mathbb{E} (A_3 | \|\mathbf{P}_i\|) + \mathbb{E} (A_4 | \|\mathbf{P}_i\|) \right\},
\end{aligned}$$

em que

$$\begin{aligned}
\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_j} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_\ell} &= \frac{1}{4} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \} \\
&\quad + \frac{1}{2} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \} W_g(\delta_i) \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\
&\quad + \frac{1}{2} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \} W_g(\delta_i) \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\
&\quad + W_g^2(\delta_i) \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\
&= A_1 + A_2 + A_3 + A_4,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(A_1 | \|\mathbf{P}_i\|) &= \frac{1}{4} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \} \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \} \\
&= \frac{b_{ij\ell}}{4},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(A_2 | \|\mathbf{P}_i\|) &= \frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\} W_g(\delta_i) \mathbb{E}\left(\mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \middle| \|\mathbf{P}_i\|\right) \\
&= \frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\} W_g(\delta_i) \|\mathbf{P}_i\|^2 \mathbb{E}\left(\frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\|\right) \\
&= \frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\} W_g(\delta_i) \|\mathbf{P}_i\|^2 \frac{1}{m_i} \text{tr}\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\},
\end{aligned}$$

$$\mathbb{E}(A_3 | \|\mathbf{P}_i\|) = \mathbb{E}(A_2 | \|\mathbf{P}_i\|),$$

e

$$\begin{aligned}
\mathbb{E}(A_4 | \|\mathbf{P}_i\|) &= W_g^2(\delta_i) \mathbb{E}\left(\mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \mathbf{P}_i^T \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \middle| \|\mathbf{P}_i\|\right) \\
&= W_g^2(\delta_i) \|\mathbf{P}_i\|^4 \mathbb{E}\left(\frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \left[\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1/2}\right] \times \right. \\
&\quad \left. \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \frac{\mathbf{P}_i^T}{\|\mathbf{P}_i\|} \left[\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\right] \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\|\right) \\
&= \frac{W_g^2(\delta_i) \|\mathbf{P}_i\|^4}{m_i(m_i + 2)} \left[2 \text{tr}\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\} + \right. \\
&\quad \left. \text{tr}\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1/2}\} \text{tr}\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\} \right] \\
&= \frac{W_g^2(\delta_i) \|\mathbf{P}_i\|^4}{m_i(m_i + 2)} \left[\text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\} \text{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\} + \right. \\
&\quad \left. 2 \text{tr}\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\} \right] \\
&= \frac{W_g^2(\delta_i) \|\mathbf{P}_i\|^4}{m_i(m_i + 2)} \left[b_{i,j,\ell} + 2 \text{tr}\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\} \right].
\end{aligned}$$

Finalmente,

$$\begin{aligned}
\mathcal{I}_{\mathbf{P}_i}^{\tau_j \tau_\ell}(\boldsymbol{\theta}) &= \mathbb{E}\left\{\mathbb{E}(A_1|\|\mathbf{P}_i\|) + \mathbb{E}(A_2|\|\mathbf{P}_i\|) + \mathbb{E}(A_3|\|\mathbf{P}_i\|) + \mathbb{E}(A_4|\|\mathbf{P}_i\|)\right\} \\
&= \mathbb{E}\left\{\mathbb{E}(A_1|\|\mathbf{P}_i\|)\right\} + \mathbb{E}\left\{\mathbb{E}(A_2|\|\mathbf{P}_i\|)\right\} + \\
&\quad \mathbb{E}\left\{\mathbb{E}(A_3|\|\mathbf{P}_i\|)\right\} + \mathbb{E}\left\{\mathbb{E}(A_4|\|\mathbf{P}_i\|)\right\},
\end{aligned}$$

em que

$$\mathbb{E}\left\{\mathbb{E}(A_1|\|\mathbf{P}_i\|)\right\} = \frac{b_{i_j \ell}}{4},$$

