# Modelos mistos lineares elípticos com erros de medição

Joelmir André Borssoi

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

Coorientador: Prof. Dr. Manuel Jesús Galea Rojas

O autor recebeu auxílio financeiro da CAPES e do CNPq.

São Paulo,
Abril de 2014

# Modelos Mistos Lineares Elípticos com Erros de Medição

Este exemplar corresponde à versão final
da tese devidamente corrigida,
defendida por Joelmir André Borssoi
e aprovada pela Comissão Julgadora em 20/02/2014.

Banca Examinadora:

- Prof. Dr. Gilberto Alvarenga Paula (Orientador) (IME-USP)
- Prof. Dr. Manuel Jesús Galea Rojas (Coorientador) (PUC-Chile)
- Prof. Dr. Mario de Castro Andrade Filho (ICMC-USP)
- Prof. Dr. Miguel Angel Uribe Opazo (UNIOESTE)
- Prof. Dr. Filidor Edilfonso Vilca Labra (UNICAMP)

*Ao meu pai, Adelino (in memoriam).*

# Agradecimentos

Ao final de mais uma importante etapa da minha vida tenho muito a agradecer a quem esteve comigo e participou desta caminhada.

Em primeiro lugar, quero agradecer a Deus pelo dom da vida, por iluminar-me nos momentos mais difíceis e pelas pessoas que colocou em meu caminho, antes e durante este período do doutorado.

Meus mais sinceros agradecimentos ao meu orientador, Prof. Dr. Gilberto Alvarenga Paula, pela confiança, auxílio, ensinamentos e apoio a mim dedicados durante todo o desenvolvimento deste trabalho. É uma honra poder dizer que fui orientado pelo senhor.

Agradeço também ao meu coorientador, Prof. Dr. Manuel Galea, pela amizade, apoio e ensinamentos que tenho recebido desde os tempos de mestrado. O senhor sempre foi um grande incentivador e é uma honra poder trabalharmos juntos.

Agradeço de forma muito especial a minha amada esposa, Pâmela, pelo companheirismo, paciência, compreensão e incentivo que nunca me faltaram. Você, melhor do que ninguém, sabe tudo o que passamos até a conclusão deste trabalho. Não tenho palavras para expressar o quanto você foi e é importante para mim nessa caminhada... te amo!

Quero agradecer, também de forma especial, aos meus familiares: à minha mãe, Terezinha, que junto com meu pai são minha base, minha referência, minha inspiração e meus grandes incentivadores desde antes das séries iniciais. Sem o incentivo de vocês não teria chegado até aqui. Também às minhas irmãs (e cunhados) e meus irmãos (e cunhadas): Adriana (e Robinson), Tatiani (e Denis), Adilson (e Andréia), Marinho (e Nelsy) pelo carinho, apoio e incentivo que sempre recebi de vocês.

Expresso meus agradecimentos ao Prof. Dr. Miguel Angel Uribe Opazo, pela amizade e ensinamentos desde a graduação e pelos incentivos para que seguisse a carreira acadêmica e ingressasse neste doutorado.

Gostaria de agradecer a todos os colegas e amigos que fiz no Instituto de Matemática e Estatística da USP, pela amizade e companheirismo, tanto nos estudos quanto nos agradáveis momentos do "café"e do futebol. Em especial ao Wagner Souza, Tiago Vargas, Michel Helcias, Camila Bertini, Alice Morais e Tiago Magalhães.

Finalmente, agradeço à Universidade de São Paulo, pela oportunidade da formação

acadêmica; aos professores do Instituto de Matemática e Estatística, pelos ensinamentos durante o doutorado; e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio financeiro por meio de bolsa de estudos.

# Resumo

O objetivo principal deste trabalho é estudar modelos mistos lineares elípticos em que uma das variáveis explicativas ou covariáveis é medida com erros, sob a abordagem estrutural. O trabalho é apresentado numa notação longitudinal, todavia a covariável medida com erros pode ser observada temporalmente ou como medidas repetidas. Assumimos uma estrutura hierárquica apropriada com distribuição elíptica conjunta para os erros envolvidos, porém a inferência é desenvolvida sob uma abordagem marginal em que consideramos a distribuição marginal da resposta e da variável medida com erros. Procedimentos de influência local em que o esquema de perturbação é escolhido de forma apropriada são desenvolvidos. Um exemplo para motivação é apresentado e analisado através dos procedimentos apresentados neste trabalho. Detalhamos nos apêndices os principais procedimentos necessários para o desenvolvimento do modelo proposto.

**Palavras-chave:** Métodos de diagnóstico, métodos robustos, modelos com erros nas variáveis, modelos elípticos, modelos mistos.

# Abstract

The aim of this thesis is to study elliptical linear mixed models in which one of the explanatory variables is subject to measurement error under the structural assumption. The work is presented by assuming a longitudinal structure, however the explanatory variable may be observed along the time or as repeated measures. A joint hierarchical structure is assumed for the elliptical errors, but the inference is made under the marginal structure. The methodology of local influence is applied with the perturbation schemes being selected appropriately. A motivation example is presented and analysed by the procedures developed in this work. All the main derivations for the development of the proposed model are presented in the appendices.

**Keywords:** Elliptical models, diagnostic methods, measurement error models, mixed models, robust methods.

# Sumário

# Lista de Figuras

# Lista de Tabelas

# Capítulo 1

# Introdução

## 1.1 Introdução

Em muitos estudos se faz necessário acompanhar uma amostra de indivíduos por um período de tempo e, para cada indivíduo, algumas ou todas as variáveis são medidas em múltiplos períodos de tempo. Esses estudos são chamados de estudos longitudinais. Dados longitudinais são um tipo especial de dados agrupados, aqueles em que o tempo é um componente importante. Modelos para dados longitudinais são frequentemente analisados usando técnicas de modelos lineares mistos. Esses modelos têm sido utilizados para estudar problemas em diversas áreas de pesquisa como agricultura, biologia, economia, geofísica e ciências sociais (Diggle et al., 1994), e o potencial de aplicação é explicado pela flexibilidade que oferecem para estudar a correlação entre e intraunidades amostrais, comumente encontradas em dados longitudinais (Laird e Ware, 1982).

Uma classe de modelos denominada "modelos mistos lineares elípticos" foi proposta por Savalli *et al.* (2006), em que o modelo marginal é também elíptico. Essa proposta traz inúmeras vantagens, por exemplo, no desenvolvimento de procedimentos de estimação, metodologias de diagnóstico e testes para os componentes de variância e pode ser interpretada como uma generalização do modelo misto linear normal no sentido de flexibilização da curtose da distribuição dos erros. Outra vantagem é que quando os erros têm distribuição com caudas mais pesadas do que a distribuição normal, as estimativas de máxima verossimilhança dos parâmetros envolvidos são mais robustas contra observações aberrantes, no sentido da distância de Mahalanobis. Em Osorio *et al.* (2007) foram derivadas as curvaturas normais de influência local para vários esquemas de perturbação para a classe de modelos mistos lineares elípticos. Russo *et al.* (2009) estenderam a classe proposta por Savalli *et al.* (2006) substituindo o efeito fixo linear por um efeito fixo não linear, criando assim a classe denominada "modelos mistos parcialmente não lineares elípticos", para os quais desenvolveram procedimentos de estimação e metodologias de diagnóstico. Russo *et al.* (2012) desenvolveram para

essa mesma classe testes para os componentes de variância através de uma estatística tipo escore proposta por Silvapulle e Silvapulle (1995); estudos de sensibilidade da estatística do teste e vários estudos de simulação para avaliar os impactos da classificação incorreta da curtose no tamanho e poder do teste foram apresentados.

Ainda para os modelos mistos parcialmente não lineares elípticos, Russo *et al.* (2012) propuseram uma estrutura geral para as matrizes de variância-covariância dos erros e efeitos aleatórios, incluindo como casos particulares estruturas autoregressivas e heteroscedásticas; procedimentos de estimação e metodologias de diagnóstico foram também desenvolvidos. Ibacache-Pulgar *et al.* (2012) apresentaram recentemente uma outra extensão da classe proposta por Savalli *et al.* (2006), em que um componente fixo não paramétrico é adicionado aos efeitos fixos e aleatórios criando assim os "modelos mistos semiparamétricos elípticos", assumindo que o componente não paramétrico é do tipo *B-spline* cúbica. Um procedimento de estimação tipo *back-fitting* foi desenvolvido para a estimação dos parâmetros envolvidos e verificou-se que as estimativas, inclusive do componente não paramétrico, são robustas contra observações aberrantes como no caso paramétrico. Curvaturas de influência local foram derivadas para alguns esquemas de perturbação e algumas aplicações foram apresentadas. Ibacache-Pulgar e Paula (2011) apresentaram um estudo sobre a existência e unicidade das estimativas de máxima verossimilhança em modelos semiparamétricos $t$ de Student.