$$\begin{aligned}
\mathbb{E}\left\{\mathbb{E}(A_2|\|\mathbf{P}_i\|)\right\} &= \frac{1}{2m_i} \text{tr}\left\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\right\} \text{tr}\left\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\right\} \mathbb{E}\left\{W_g(\delta_i) \|\mathbf{P}_i\|^2\right\} \\
&= \frac{1}{2m_i} \text{tr}\left\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\right\} \text{tr}\left\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j)\right\} \mathbb{E}\left\{W_g(\delta_i) \|\mathbf{P}_i\|^2\right\} \\
&= \frac{d_{i_j \ell}}{2m_i} \mathbb{E}\left\{W_g(\delta_i) \|\mathbf{P}_i\|^2\right\} \\
&= \frac{d_{i_j \ell}}{2m_i} \left(-\frac{m_i}{2}\right) \\
&= -\frac{b_{i_j \ell}}{4},
\end{aligned}$$

$$\mathbb{E}\left\{\mathbb{E}(A_3|\|\mathbf{P}_i\|)\right\} = \mathbb{E}\left\{\mathbb{E}(A_2|\|\mathbf{P}_i\|)\right\},$$

e

$$\begin{aligned}
\mathbb{E}\left\{\mathbb{E}(A_4|\|\mathbf{P}_i\|)\right\} &= \frac{\mathbb{E}\left\{W_g^2(\delta_i) \|\mathbf{P}_i\|^4\right\}}{m_i(m_i + 2)} \left[b_{i_j \ell} + 2 \text{tr}\left\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\right\}\right] \\
&= \frac{f_{g_i}}{m_i(m_i + 2)} \left[b_{i_j \ell} + 2 \text{tr}\left\{\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1/2}\right\}\right].
\end{aligned}$$

Logo, após algumas manipulações de álgebra, obtemos que

$$\mathcal{I}_{\mathbf{P}_i}^{\tau_j \tau_\ell}(\boldsymbol{\theta}) = \frac{b_{i_j \ell}}{4} \left(\frac{4f_{g_i}}{m_i(m_i + 2)} - 1 \right) + \frac{2f_{g_i}}{m_i(m_i + 2)} \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \right\}.$$

C.3.5 Prova de (3.22) e (3.23)

A partir de

$$\begin{aligned} \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \beta_j \partial \tau_\ell} &= \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_j \partial \tau_\ell} \\ &= -2W_g(\delta_i) \mathbf{x}_{i_j}^{*T} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(j) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i, \end{aligned}$$

temos que

$$\begin{aligned} \mathcal{I}_{\mathbf{P}_i}^{\beta_j \tau_\ell}(\boldsymbol{\theta}) &= \mathbb{E} \left(- \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \beta_j \partial \tau_\ell} \right) \\ &= \mathbb{E} \left\{ \mathbb{E} \left(- \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_j \partial \tau_\ell} \middle| \|\mathbf{P}_i\| \right) \right\} \\ &= \mathbb{E} \left\{ \mathbb{E} \left(\frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \beta_j} \frac{\partial L_{\mathbf{P}_i}(\boldsymbol{\theta})}{\partial \tau_\ell} \middle| \|\mathbf{P}_i\| \right) \right\} \\ &= 0. \end{aligned}$$

Da mesma maneira podemos obter que

$$\mathcal{I}_{\mathbf{P}_i}^{\zeta_{k_j} \tau_\ell}(\boldsymbol{\theta}) = \mathbb{E} \left(- \frac{\partial^2 L_{\mathbf{P}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \zeta_{k_j} \partial \tau_\ell} \right). \quad (\text{C.16})$$

Apêndice D

Dados de fosfato

Tabela D.1: Medições de fosfato inorgânico no plasma obtidas a partir de amostras de sangue de 20 pacientes obesos, registradas após da aplicação oral de uma dose de glucose padrão (mg/dd).