Erros de medidas nas variáveis são comuns na prática. Por exemplo, pressão arterial, ingestão de gordura em estudos nutricionais e registros de depressão podem todos serem medidos com erros, uma vez que é, em geral, difícil medir com precisão essas variáveis. Na presença de erros de medição, os dados observados não são os verdadeiros valores, mas os "mal medidos". Se as covariáveis são medidas com erros e são tratadas como variáveis verdadeiras, a inferência estatística poderá ser enganosa na medida em que, por exemplo, uma covariável significativa pode ser considerada não significativa (Wu, 2010). Além disso, se os erros de medição não forem levados em conta, os estimadores de máxima verossimilhança são geralmente viesados e inconsistentes (Fuller, 1987).

Portanto, é importante investigar a combinação de efeitos aleatórios e erros de medição em modelos lineares. Neste sentido, Zhong *et al.* (2002), por exemplo, estudaram a combinação dos modelos lineares mistos com variáveis medidas com erros utilizando a função escore corrigida de Nakamura (1990) para o modelo normal.

## 1.2 Exemplo para motivação

Uma amostra de 506 observações de setores censitários de 92 distritos da região metropolitana de Boston foi obtida em 1970. A amostra possui variáveis de atributos como valor das habitações, variáveis de bairro, variáveis de acessibilidade, além de uma variável de poluição do ar, medida pela concentração de óxido de nitrogênio ao quadrado (NOXSQ). Essa amostra completa foi analisada por Harrison e Rubinfeld (1978) e por Belsley *et al.* (1980), mas, como em Zhong *et al.* (2002), vamos considerar dados de apenas 132 setores censitários de 15 distritos da cidade de Boston.

Para ilustrar a estrutura dos dados do censo apresentamos os dados completos na Tabela D.1 no Apêndice D. Uma breve descrição de cada variável é apresentada na Tabela 1.1. Mais detalhes podem ser encontrados em Harrison e Rubinfeld (1978).

Tabela 1.1: Descrição das variáveis utilizadas para analisar os dados dos setores censitários de Boston.

| Símbolo | Descrição |
|---|---|
| LMV | Logaritmo do valor mediano das casas ocupadas pelos proprietários, em USD |
| CRIM | Taxa de criminalidade por cidade |
| ZN | Proporção de terrenos residenciais para os lotes com mais de 25000 pés$^2$ |
| INDUS | Proporção de hectares não comerciais de varejo por cidade |
| CHAS | Variável auxiliar do Rio Charles, com valor 1 se limites das vias sobre o rio e 0 caso contrário |
| NOXSQ | Concentração de óxido de nitrogênio (partes por 100 milhões) ao quadrado |
| ROOM | Número médio de quartos ao quadrado |
| AGE | Proporção de unidades ocupadas pelos proprietários construídas antes de 1940 |
| DIST | Logaritmo das distâncias ponderadas para cinco centros de emprego da região de Boston |
| RAD | Logaritmo do índice de acessibilidade para rodovias radiais |
| TAX | Valor total da taxa de imposto sobre a propriedade (por 10000 USD) |
| PTRATIO | Relação aluno-professor por cidade |
| BLACK | $(Bk - 0,63)^2$, em que $Bk$ é a proporção de negros na população |
| LSTAT | Logaritmo da proporção da população com baixa renda |

Harrison e Rubinfeld (1978) utilizaram os dados de setores censitários de Boston e construíram um modelo hedônico de preços da habitação para medir a disposição de se pagar por ar limpo. Os autores estavam principalmente interessados em examinar o impacto da poluição do ar (medida pelo quadrado da concentração de óxido de nitrogênio (NOXSQ)) no preço das casas ocupadas pelos proprietários e para isso incluiram NOXSQ e outras treze

variáveis como indicadores relevantes para o estudo. O (logaritmo) do valor mediano das casas ocupadas pelos proprietários no setor censitário foi tomado como a variável dependente em um modelo de regressão de efeitos fixos. Uma descrição completa destes dados pode ser encontrada em Harrison e Rubifeld (1978) e Belsley *et al.* (1980).

O modelo utilizado por Harrison e Rubinfeld (1978), apresentado e discutido por Belsley *et al.* (1980) é um modelo de regressão linear múltipla que, em termos das variáveis apresentadas na Tabela 1.1, fica dado por

$$LMV_{ij} = \beta_1 + \beta_2 CRIM_{ij} + \cdots + \beta_{13} BLACK_{ij} + \beta_{14} LSTAT_{ij} + \boldsymbol{\epsilon}_{ij},$$

para $i = 1, ..., 92$ e $j = 1, ..., m_i$. Esse modelo pode ser reescrito como

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} + ... + \beta_{13} x_{13ij} + \beta_{14} x_{14ij} + \epsilon_{ij},$$

em que $i = 1, ..., 92$ e $j = 1, ..., m_i$. Em forma matricial, o modelo fica dado por

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i,$$

$i = 1, ..., 92$, com $\boldsymbol{X}_i$ sendo uma matriz $m_i \times p$, $\boldsymbol{\beta}$ é um vetor $p \times 1$ e $\boldsymbol{\epsilon}_i$ um vetor $m_i \times 1$, assumindo $\boldsymbol{\epsilon}_i \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{m_i})$ mutuamente independentes.

Zhong *et al.* (2002) utilizaram o mesmo modelo e conjunto de dados que Harrison e Rubinfeld (1978), porém selecionaram dados de apenas 132 setores censitários de 15 distritos da cidade de Boston. Além disso, os setores censitários dos distritos são tomados como medidas repetidas e, por isso, os autores ajustaram um modelo linear de efeitos mistos.

Neste conjunto de dados, todas a variáveis independentes podem ser medidas precisamente, com excessão da variável que mede a poluição (NOXSQ), a qual é considerada com erros de medição. O modelo misto linear de Zhong *et al.* (2002) para esta aplicação fica dado por

$$\begin{aligned} LMV_{ij} &= \beta_1 + \beta_2 ROOM_{ij} + \beta_3 AGE_{ij} + \beta_4 DIST_{ij} + \beta_5 BLACK_{ij} \\ &\quad \beta_6 LSTAT_{ij} + \beta_7 CRIM_{ij} + \beta_8 CHAS_{ij} + \beta_9 NOXSQ_{ij} + b_i \mathbf{1}_{m_i} + \epsilon_{ij}, \quad (1.1) \end{aligned}$$

em que $i = 1, ..., 15$ e $j = 1, ..., m_i$ e $\mathbf{1}_{m_i}$ denota um vetor de uns $m_i \times 1$, assumindo $\boldsymbol{\epsilon}_i \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{m_i})$ e $b_i \sim \boldsymbol{N}(0, \sigma^2)$, sendo esses erros mutuamente independentes.

Podemos observar que tanto em Belsley *et al.* (1980) quanto em Zhong *et al.* (2002) foram considerados erros com distribuição normal. Em nosso caso, utilizaremos o mesmo subconjunto de dados analisado por Zhong *et al.* (2002), mas a abordagem será uma extensão

do modelo ajustado por esses autores, ajustando um modelo misto linear, com uma variável explicativa sujeita a erros de medição e supondo uma distribuição elíptica tanto para os efeitos aleatórios quanto para os erros aleatórios. Para efeito de comparação, utilizaremos as mesmas variáveis do modelo (1.1).

### 1.2.1 Análise descritiva

Na Tabela 1.2 são apresentadas algumas estatísticas descritivas para a variável resposta LMV, que é o logaritmo do valor mediano das casas ocupadas pelos proprietários, em USD, para os 15 grupos formados na cidade de Boston. Podemos observar que as médias de LMV nos distritos 76 ($\bar{x} = 10,316$), 77 ($\bar{x} = 10,820$) e 88 ($\bar{x} = 10,058$) são as mais elevadas. Pelo coeficiente de variação (CV) vemos que não há grande variabilidade entre os distritos, visto que o maior valor foi $CV = 3,87\%$ (para o distrito 76).