Paciente	Tempo (em horas)							
	0	0.5	1	1.5	2	3	4	5
1	4,3	3,3	3	2,6	2,2	2,5	2,4	3,4
2	5	4,9	4,1	3,7	3,7	4,1	4,7	4,9
3	4,6	4,4	3,9	3,9	3,7	4,2	4,8	5
4	4,3	3,9	3,1	3,1	3,1	3,1	3,6	4
5	3,1	3,1	3,3	2,6	2,6	1,9	2,3	2,7
6	4,8	5	2,9	2,8	2,2	3,1	3,5	3,6
7	3,7	3,1	3,3	2,8	2,9	3,6	4,3	4,4
8	5,4	4,7	3,9	4,1	2,8	3,7	3,5	3,7
9	3	2,5	2,3	2,2	2,1	2,6	3,2	3,5
10	4,9	5	4,1	3,7	3,7	4,1	4,7	4,9
11	4,8	4,3	4,7	4,6	4,7	3,7	3,6	3,9
12	4,4	4,2	4,2	3,4	3,5	3,4	3,9	4
13	4,9	4,3	4	4	3,3	4,1	4,2	4,3
14	5,1	4,1	4,6	4,1	3,4	4,2	4,4	4,9
15	4,8	4,6	4,6	4,4	4,1	4	3,8	3,8
16	4,2	3,5	3,8	3,6	3,3	3,1	3,5	3,9
17	6,6	6,1	5,2	4,1	4,3	3,8	4,2	4,8
18	3,6	3,4	3,1	2,8	2,1	2,4	2,5	3,5
19	4,5	4	3,7	3,3	2,4	2,3	3,1	3,3
20	4,6	4,4	3,8	3,8	3,8	3,6	3,8	3,8

Apêndice E

Dados pressão ocular

Tabela E.1: Medições da pressão ocular do olho esquerdo de 20 pacientes registradas em intervalos de tempo de três horas.

Paciente	Tempo (em horas)						
	6	9	12	15	18	21	24
1	22	18	17	15	17	18	16
2	20	18	19	17	16	18	14
3	22	25	21	29	27	21	19
4	20	23	23	18	24	18	19
5	24	24	21	24	21	17	17
6	21	18	19	17	18	17	16
7	22	20	20	22	22	18	19
8	28	22	19	19	20	20	23
9	15	18	16	16	18	15	17
10	25	19	22	20	20	18	20
11	25	23	20	22	23	20	23
12	23	22	22	20	22	20	22
13	17	21	19	18	18	18	17
14	19	16	15	16	16	16	17
15	21	13	18	16	18	18	22
16	22	17	20	18	18	19	20
17	15	14	14	14	14	13	17
18	17	16	14	15	16	16	14
19	18	20	18	16	17	20	13
20	17	17	15	19	13	16	13
21	15	14	14	14	14	17	15
22	17	19	18	16	20	18	14
23	22	18	24	19	20	21	20
24	27	24	23	23	20	23	27
25	24	16	22	16	14	20	16
26	30	21	16	18	18	18	21
27	17	22	20	20	20	18	15
28	30	24	24	20	22	22	30
29	17	22	22	20	20	14	16

Referências Bibliográficas

- [1] Akaike, H. (1974). A new look at statistical models identification. *IEEE Transactions on Automatic Control* **AU-19**, 716-722.
- [2] Arellano, R. (1994). *Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão*. Tese de Doutorado, Departamento de Estatística, Universidade de São Paulo.
- [3] Banerjee, M. e Frees, E. W. (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association*, **92**, 999-1005.
- [4] Barnett, V. e Lewis, T. (1994). *Outliers in Statistical Data*. John Wiley, Chichester.
- [5] Belsley, D. A., Kuh, E. e Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley, New York.
- [6] Berhane, K. e Tibshirani, J. (1998). Generalized additive models for longitudinal data. *The Canadian Journal of Statistics*, **26**, 517-535.
- [7] Bianco, A., Boente, G. e Martínez, E. (2006). Robust tests in semiparametric partly linear models. *Scandinavian Journal of Statistics*, **33**, 435-450.
- [8] Billor, N. e Loynes, R. M. (1993). Local influence: a new approach. *Communications in Statistics, Theory and Methods*, **22**, 1595-1611.
- [9] Breslow, N. E. e Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.
- [10] Buja, A., Hastie, T. e Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, **17**, 453-555.
- [11] Burnham, K. P. e Anderson, D. R. (1998). *Model Selection and Inference*. Springer-Verlag, New York.