Tabela 1.2: Estatísticas descritivas para a variável resposta LMV nos 15 distritos.

| Setor | $m_i$ | $\bar{x}$ | DP | CV (%) | Min | Q1 | Md | Q3 | Max |
|---|---|---|---|---|---|---|---|---|---|
| 75 | 8 | 9,941 | 0,132 | 1,330 | 9,729 | 9,815 | 9,964 | 10,029 | 10,127 |
| 76 | 6 | 10,316 | 0,399 | 3,870 | 9,994 | 9,994 | 10,135 | 10,820 | 10,820 |
| 77 | 3 | 10,820 | 0,000 | 0,000 | 10,820 | 10,820 | 10,820 | 10,820 | 10,820 |
| 78 | 2 | 9,532 | 0,000 | 0,000 | 9,532 | 9,532 | 9,532 | 9,532 | 9,53274 |
| 79 | 7 | 9,415 | 0,154 | 1,640 | 9,230 | 9,249 | 9,480 | 9,539 | 9,616 |
| 80 | 11 | 9,302 | 0,329 | 3,54 | 8,882 | 9,082 | 9,259 | 9,417 | 10,052 |
| 81 | 13 | 9,055 | 0,378 | 4,170 | 8,517 | 8,689 | 9,048 | 9,441 | 9,532 |
| 82 | 8 | 9,807 | 0,290 | 2,960 | 9,384 | 9,637 | 9,753 | 10,115 | 10,236 |
| 83 | 19 | 9,314 | 0,310 | 3,330 | 8,854 | 9,036 | 9,296 | 9,561 | 9,943 |
| 84 | 23 | 9,485 | 0,213 | 2,240 | 9,036 | 9,367 | 9,503 | 9,629 | 9,820 |
| 85 | 11 | 9,770 | 0,177 | 1,810 | 9,449 | 9,609 | 9,852 | 9,903 | 9,971 |
| 86 | 6 | 9,909 | 0,073 | 0,730 | 9,857 | 9,857 | 9,891 | 9,944 | 10,052 |
| 87 | 7 | 9,715 | 0,319 | 3,290 | 9,393 | 9,496 | 9,589 | 9,971 | 10,302 |
| 88 | 4 | 10,058 | 0,057 | 0,570 | 9,990 | 10,003 | 10,058 | 10,113 | 10,127 |
| 89 | 4 | 9,921 | 0,045 | 0,450 | 9,857 | 9,876 | 9,933 | 9,955 | 9,962 |

$m_i$: nº de observações do grupo $i$; $\bar{x}$: média; DP: desvio padrão; CV: coeficiente de variação; Q1: 1º quartil; Md: mediana; Q3: 3º quartil; Min: mínimo; Max: máximo

Nas Figuras 1.1-1.4, são apresentados os gráficos *box-plot* para cada variável, segundo cada distrito. Por meio destes gráficos podemos observar que, para todas as variáveis em estudo, há uma grande variação do valor mediano em cada um dos distritos. Além disso, há a presença de diversos pontos discrepantes que precisam ser analisados com maior cuidado. A presença de pontos discrepantes deve ser levada em conta especialmente no ajustes de

modelos, no sentido de buscar distribuições que acomodem estes pontos para que não tenham uma influência desproporcional nos resultados inferenciais do estudo.

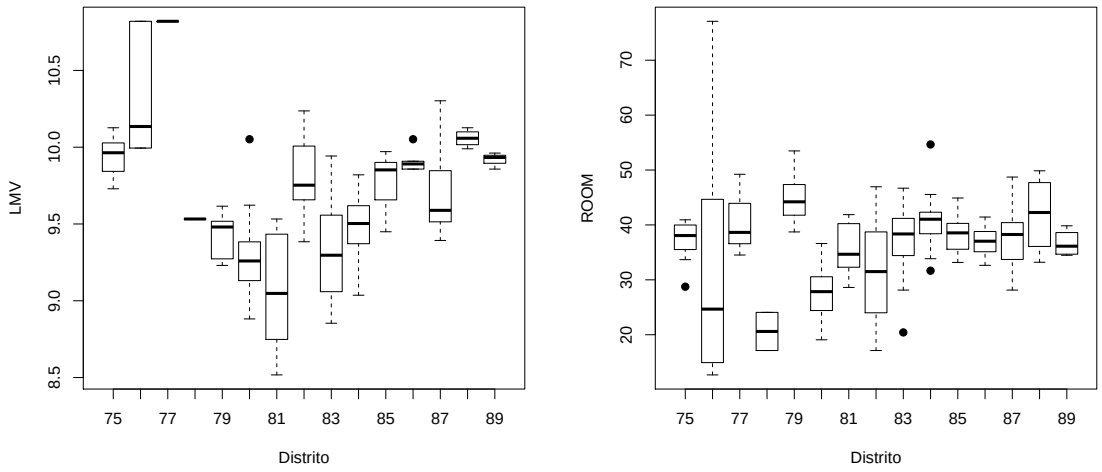


Figura 1.1: Gráficos *box-plot* das variáveis LMV e ROOM, segundo cada distrito.

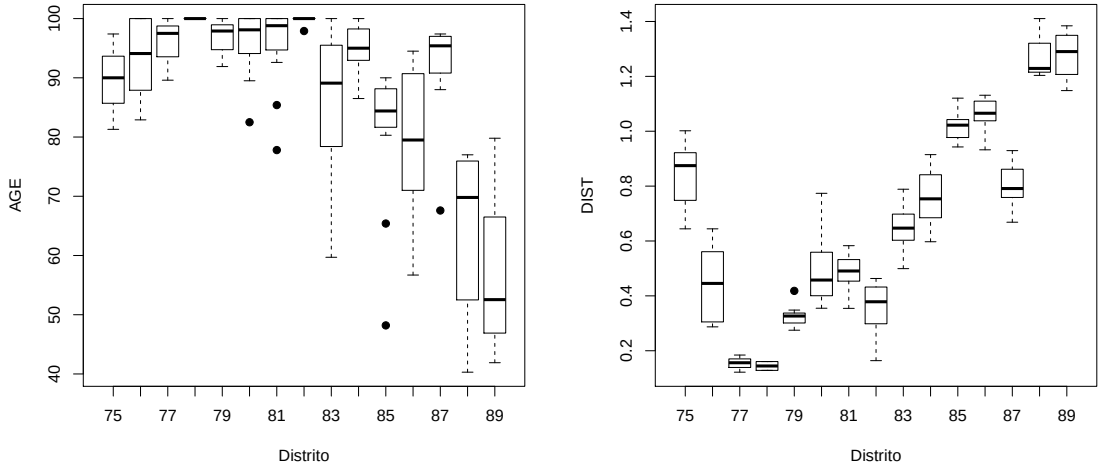


Figura 1.2: Gráficos *box-plot* das variáveis AGE e DIST, segundo cada distrito.

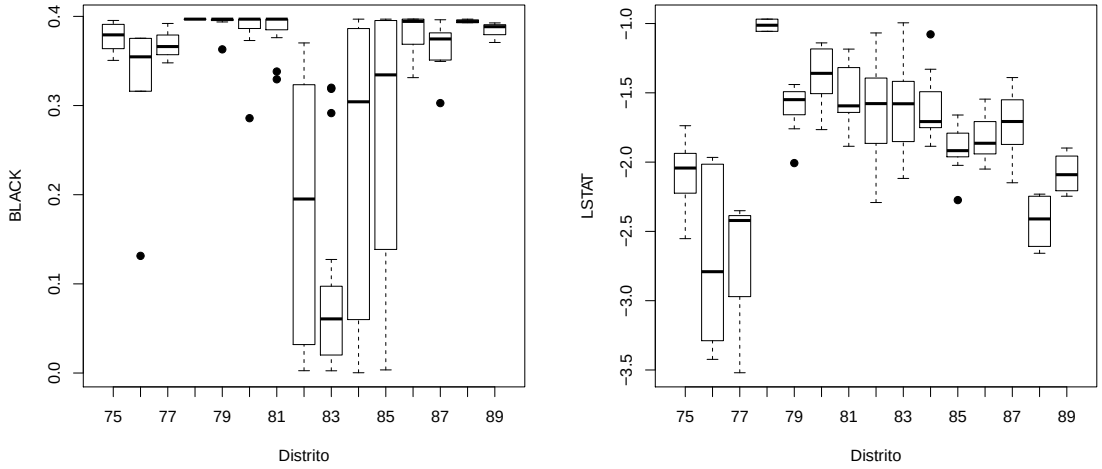


Figura 1.3: Gráficos *box-plot* das variáveis BLACK e LSTAT, segundo cada distrito.

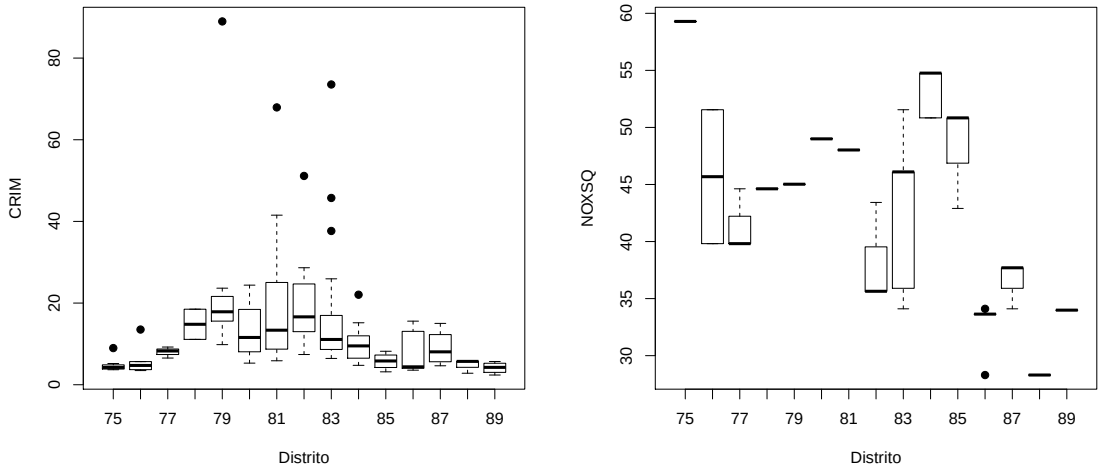


Figura 1.4: Gráficos *box-plot* das variáveis CRIM e NOXSQ, segundo cada distrito.

Nas Figuras 1.5 e 1.6 apresentamos os diagramas de dispersão entre a variável resposta LMV e cada uma das covariáveis. Apesar de alguns pontos discrepantes, a dispersão das variáveis ROOM, DIST, LSTAT, CRIM e NOXSQ *versus* LMV mostra uma nuvem de pontos que indica uma dependência linear entre as variáveis. Já para a dispersão das variáveis AGE e BLACK *versus* LMV, a nuvem de pontos não indica claramente uma dependência linear entre as variáveis.

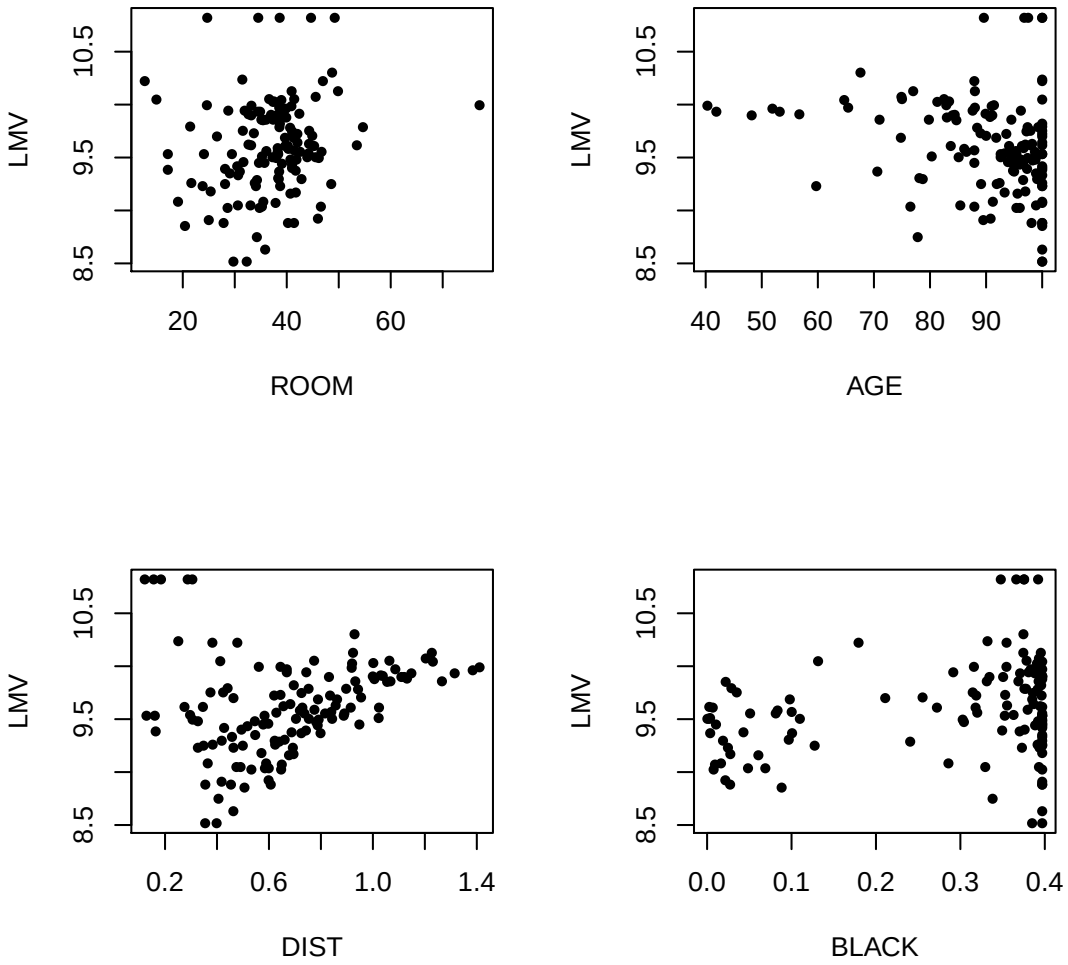


Figura 1.5: Diagramas de dispersão entre a variável resposta LMV e as covariáveis ROOM, AGE, DIST e BLACK.

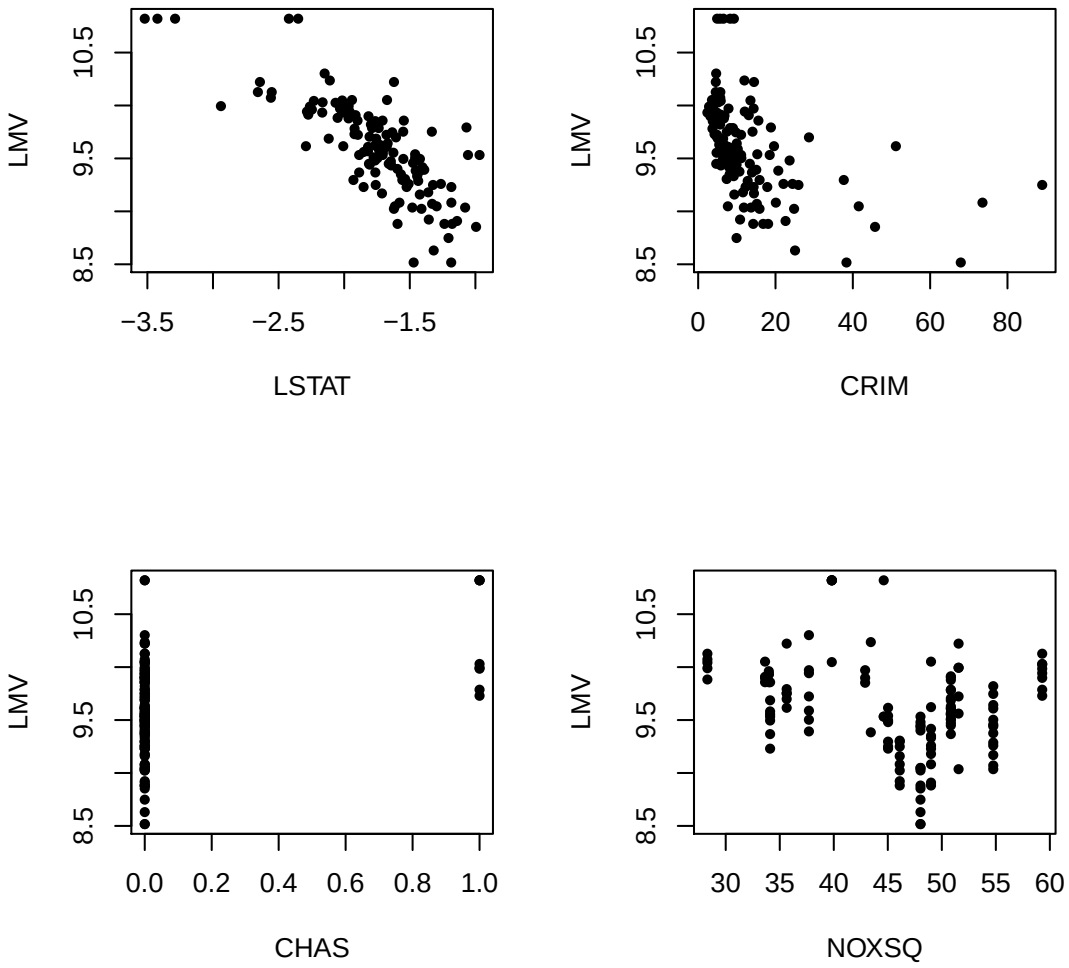


Figura 1.6: Diagramas de dispersão entre a variável resposta LMV e as covariáveis LSTAT, CRIM, CHAS e NOXSQ.

## 1.3 Proposta da tese e objetivos

A proposta deste trabalho é estender o modelo proposto por Savalli *et al.* (2006) no sentido de incluir um componente aleatório para as variáveis sujeitas a erro. Nesta proposta pretendemos estudar e desenvolver procedimentos para análises em modelos mistos lineares elípticos com erros de medição, como estimação e inferência robustas, no sentido de utilizar distribuições que acomodem pontos aberrantes de forma mais eficiente do que a distribuição normal. Além disso, pretendemos realizar um estudo de sensibilidade das estimativas por meio de métodos de diagnóstico de influência local, utilizando a metodologia proposta por Zhu *et al.* (2007) para seleção de perturbações apropriadas. Após o desenvolvimento da

teoria faremos uma reanálise dos dados dos setores censitários da cidade de Boston.