-
- [12] Caroni, C. (1987). Residuals and influence in the multivariate linear model. *The Statistician*, **36**, 365-370.
- [13] Cao, C.-Z. e Lin, J.-G. (2009). Diagnostics for elliptical linear mixed models with first-order autoregressive errors. *Journal of Statistical Computation and Simulation*. Em impressão.
- [14] Chang, X. e Qu, L. (2004). Wavelet estimation of partially linear models. *Computational Statistics and Data Analysis*, **47**, 31-48.
- [15] Chatterjee, S. e Hadi, A. (1988). *Sensitivity Analysis in Linear Regression*. John Wiley, New York.
- [16] Chi, E. M. e Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association, Theory and Methods*, **84**, 452-459.
- [17] Chi, E. M. e Reinsel, G. C. (1991). Asymptotic properties of the score test for autocorrelation in a random effects with AR(1) errors models. *Statistics and Probability Letters*, **11**, 453-457.
- [18] Christensen, R., Pearson, L. M. e Johnson, W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, **34**, 38-45.
- [19] Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15-18.
- [20] Cook, R. D. e Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, New York.
- [21] Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B*, **48**, 133-169.
- [22] Cook, R. D. (1997). Local influence. Em Kotz, S., Read, C. B. e Banks, D. L. (Eds.), *Encyclopedia of Statistical Sciences, Update, Vol. 1*, 380-385, Wiley.
- [23] Cordeiro, G. M. e Paula, G. A. (1992). Estimation, large-sample parametric tests and diagnostics for non-exponential family nonlinear models. *Communications in Statistics, Simulation and Computation*, **21**, 149-172.
- [24] Cox, D. R. e Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B*, **30**, 248-275.

-
- [25] Craven, P. e Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerical Mathematical*, **31**, 377-403.
- [26] De Montricher, G. F., Tapia, R. A. e Thompson, J. R. (1975). Nonparametric maximum likelihood estimation of probability densities by penalty function methods. *The Annals of Statistics*, **3**, 1329-1348.
- [27] Dempster, A. P., Laird, N. M. e Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, **39**, 1-38.
- [28] Dempster, A. P., Rubin, D. B. e Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, **76**, 341-353.
- [29] Díaz, G. J., Galea, R. M. e Leiva, S. V. (2003). Influence diagnostics for elliptical multivariate linear regression models. *Communications in Statistics, Theory and Methods*, **32**, 625-641.
- [30] Durban, M., Hackett, C. A. e Currie, I. D. (1999). Approximate standard errors in semiparametric models. *Biometrics*, **55**, 699-703.
- [31] Durban, M., Hackett, C. A., McNicol, J. W., Newton, A. C., Thomas, W. T. B. e Currie, I. D. (2003). The practical use of semiparametric models in field trials. *Journal of Agricultural, Biological, and Environmental Statistics*, **8**, 48-66.
- [32] Eaton, M. L. (1983). *Multivariate Statistics*. John Wiley, New York.
- [33] Engle, R. F., Granger, C. W. J., Rice, J. e Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, **81**, 310-320.
- [34] Escobar, E. e Meeker, W. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, **48**, 507-528.
- [35] Eubank, R. L. (1984). The hat matrix for smoothing splines. *Statistics and Probability Letters*, **2**, 9-14.
- [36] Eubank, R. L. (1985). Diagnostics for smoothing splines. *Journal of the Royal Statistical Society B*, **47**, 332-341.
- [37] Eubank, R. L. e Gunst, R. F. (1986). Diagnostics for penalized least-squares estimators. *Statistics and Probability Letters*, **4**, 265-272.