Nossos objetivos específicos para este trabalho de tese são os seguintes:

i) desenvolver a estimação por máxima verossimilhança em um modelo de regressão linear com efeitos mistos com erros em uma variável explicativa, na abordagem estrutural;

ii) aplicar o método de influência local em que o esquema de perturbação é escolhido de forma apropriada e

iii) fazer a aplicação da teoria desenvolvida utilizando as distribuições normal e $t$ de Student ao conjunto de dados dos setores censitários da cidade de Boston.

## 1.4 Organização do trabalho

Este trabalho de tese está organizado em seis capítulos, cuja descrição é apresentada a seguir. No Capítulo 2 revisamos o modelo misto linear normal proposto por Laird e Ware (1982) e apresentamos a proposta de inclusão de uma variável explicativa ou covariável contínua sujeita a erros de medição. Para o modelo proposto, discutimos e apresentamos as condições de identificabilidade do modelo, apresentamos as expressões das funções escore, a matriz de informação de Fisher, discutimos o processo de estimação de máxima verossimilhança, a predição dos efeitos aleatórios, além de apresentar uma proposta de verificação da qualidade do ajuste por meio de distâncias de Mahalanobis transformadas.

No Capítulo 3 descrevemos o modelo misto linear elíptico proposto por Savalli *et al.* (2006) e propomos a inclusão de uma covariável contínua sujeita a erros de medição. A partir disso, estendemos os resultados obtidos para o modelo normal, obtidos no Capítulo 2.

No Capítulo 4 desenvolvemos métodos de diagnóstico baseados na influência local considerando as perturbações: ponderação de casos, na matriz de escala e nas respostas observadas. Para isso, utilizamos a metodologia proposta por Zhu *et al.* (2007) para seleção de perturbações apropriadas segundo o modelo proposto.

No Capítulo 5 utilizamos os dados dos setores censitários da cidade de Boston para ilustrar a aplicação dos resultados inferenciais e de diagnóstico, particularizando os resultados obtidos e apresentados nos capítulos anteriores para o modelo misto linear normal e $t$-Student multivariados.

E, finalmente, apresentamos no Capítulo 6 uma discussão sobre os principais resultados obtidos, as principais conclusões e trabalhos futuros.

## 1.5 Aspectos preliminares das distribuição elípticas

Nos últimos anos várias abordagens surgiram como alternativas à modelagem com erros normais, entre elas, a utilização de distribuições simétricas ou elípticas. Várias dessas abordagens encontram-se em Fang *et al.* (1990) e Fang e Anderson (1990).

A classe das distribuições elípticas reúne distribuições com caudas mais leves e mais pesadas do que a normal, mas que preservam a estrutura simétrica da distribuição normal, que é um caso particular desta classe. Outras distribuições muito usadas e que são também casos particulares das elípticas são a distribuição $t$ de Student, a exponencial potência, a logística e a normal contaminada, entre outras.

Com o objetivo de introduzir o modelo misto elíptico com erros nas variáveis é apresentada nesta seção a classe de distribuições elípticas com algumas definições e propriedades úteis para o desenvolvimento deste trabalho. Um estudo mais detalhado das propriedades das distribuições elípticas pode ser encontrado, por exemplo, em Fang *et al.* (1990) e Arellano-Valle (1994).

### 1.5.1 Distribuição elíptica multivariada

**Definição 1** *Diz-se que o vetor aleatório* $\boldsymbol{y} \in \Re^n$ $(n \geq 2)$ *segue uma distribuição elíptica se sua função característica assume a forma*

$$\psi_{\boldsymbol{y}}(\boldsymbol{t}) = (\exp)\{\mathrm{i}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\mu}\}\mathrm{g}(\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}), \tag{1.2}$$

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

Se $\boldsymbol{y}$ tem distribuição elíptica com função característica dada por (1.2), usamos a notação $\boldsymbol{y} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, ou simplesmente $\boldsymbol{y} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

**Propriedade 1** *Suponha que* $\boldsymbol{y} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, *com* $\mathrm{posto}(\boldsymbol{\Sigma}) < n$. *Se* $\boldsymbol{B}$ *é uma matriz* $(n \times m)$ *e* $\boldsymbol{\vartheta}$ *é um vetor* $(m \times 1)$, *então*

$$\boldsymbol{\vartheta} + \boldsymbol{B}^T\boldsymbol{y} \sim El_m(\boldsymbol{\vartheta} + \boldsymbol{B}^T\boldsymbol{\mu}, \boldsymbol{B}^T\boldsymbol{\Sigma}\boldsymbol{B}, g).$$

*Em particular, se considerarmos as partições*

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

*obtém-se as distribuições marginais:*

*a)* $\boldsymbol{y}^{(1)} \sim El_m(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}, g)$ *e*

*b)* $\boldsymbol{y}^{(2)} \sim El_{(n-m)}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}, g)$.

Isso quer dizer que uma transformação linear de um vetor aleatório com distribuição elíptica segue também uma distribuição elíptica e que cada elemento do vetor aleatório $\boldsymbol{y}$ tem uma distribuição marginal elíptica.

**Propriedade 2** *Suponha que* $\boldsymbol{y} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ *com* $\boldsymbol{\Sigma} \geq 0$. *Considerando a partição apresentada na Propriedade 1, temos que*

$$\left(\boldsymbol{y}^{(1)} \middle| \boldsymbol{y}_0^{(2)}\right) \sim El_m\left(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}, g_{q(y_0^{(2)})}\right),$$

*em que*

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_0^{(2)} - \boldsymbol{\mu}^{(2)}),$$

$$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \text{ e}$$

$$q(\boldsymbol{y}_0^{(2)}) = (\boldsymbol{y}_0^{(2)} - \boldsymbol{\mu}^{(2)})^T\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_0^{(2)} - \boldsymbol{\mu}^{(2)}).$$

*Analogamente, temos que*

$$\left(\boldsymbol{y}^{(2)} \middle| \boldsymbol{y}_0^{(1)}\right) \sim El_m\left(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}, g_{q(y_0^{(1)})}\right),$$

*em que*

$$\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}^{(2)} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_0^{(1)} - \boldsymbol{\mu}^{(1)}),$$

$$\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \text{ e}$$

$$q(\boldsymbol{y}_0^{(1)}) = (\boldsymbol{y}_0^{(1)} - \boldsymbol{\mu}^{(1)})^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_0^{(1)} - \boldsymbol{\mu}^{(1)}).$$

Isso quer dizer que distribuições condicionais do vetor aleatório $\boldsymbol{y}$ dados valores de subvetores de $\boldsymbol{y}$ são também elípticos.

**Definição 2** *Assumindo que* $\text{posto}(\boldsymbol{\Sigma}) = n$, *dizemos que o vetor aleatório* $\boldsymbol{y}$ *tem distribuição*

*elíptica multivariada com função densidade da forma*

$$f_{\boldsymbol{y}}(\boldsymbol{y}) = |\boldsymbol{\Sigma}|^{-1/2} g(\delta), \tag{1.3}$$

*em que* $\delta = (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$ *e* $g(\cdot)$ *é uma função escalar contínua e diferenciável de* $\Re \to [0, \infty)$ *tal que*

$$\int_0^\infty \delta^{-1/2} g(\delta) d\delta < \infty.$$

*A função* $g(\cdot)$ *é conhecida como função geradora de densidades.*

Na Tabela 1.3 apresentamos algumas distribuições que pertencem à classe das distribuições elípticas.

Tabela 1.3: Exemplos de distribuições pertencentes à classe das elípticas.

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

$c$ é uma constante de normalização.

Assumindo que a função geradora de densidades $g(\cdot)$, definida em (1.3), é contínua e diferenciável, podemos definir as quantidades a seguir:

$$W_g(\delta) = \frac{d}{d\delta} \log g(\delta) = \frac{g'(\delta)}{g(\delta)} \quad \text{e} \quad W_g'(\delta) = \frac{d}{d\delta} W_g(\delta).$$

Exemplos de $W_g(\delta)$ e $W_g'(\delta)$ para algumas distribuições elípticas multivariadas são apresentados a seguir.

- *t de Student* com graus de liberdade $\nu > 0$

$$W_g(\delta) = -\frac{1}{2}\left(\frac{\nu + m}{\nu + \delta_i}\right) \quad \text{e} \quad W_g'(\delta) = \frac{1}{2}\frac{\nu + m}{(\nu + \delta_i)^2}.$$

- *Exponencial potência* com parâmetro de forma $\lambda > 0$

$$W_g(\delta) = -\frac{1}{2}\lambda\delta_i^{\lambda-1} \quad \text{e} \quad W_g'(\delta) = -\frac{1}{2}\lambda(\lambda-1)u^{\lambda-2}$$

  com $\delta_i \neq 0$ e $\lambda \neq \frac{1}{2}$.