-
- [38] Eubank, R. L. (1988). *Smoothing Splines and Nonparametric Regression*. Marcel Dekker, New York and Basel.
- [39] Eubank, R. L. e Thomas, W. (1993). Detecting heteroscedasticity in nonparametric regression. *Journal of the Royal Statistical Society B*, **55**, 145-155.
- [40] Fahrmeir, L. e Lang, S. (2001). Bayesian inference for generalized additive mixed models based on markov random field priors. *Applied Statistics*, **50**, 201-220.
- [41] Fang, K.T., Kotz, S. e Ng, K.W.(1990). *Symmetric Multivariate and Related Distribution*. Chapman and Hall, London.
- [42] Fang, K. T. e Zhang, Y. T. (1990). *Generalized Multivariate Analysis*. Springer-Verlag e Science Press, Berlin e Beijing.
- [43] Flanders, W. D., Klein, M. e Tolbert, P. (2005). *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 246-257.
- [44] Fuller, W. (1987). *Measurement Error Models*. Wiley, New York.
- [45] Fung, W. K. e Kwan, C. W. (1997). A note on local influence based on normal curvature. *Journal of the Royal Statistical Society B*, **59**, 839-843.
- [46] Fung, W., Zhu, Z., Wei, B. e He, X. (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Association B*, **64**, 565-579.
- [47] Galea, M. (1995). *Calibração Comparativa Estrutural e Funcional*. Tese de Doutorado, Departamento de Estatística, Universidade de São Paulo.
- [48] Galea, M., Paula, G. e Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, **46**, 71-79.
- [49] Galea, M., Riquelme, M. e Paula, G. A. (2000). Diagnostic methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics*, **14**, 167-184.
- [50] Galea, M., Bolfarine, H. e Vilca, F. (2005). Local influence in comparative calibration models under elliptical t-distributions. *Biometrical Journal*, **47**, 691-706
- [51] Gannaz, I. (2007). Robust estimation and wavelet thresholding in partially models. *Stat Comput* (2007), **17**, 239-310.

-
- [52] Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley, New York.
- [53] Good, I. J. e Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika*, **58**, 255-277.
- [54] Green, P. J. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika*, **72**, 527-537.
- [55] Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, **55**, 245-259.
- [56] Green, P. J. (1990). On use of the EM algorithm for penalized likelihood estimation. *Journal of the Royal Statistical Society B*, **52**, 443-452.
- [57] Green, P. J. e Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, Boca Raton.
- [58] Gu, C. (1992). Diagnostics for nonparametric regression models with additive terms. *Journal of the American Statistical Association*, **87**, 1051-1058.
- [59] Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer-Verlag, New York.
- [60] Gu, C. e Ma, P. (2005). Optimal smoothing in nonparametric mixed-effect models. *The Annals of Statistics*, **33**, 1357-1379.
- [61] Hall, P. e Titterton, D. (1987). Common structure of techniques for choosing smoothing parameters in regression problems. *Journal of the Royal Statistical Society B*, **49**, 184-198.
- [62] Hamilton, S. Truong, Y. (1997). Local estimation in partly linear models. *Journal Multivariate Analysis*, **60**, 1-19.
- [63] Hand, D. e Crowder, M. (1996). *Practical Longitudinal Data Analysis*. Chapman and Hall, Boca Raton.
- [64] Hardle, W., Mammen, E. e Muller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, **93**, 1461-1474.
- [65] Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*, **4**, 384-395.