- *Normal contaminada* com $0 < \alpha < 1$ e $0 \leq \kappa < 1$

$$W_g(\delta) = -\frac{1}{2}\frac{1-\alpha+\alpha\gamma^{m/2+1}e^{(1-\kappa)\delta_i/2}}{1-\alpha+\alpha\kappa^{m/2}e^{(1-\kappa)\delta_i/2}}$$

  e

$$W_g'(\delta) = -\frac{1}{2}\frac{\alpha\kappa^{m/2}(1-\kappa)[W_g(\delta_i)+\kappa/2]e^{(1-\kappa)u/2}}{1-\alpha+\alpha\kappa^{m/2}e^{(1-\kappa)\delta_i/2}},$$

  em que $\alpha$ representa a porcentagem de pontos discrepantes enquanto $\kappa$ pode ser interpretado como um fator de escala.

# Capítulo 2

# Modelo Misto Linear Normal com Erros de Medição

## 2.1 Introdução

Neste capítulo vamos apresentar um resumo dos modelos mistos sob a abordagem proposta por Laird e Ware (1982) e introduzir a proposta de inclusão de uma covariável contínua sujeita a erros de medição, considerando uma abordagem com a distribuição conjunta da resposta e da variável observada sujeita a erros de medição. Para este modelo proposto, discutimos as condições de identificabilidade, apresentamos as expressões das funções escore, a matriz de informação de Fisher, a estimação por máxima verossimilhança, além da predição dos efeitos aleatórios $\boldsymbol{b}_i$ e da verdadeira variável, chamada de $\boldsymbol{u}_i^*$. Outra discussão importante apresentada é acerca de testes de hipóteses para a inclusão da variável medida com erros, por meio de quatro testes conhecidos: teste da razão de verossimilhanças, teste de Wald, teste do escore de Rao (1948) e teste de Neyman (1959). Finalmente, apresentamos uma proposta de verificação da qualidade do ajuste por meio da distância de Mahalanobis transformada (Johnson *et al.*, 1994).

## 2.2 Modelo misto linear normal

Dados longitudinais (ou dados de medidas repetidas) são muito comuns na prática, obtidos, por exemplo, com estudos experimentais. Em um estudo longitudinal, indivíduos são acompanhados por um período de tempo e, para cada indivíduo, dados são coletados em diversos espaços de tempo. Assim, a definição característica de estudo longitudinal é que múltiplas ou repetidas medições de uma mesma variável são efetuadas para cada indivíduo no estudo, sobre um período de tempo (Wu, 2010).

Um dos objetivos principais de análises estatísticas é estudar as variações dos dados. Para dados longitudinais existem duas fontes de variação: 1) variação intraindivíduo, isto é, a variação nas medições repetidas em cada indivíduo; 2) variação entre indivíduos, isto é, a variação dos dados entre os diferentes indivíduos. Assim, para dados com essa estrutura, o modelo de regressão linear clássico é inapropriado uma vez que as observações em cada indivíduo podem ser correlacionadas e a suposição de independência não é válida. Para incorporar a correlação intraindivíduos e a variação entre indivíduos, podemos estender o modelo linear clássico introduzindo efeitos aleatórios e assim obter o modelo misto linear (MML). Em modelos com efeitos mistos paramétricos, geralmente assume-se que os efeitos aleatórios seguem uma distribuição normal multivariada e assume-se que os erros aleatórios intraindivíduos (grupo) seguem distribuições paramétricas na família exponencial.

O modelo misto linear normal para respostas contínuas proposto por Laird e Ware (1982) assume a seguinte forma:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{2.1}$$

$$\boldsymbol{b}_i \overset{\text{iid}}{\sim} \boldsymbol{N}_q(\boldsymbol{0}, \boldsymbol{D}) \quad \text{e} \quad \boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} \boldsymbol{N}_{m_i}(\boldsymbol{0}, \boldsymbol{R}_i), \quad i = 1, ..., n, \tag{2.2}$$

em que $\boldsymbol{y}_i$ representa o vetor aleatório $m_i$-dimensional das respostas observadas para o $i$-ésimo indivíduo ou grupo, $\boldsymbol{X}_i$ e $\boldsymbol{Z}_i$ são matrizes de planejamento $(m_i \times p)$ e $(m_i \times q)$, respectivamente, $\boldsymbol{\beta}$ é um vetor $p$-dimensional de efeitos fixos, $\boldsymbol{b}_i$ denota um vetor $q$-dimensional de efeitos aleatórios e $\boldsymbol{\epsilon}_i$ representa um vetor de erros. Em geral, assume-se que os efeitos aleatórios $\boldsymbol{b}_i$ e os erros $\boldsymbol{\epsilon}_i$ são independentes e que $\boldsymbol{D}$ e $\boldsymbol{R}_i$ são matrizes de variância-covariância de ordens $(q \times q)$ e $(m_i \times m_i)$, positivas definidas, que correspondem, respectivamente, às variabilidades entre e intraunidades amostrais. Note que o MML (2.1) difere do modelo de regressão linear clássico somente pelo termo $\boldsymbol{Z}_i\boldsymbol{b}_i$, que liga os efeitos aleatórios à resposta.

Alternativamente, podemos escrever o modelo usando a distribuição conjunta de $(\boldsymbol{y}_i^T, \boldsymbol{b}_i^T)^T$ que é dada por

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{pmatrix} \overset{\text{ind}}{\sim} \boldsymbol{N}_{m_i+q}\left[\begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i & \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{Z}_i^T & \boldsymbol{D} \end{pmatrix}\right], \tag{2.3}$$

e a inferência clássica é usualmente baseada na função de verossimilhança do modelo marginal $\boldsymbol{y}_i \sim \boldsymbol{N}_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i)$.

Geralmente, assume-se que a matriz de variâncias-covariâncias $\boldsymbol{R}_i$ depende de $i$ somente através de suas dimensões. Por exemplo, é comum assumir que $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}_{m_i}$, com $\boldsymbol{I}_{m_i}$ sendo a matriz identidade de ordem $m_i$. Isso sugere que as medições intraindivíduos são frequen-

temente assumidas ser condicionalmente independentes dados os efeitos aleatórios. Esta suposição pode ser razoável quando as medições intraindivíduos são distantes de modo que a correlação intraindivíduos seja praticamente desprezível, ou que a variação entre indivíduos seja dominante. Em muitos casos, uma caracterização precisa de $\boldsymbol{R}_i$ é menos exigente. Davidian e Giltinan (2003) forneceram uma discussão detalhada sobre a especificação da matriz $\boldsymbol{R}_i$.

A distribuição marginal da resposta $\boldsymbol{y}_i$ é dada por

$$\boldsymbol{y}_i \overset{\text{ind}}{\sim} \boldsymbol{N}_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i).$$

Assim, a estrutura de variância-covariância das observações repetidas no indivíduo $i$ fica dada por

$$Var(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i.$$

Os efeitos aleatórios $\boldsymbol{b}_i's$ podem ser preditos através do método de Bayes empírico (Verbeke e Molemberghs, 2000), dado por

$$\widehat{\boldsymbol{b}}_i = E(\widehat{\boldsymbol{b}_i|\boldsymbol{y}_i}) = \widehat{\boldsymbol{D}}\boldsymbol{Z}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}).$$

A média marginal $E(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$ pode ser interpretada como uma média sobre todos os efeitos aleatórios, por isso não reflete trajetórias individuais longitudinais. Ao invés disso, a inferência individual é realizada condicionando sobre os efeitos aleatórios $\boldsymbol{b}_i's$. A inferência para os parâmetros populacionais $\boldsymbol{\beta}$ é baseada na distribuição marginal apresentada.

Note que a capacidade para obter a distribuição marginal de $\boldsymbol{y}_i$ em forma fechada depende das seguintes suposições: os efeitos aleatórios $\boldsymbol{b}_i's$ e os erros aleatórios $\boldsymbol{\epsilon}_i's$ sejam lineares no modelo (2.1)-(2.2) e que $\boldsymbol{b}_i$ e $\boldsymbol{\epsilon}_i$ sejam independentes e normalmente distribuídos.

A identificabilidade dos parâmetros ou do modelo é um problema importante em modelos com efeitos mistos. Diz-se que parâmetros ou modelos são ditos inidentificáveis se dois conjuntos diferentes de parâmetros levam à mesma distribuição de probabilidade. Demidenko (2004) e Wang e Heckman (2009) discutiram a identificabilidade dos parâmetros para MMLs. Em particular, Wang e Heckman (2009) mostraram que o modelo (2.1)-(2.2) é sempre identificável quando $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}_{m_i}$.