-
- [66] Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association, Theory and Methods*, **72**, 320-342.
- [67] Hastie, T. e Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, **1**, 293-310.
- [68] Hastie, T. e Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, **82**, 371-386.
- [69] Hastie, T. e Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B*, **55**, 757-796.
- [70] Hastie, T. e Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- [71] Hawkins, D. H. (1980). *Identification of Outliers*. Chapman and Hall, London.
- [72] He, X. e Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *Journal of Multivariate Analysis*, **58**, 162-181.
- [73] He, X., Zhu, Z. e Fung, W. K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, **89**, 579-590.
- [74] He, X., Fung, W. K. e Zhu, Z. (2005). Robust estimation in generalied partial linear models for clustered data. *Journal of the American Statistical Association*, **100**, 1176-1184.
- [75] Heckman, N. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society B*, **48**, 244-248.
- [76] Heckman, N. (1988). Minimax estimates in a semiparametric model. *Journal of the American Statistical Association*, **83**, 1090-1096.
- [77] Hurvich, C. M., Simonoff, J. S. e Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society B*, **60**, 271-293.
- [78] Ibacache, G. (2004). *Influencia Local en el Modelo de Regresión Multivariado Simple Equicorrelacionado*. Tesis para Optar al Grado Académico de Magíster en Estadística.

-
- [79] Johnson, N. L., Kotz, S. e Balakrishnan, N. (1975). *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley, New York.
- [80] Ke, C. e Wang, Y. (2001). Semiparametric nonlinear mixed-effects models and their applications. *Journal of the American Statistical Association*, **96**, 1272-1281.
- [81] Keramides, E. M. e Lee, J. C. (1995). Selection of a covariance structure for growth-curves. *Biometrical Journal*, **37**, 783-797.
- [82] Kim, C. (1996). Cook's distance in spline smoothing. *Statistics and Probability Letters*, **31**, 139-144.
- [83] Kim, C., Park, B. e Kim, W. (2002). Influence diagnostics in semiparametric regression models. *Statistics and Probability Papers*, **60**, 49-58.
- [84] Kohn, R., Ansley, C. F. e Tharm, D. (1991). The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of the American Statistical Association*, **86**, 1042-1050.
- [85] Kowalski, J., Mendoza-Blanco, J. R., Tu, X. M. e Gleser, L. J. (1999). On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regressions. *Communications in Statistics, Theory and Methods*, **28**, 2119-2140.
- [86] Laird, N. M e Ware, J. M. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.
- [87] Laird, N., Lange, N. e Stram, D. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association*, **82**, 97-105.
- [88] Lange, K. L., Little, R. J. A. e Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**, 881-896.
- [89] Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer, New York.
- [90] Lesaffre, E. e Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570-582.

-
- [91] Liang, H. (2006). Checking linearity of non-parametric component in partially linear models with an application in systemic inflammatory response syndrome study. *Statistical Methods in Medical Research*, **15**, 273-284.
- [92] Lin, X. e Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, **91**, 1007-1016.
- [93] Lin, X. e Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society B*, **61**, 381-400.
- [94] Lindstrom, M. J. e Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673-687.
- [95] Little, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, **37**, 23-38.
- [96] Liu, S. Z. (2000). On local influence for elliptical linear models. *Statistical Papers*, **41**, 211-224.
- [97] Liu, S. (2002). Local influence in multivariate elliptical linear regression models. *Linear Algebra and its Applications*, **354**, 159-174.
- [98] Liu, S. (2004). On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. *Journal of Applied Probability, Stochastic Methods and their Applications*, **41**, 393-405.
- [99] Lombardía, M. J. e Sperlich, S. (2008). Semiparametric inference in generalized mixed effects models. *Journal of the Royal Statistical Society B*, **70**, 913-930.
- [100] Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society B*, **44**, 226-233.
- [101] Lu, B. e Song, X. Y. (2006). Local influence analysis of multivariate probit latent variable models. *Journal of Multivariate Analysis*, **97**, 1783-1798.
- [102] Ma, Y., Chiou, J. M. e Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika*, **93**, 75-84.
- [103] Magnus, J. R. e Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.

-
- [104] Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, **4**, 51-67.
- [105] Meng, X.-L. e Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association*, **86**, 899-909.
- [106] Nychka, D. (1990). Some properties of adding a smoothing step to the EM algorithm. *Statistics and Letters*, **9**, 187-193.
- [107] Nobre, J. S. (2004). *Métodos de Diagnóstico para Modelos Lineares Mistos*. Dissertação de Mestrado, Departamento de Estatística, Universidade de São Paulo.
- [108] Osorio, F. (2006). *Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos*. Tese de Doutorado, Departamento de Estatística, Universidade de São Paulo.
- [109] Osorio, F., Paula, G. A. e Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, **51**, 4354-4368.
- [110] Opsomer, J. D. e Ruppert, D. (1999). A root-n consistent backfitting estimator for semiparametric additive modeling. *Journal of Computational and Graphical Statistics*, **8**, 715-732.
- [111] O'Sullivan, F., Yandell, B. e Raynor, W. (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association, Theory and Methods*, **81**, 96-103.
- [112] Ouwens, M., Tan, F. e Berger, M. (2001). Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, **57**, 1166-1172.
- [113] Pan, J. X. e Fang, K.T. (2000). *Growth Curve Models and Statistical Diagnostics*. Springer, New York.
- [114] Parise, H., Wand, M. P., Ruppert, D. e Ryan, L. (2001). Incorporation of historical controls using semiparametric mixed models. *Applied Statistics*, **50**, 31-42.
- [115] Paula, G. A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis*, **16**, 63-79.

-
- [116] Paula, G. A. e Peres, C. A. (1988). Diagnostics for GLMs with linear inequality parameter constraints. *Communications in Statistics, Theory and Methods*, **17**, 4205-4219.
- [117] Pauler, Donna. K. (1998). The Schwarz criterion and related methods for normal linear models. *Biometrika*, **85**, 13-27.
- [118] Peña, D. (2005). A new statistic for influence in linear regression. *Journal of the American Statistical Association*, **47**, 1-11.
- [119] Pinheiro, J. e Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, **4**, 12-35.
- [120] Pinheiro, J. e Bates, D. (2000). *Mixed-effects Models in S and S-PLUS*. Springer-Verlag, New York.
- [121] Pinheiro, J., Liu, C. e Wu, Y. N. (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.
- [122] Pitrun, I., King, M. L. e Zhang, X. (2006). Smoothing spline based tests for non-linearity in a partially linear model. *Journal of Statistical Planning and inference*, **136**, 2446-2469.
- [123] Polasek, W. (1984). Regression diagnostics for general linear regression models. *Journal of the American Statistical Association*, **79**, 336-340.
- [124] Poon, W. e Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B*, **61**, 51-61.
- [125] Reinsch, C. (1967). Smoothing by spline functions. *Numerical Mathematical*, **10**, 177-183.
- [126] Reinsel, G. (1984). Estimation and prediction in a multivariate random effects generalized linear model. *Journal of the American Statistical Association*, **79**, 406-414.
- [127] Rice, J. (1986). Convergence rates for partially splines models. *Statistics and Probability Letters*, **4**, 203-208.
- [128] Richardson, A. M. (1997). Bounded influence estimation in the mixed linear model. *Journal of the American Statistical Association*, **92**, 154-161.

-
- [129] Rigby, R. e Stasinopoulos, D. (2005). Generalized additive models for location, scale and shape. *Applied Statistical*, **54**, 507-554.
- [130] Robinson, P. (1988). Root n-consistent semiparametric regression. *Econometrica*, **56**, 931-954.
- [131] Rousseeuw, P. J. e Leroy, A. M. (1987). *Robust Regression and Outliers Detection*. John Wiley, New York.
- [132] Rubin, D. B. (1983). *Iteratively Reweighted Least Squares*. Encyclopedia of the Statistical Sciences (Vol 4), John Wiley, New York.
- [133] Savalli, C. (2006). *Teste do Tipo Escore para componentes de Variância em modelos Elípticos Lineares Mistos*. Tese de Doutorado, Departamento de Estatística, Universidade de São Paulo.
- [134] Savalli, C., Paula, G. A. e Cysneiros, F. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, **6**, 59-76.
- [135] Schoenberg, I. J. (1964). Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences of the United States of America*, **52**, 947-950.
- [136] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.
- [137] Segal, M. R., Bacchetti, P. e Jewell, N. P. (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm. *Journal of the Royal Statistical Society B*, **56**, 345-352.
- [138] Shen, X. (1997). On methods of sieves and penalization. *The Annals of Statistics*, **25**, 2555-2591.
- [139] Shiau, Jyh-Jen. H. e Wahba. G. (1988). Rates of convergence of some estimators for a semiparametric model. *Communications in Statistics, Simulation and Computation*, **17**, 1117-1133.
- [140] Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *The Annals of Statistics*, **10**, 795-810.
- [141] Silverman, B. W. (1984). A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *Journal of the American Statistical Association*, **79**, 584-589.