Uma deficiência que pode ocorrer com o modelo (2.1)-(2.2) é a sensibilidade das estimativas de máxima verossimilhança a pontos aberrantes. Uma maneira de amenizar esse problema é através da aplicação de métodos robustos, conforme descrito por Copt e Victoria-Ferrer (2006). Todavia, em algumas situações pode haver indícios (por meio de análises de resíduos) de que os erros apresentam caudas mais leves ou mais pesadas do que os erros

normais. Nesses casos, distribuições para os erros que flexibilizem a curtose podem ser assumidos e, em particular, no caso de erros com distribuições elípticas (Fang *et al.*, 1990) com caudas pesadas, as estimativas de máxima verossimilhança são robustas contra observações aberrantes. Outra alternativa é a aplicação de métodos robustos em modelos com erros elípticos com caudas mais pesadas do que os erros normais. Não há, contudo, uma garantia de que esses métodos levem a uma proteção maior contra observações extremas do que o método de máxima verossimilhança, além do custo computacional ser mais expendioso.

### 2.2.1 Inclusão de uma covariável medida com erros

Vamos supor agora que uma variável explicativa ou covariável contínua sujeita a erros de medição é incluída no modelo (2.1)-(2.2) da seguinte forma:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \gamma\boldsymbol{u}_i^* + \boldsymbol{\epsilon}_i, \tag{2.4}$$

$$\boldsymbol{b}_i \overset{\text{iid}}{\sim} \boldsymbol{N}_q(\boldsymbol{0}, \boldsymbol{D}) \text{ e} \tag{2.5}$$

$$\boldsymbol{u}_i = \boldsymbol{u}_i^* + \boldsymbol{e}_i,\ i = 1, ..., n, \tag{2.6}$$

em que
$\boldsymbol{y}_i = (y_{i1}, ..., y_{im_i})^T$;
$\boldsymbol{X}_i$ é uma matriz $(m_i \times p)$ que contém valores de variáveis explicativas ou covariáveis;
$\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ é um vetor $p$-dimensional de efeitos fixos;
$\boldsymbol{Z}_i$ é uma matriz de planejamento $(m_i \times q)$ de efeitos aleatórios $\boldsymbol{b}_i$;
$\boldsymbol{b}_i = (b_{i1}, ..., b_{iq})^T$ denota um vetor $q$-dimensional de efeitos aleatórios;
$\boldsymbol{\epsilon}_i$ denota um vetor de erros aleatórios $m_i$-dimensional;
$\boldsymbol{u}_i$, $i = 1, ..., n$, é uma variável observada, com medidas repetidas e sujeita a erros de medição, da variável verdadeira $\boldsymbol{u}_i^*$, não observável; e
$\boldsymbol{e}_i$ é um vetor de erros de medição $(m_i \times 1)$.

Este modelo fica, alternativamente, dado por

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \\ \boldsymbol{u}_i \\ \boldsymbol{u}_i^* \end{pmatrix} \overset{\text{ind}}{\sim} \boldsymbol{N}_{3m_i+q} \left[ \begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} + \gamma\boldsymbol{\mu}_i^* \\ \boldsymbol{0} \\ \boldsymbol{\mu}_i^* \\ \boldsymbol{\mu}_i^* \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_i & \boldsymbol{Z}_i\boldsymbol{D} & \gamma\boldsymbol{\Sigma}_u & \gamma\boldsymbol{\Sigma}_u \\ \boldsymbol{D}\boldsymbol{Z}_i^T & \boldsymbol{D} & \boldsymbol{0} & \boldsymbol{0} \\ \gamma\boldsymbol{\Sigma}_u & \boldsymbol{0} & \boldsymbol{\Sigma}_u + \sigma_e^2\boldsymbol{I}_{m_i} & \boldsymbol{\Sigma}_u \\ \gamma\boldsymbol{\Sigma}_u & \boldsymbol{0} & \boldsymbol{\Sigma}_u & \boldsymbol{\Sigma}_u \end{pmatrix} \right], \tag{2.7}$$

com $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \gamma^2\boldsymbol{\Sigma}_u + \sigma^2\boldsymbol{I}_{m_i}$, i=1, ..., n.

Assim, sob uma abordagem marginal temos que a distribuição dos dados observados $\boldsymbol{W}_i = (\boldsymbol{y}_i^T, \boldsymbol{u}_i^T)^T$ fica dada por

$$\boldsymbol{W}_i \overset{\text{ind}}{\sim} \boldsymbol{N}_{2m_i}\left[\begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} + \gamma\boldsymbol{\mu}_i^* \\ \boldsymbol{\mu}_i^* \end{pmatrix}, \begin{pmatrix} \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \gamma^2\boldsymbol{\Sigma}_u + \sigma^2\boldsymbol{I}_{m_i} & \gamma\boldsymbol{\Sigma}_u \\ \gamma\boldsymbol{\Sigma}_u & \boldsymbol{\Sigma}_u + \sigma_e^2\boldsymbol{I}_{m_i} \end{pmatrix}\right]. \quad (2.8)$$

Por simplicidade, vamos assumir que $\boldsymbol{\mu}_i^* = \mu\mathbf{1}_{m_i}$ e $\boldsymbol{\Sigma}_u = \sigma_u^2\boldsymbol{I}_{m_i}$, em que $\mathbf{1}_{m_i}$ é um vetor $(m_i \times 1)$ de $1's$ e $\boldsymbol{I}_{m_i}$ é a matriz identidade de ordem $m_i$. Portanto, a distribuição dos dados observados fica dada por

$$\boldsymbol{W}_i \overset{\text{ind}}{\sim} \boldsymbol{N}_{2m_i}\left(\boldsymbol{\mu}_{iW}, \boldsymbol{V}_i\right), \quad (2.9)$$

em que:

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

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

A identificabilidade em modelos que consideram erros de medição é um assunto muito importante e sua verificação é uma etapa fundamental para a definição do modelo.

**Definição 3** *Seja um modelo estatístico definido por uma família de distribuições para $\boldsymbol{W}$, parametrizado pelo vetor $\boldsymbol{\theta}$, $\{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, em que $\boldsymbol{\Theta}$ é o espaço paramétrico e $P_{\boldsymbol{\theta}}$ denota a distribuição associada a $\boldsymbol{\theta}$. Dizemos que o modelo é identificável em $\boldsymbol{\Theta}$ se $P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2}$ implica $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.*

Segundo Demidenko (2004), para modelos de regressão com distribuição normal a condição

$$E_{\boldsymbol{\theta}_1}(\boldsymbol{W}_i) = E_{\boldsymbol{\theta}_2}(\boldsymbol{W}_i) \text{ e } cov_{\boldsymbol{\theta}_1}(\boldsymbol{W}_i) = cov_{\boldsymbol{\theta}_2}(\boldsymbol{W}_i) \text{ implicam } \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2,$$

é uma condição necessária e suficiente para garantir a identificabilidade, visto que a distribuição normal é especificada unicamente pelos dois primeiros momentos. Mais formalmente temos a seguinte propriedade.

**Propriedade 3** *Considere um modelo definido como*

$$\boldsymbol{W}_i \overset{\text{ind}}{\sim} \boldsymbol{N}_{2m_i}(f(\tilde{\boldsymbol{\beta}}), \boldsymbol{V}_i(\boldsymbol{\theta})),$$

*em que $\boldsymbol{W}_i$ é um vetor $(2m_i \times 1)$ de dados observados, $f(\tilde{\boldsymbol{\beta}})$ é uma função vetor linear $(2m_i \times 1)$ de parâmetros $\tilde{\boldsymbol{\beta}}$ e $\boldsymbol{V}_i(\boldsymbol{\theta})$ é uma matriz $(2m_i \times 2m_i)$ de variâncias-covariâncias que depende do vetor de parâmetros $\boldsymbol{\theta}$. Então, o modelo é identificável se, e somente se, $f(\tilde{\boldsymbol{\beta}}_1) = f(\tilde{\boldsymbol{\beta}}_2)$ e $\boldsymbol{V}_i(\boldsymbol{\theta}_1) = \boldsymbol{V}_i(\boldsymbol{\theta}_2)$ implicam que $\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_2$ e $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.*

Na sequência, aplicamos este resultado para o modelo (2.9) proposto, para verificar as condições para que este modelo seja identificável.