-
- [142] Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society B*, **47**, 1-52.
- [143] Simonoff, J. S. e Tsai, C.-L. (1999). Semiparametric and additive model selection using an improved akaike information criterion. *Journal of Computational and Graphical Statistics*, **8**, 22-40.
- [144] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society B*, **50**, 413-436.
- [145] Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, **13**, 689-705.
- [146] Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, **14**, 590-606.
- [147] Tan, F. E. S., Ouwens, M. J. N. e Berger, M. P. F. (2001). Detection of influential observations in longitudinal mixed effects regression models. *The Statistician*, **50**, 271-284.
- [148] Tang, N. S., Wei, B. C., e Zhang, W. Z. (2006). Influence diagnostics in nonlinear reproductive dispersion mixed models. *Statistics*, **40**, 227-246.
- [149] Tapia, R. A. e Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*. The Johns Hopkins University Press, Baltimore e London.
- [150] Thomas, W. (1991). Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *Journal of the American Statistical Association*, **86**, 693-698.
- [151] Verbeke, G. e Molenberghs, G. (2001). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- [152] Villegas, C. (2002). *Influencia Local en Modelos Lineales Generalizados*. Tesis para Optar al Grado Académico de Magíster en Estadística.
- [153] Vonesh, E. F. e Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, **48**, 1-17.
- [154] Wahba, G. e Wold, S. (1975). A completely automatic french curve: fitting splines functions by cross-validation. *Communications in Statistics*, **4**, 1-17.

-
- [155] Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society B*, **40**, 364-372.
- [156] Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society B*, **45**, 133-150.
- [157] Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, **13**, 1378-1402.
- [158] Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society B*, **60**, 159-174.
- [159] Wecker, W. E. e Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, **78**, 81-89.
- [160] Wei, W. H. (2004). Derivatives diagnostics and robustness for smoothing splines. *Computational Statistics and Data Analysis*, **46**, 335-356.
- [161] Wegman, E. J. e Wright, I. W. (1983). Splines in statistics. *Journal of the American Statistical Association*, **78**, 351-365.
- [162] Zeger, S. e Diggle, P. (1994). Semiparametric models for longitudinal data with application to CDA cell numbers in HIV Seroconverters. *Biometrics*, **50**, 689-699.
- [163] Zerbe, G. O. (1979). Randomization analysis of the completely randomized design extended to growth and response curves. *Journal of the American Statistical Association, Theory and Methods*, **74**, 215-221.
- [164] Zhang, D., Lin, X. e Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association, Theory and Methods*, **93**, 710-719.
- [165] Zhang, D., Lin, X., Raz, Y. e Sowers, M. (2000). Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics*, **56**, 31-39.
- [166] Zhu, H., Lee, S. Y., Wei, B. C. e Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika*, **88**, 727-737.

- [167] Zhu, H. T. e Lee, S. Y. (2003). Local influence for generalized linear mixed models. *The Canadian Journal of Statistics*, **31**, 293-309.
- [168] Zhu, H. e Zhang, H. (2004). A diagnostic procedure based on local influence. *Biometrika*, **91**, 579-589.
- [169] Zhu, H., Ibrahim, J. G., Lee, S. e Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics*, **35**, 2565-2588.
- [170] Zhu, Z., He, X. e Fung, W. (2003). Local influence analysis for penalized gaussian likelihood estimators in partially linear models. *Scandinavian Journal of Statistics*, **30**, 767-780.