**I) Identificabilidade da média de $\boldsymbol{W}_i$**

De (2.9) temos que

$$\begin{aligned} E(\boldsymbol{W}_i) = \boldsymbol{\mu}_{iW} &= \begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} + \gamma\mu\mathbf{1}_{m_i} \\ \mu\mathbf{1}_{m_i} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{X}_i^*\boldsymbol{\beta}^* \\ \mu\mathbf{1}_{m_i} \end{pmatrix} \\ &= \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}, \end{aligned}$$

em que

$$\boldsymbol{X}_i^* = [\boldsymbol{X}_i, \mathbf{1}_{m_i}],\ \boldsymbol{\beta}^* = \begin{pmatrix} \boldsymbol{\beta} \\ \gamma\mu \end{pmatrix},\ \tilde{\boldsymbol{X}}_i = \begin{bmatrix} \boldsymbol{X}_i & \mathbf{1}_{m_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_{m_i} \end{bmatrix}_{(2m_i\times p+2)} \text{ e } \tilde{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{\beta} \\ \gamma\mu \\ \mu \end{pmatrix}_{(p+2\times 1)}.$$

Logo, se $\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_2$, então $\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_2$, ou seja, se

$$\begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta}_1 + \gamma_1\mu_1\mathbf{1}_{m_i} \\ \mu_1\mathbf{1}_{m_i} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta}_2 + \gamma_2\mu_2\mathbf{1}_{m_i} \\ \mu_2\mathbf{1}_{m_i} \end{pmatrix},$$

então

$$\mu_1 = \mu_2 \text{ e } \boldsymbol{X}_i\boldsymbol{\beta}_1 + \gamma_1\mu_1\mathbf{1}_{m_i} = \boldsymbol{X}_i\boldsymbol{\beta}_2 + \gamma_2\mu_2\mathbf{1}_{m_i}.$$

As condições para que isso seja válido são: *a*) $\boldsymbol{X}_i^*$ ser de posto completo;
*b*) $\mu_1 = 0$. Assim, temos que

$$\begin{aligned} \boldsymbol{X}_i\boldsymbol{\beta}_1 &= \boldsymbol{X}_i\boldsymbol{\beta}_2 \\ \boldsymbol{X}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) &= \mathbf{0} \\ \boldsymbol{\beta}_1 &= \boldsymbol{\beta}_2, \end{aligned}$$

sendo assim, identificável.

**II) Identificabilidade da matriz de variâncias-covariâncias de $\boldsymbol{W}_i$**

Considerando a matriz de variâncias-covariâncias de $\boldsymbol{W}_i$, dada em (2.9), temos que

$$\begin{bmatrix} \boldsymbol{\Sigma}_{i1} & \gamma_1\sigma_{u1}^2\boldsymbol{I}_{m_i} \\ \gamma_1\sigma_{u1}^2\boldsymbol{I}_{m_i} & (\sigma_{u1}^2+\sigma_{e1}^2)\boldsymbol{I}_{m_i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{i2} & \gamma_2\sigma_{u2}^2\boldsymbol{I}_{m_i} \\ \gamma_2\sigma_{u2}^2\boldsymbol{I}_{m_i} & (\sigma_{u2}^2+\sigma_{e2}^2)\boldsymbol{I}_{m_i} \end{bmatrix},$$

em que $\boldsymbol{\Sigma}_{i1} = \boldsymbol{Z}_i\boldsymbol{D}_1\boldsymbol{Z}_i^T + (\gamma_1^2\sigma_{u1}^2+\sigma_1^2)\boldsymbol{I}_{m_i}$, $\boldsymbol{\Sigma}_{i2} = \boldsymbol{Z}_i\boldsymbol{D}_2\boldsymbol{Z}_i^T + (\gamma_2^2\sigma_{u2}^2+\sigma_2^2)\boldsymbol{I}_{m_i}$, $\boldsymbol{D}_1$ e $\boldsymbol{D}_2$ são matrizes de variâncias-covariâncias dos efeitos aleatórios; $\gamma_1$ e $\gamma_2$ são parâmetros fixos da variável sujeita a erros de medição; $\sigma_{u1}^2$ e $\sigma_{u2}^2$ são parâmetros de escala da variável sujeita a erros de medição $\boldsymbol{u}_i^*$; $\sigma_1^2$ e $\sigma_2^2$ são parâmetros de escala dos erros aleatórios; e $\sigma_{e1}^2$ e $\sigma_{e2}^2$ são parâmetros de escala da variável longitudinal observada com erros de medição $\boldsymbol{u}_i$.

Então, temos

*a)* $\gamma_1\sigma_{u1}^2 = \gamma_2\sigma_{u2}^2$;

*b)* $\sigma_{u1}^2+\sigma_{e1}^2 = \sigma_{u2}^2+\sigma_{e2}^2$;

*c)* $\boldsymbol{Z}_i\boldsymbol{D}_1\boldsymbol{Z}_i^T + (\gamma_1^2\sigma_{u1}^2+\sigma_1^2)\boldsymbol{I}_{m_i} = \boldsymbol{Z}_i\boldsymbol{D}_2\boldsymbol{Z}_i^T + (\gamma_2^2\sigma_{u2}^2+\sigma_2^2)\boldsymbol{I}_{m_i}$.

Multiplicando a expressão em (b) por $\gamma_1$ em ambos os lados da igualdade, temos $\gamma_1\sigma_{u1}^2 + \gamma_1\sigma_{e1}^2 = \gamma_1\sigma_{u2}^2 + \gamma_1\sigma_{e2}^2$ e usando (a) temos $\gamma_2\sigma_{u2}^2 + \gamma_1\sigma_{e1}^2 = \gamma_1\sigma_{u2}^2 + \gamma_1\sigma_{e2}^2$. Assim, obtemos $(\gamma_2-\gamma_1)\sigma_{u2}^2 = \gamma_1(\sigma_{e2}^2-\sigma_{e1}^2)$.

Se supusermos $\sigma_{e1}^2$ conhecido, (b) fica provado e, assim, se $\sigma_{u1}^2 = \sigma_{u2}^2$ conclui-se que $\gamma_1 = \gamma_2$ (provando (a)).

Com os resultados anteriores, provar o item (c) reduz-se a provar

*c*)* $\boldsymbol{Z}_i\boldsymbol{D}_1\boldsymbol{Z}_i^T + \sigma_1^2\boldsymbol{I}_{m_i} = \boldsymbol{Z}_i\boldsymbol{D}_2\boldsymbol{Z}_i^T + \sigma_2^2\boldsymbol{I}_{m_i}$.

Agora, considerando $\tilde{\boldsymbol{Z}}_i = \begin{pmatrix} \boldsymbol{Z}_i \\ \boldsymbol{\alpha}_i \end{pmatrix}$ e $\tilde{\boldsymbol{D}} = \begin{bmatrix} \boldsymbol{D} & 0 \\ 0 & \sigma^2 \end{bmatrix}$, com $\boldsymbol{\alpha}_i = (0,0,...,1)^T$ de ordem $(1\times q)$, podemos escrever $\boldsymbol{Z}_i\boldsymbol{D}_1\boldsymbol{Z}_i^T + \sigma_1^2\boldsymbol{I}_{m_i} = \tilde{\boldsymbol{Z}}_i\tilde{\boldsymbol{D}}\tilde{\boldsymbol{Z}}_i^T$.

Portanto, $\tilde{\boldsymbol{Z}}_i\tilde{\boldsymbol{D}}_1\tilde{\boldsymbol{Z}}_i^T = \tilde{\boldsymbol{Z}}_i\tilde{\boldsymbol{D}}_2\tilde{\boldsymbol{Z}}_i^T$, implica $\tilde{\boldsymbol{D}}_1 = \tilde{\boldsymbol{D}}_2$, desde que $\tilde{\boldsymbol{Z}}_i$ seja de posto completo.

A partir dessas análises, para garantir que o modelo (2.9) seja identificável vamos assumir que a matriz de planejamento $\boldsymbol{X}_i^*$ é de posto completo. Para isso, temos que supor que $\beta_0$ seja conhecido. Além disso, supomos que o parâmetro de escala associado com os erros de medição, $\sigma_e^2$, é conhecido. Além desta última condição, existem outras condições usuais: supor que a variância da verdadeira variável, $\sigma_u^2$, é conhecida; supor que a variância dos erros aleatórios, $\sigma_\epsilon^2$, é conhecida; supor que duas destas variâncias sejam conhecidas; ou ainda, supor que a razão de variâncias $\sigma_\epsilon^2/\sigma_e^2$ é conhecida.

Após as considerações feitas acima e para efeito de estimação, o vetor de parâmetros a ser considerados é $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, vech(\boldsymbol{D})\right)^T$.

## 2.3 Função escore

Considerando a distribuição de $\boldsymbol{W}_i = (\boldsymbol{y}_i^T, \boldsymbol{u}_i^T)^T$ em (2.9) e seus valores observados, temos que a função de densidade normal fica dada por

$$f(\boldsymbol{W}_i; \boldsymbol{\theta}) = (2\pi)^{-\frac{2m_i}{2}} |\boldsymbol{V}_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^T \boldsymbol{V}_i^{-1} (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})\right],$$

para $i = 1, ..., n$, em que $|\boldsymbol{V}_i|$ denota o determinante da matriz $\boldsymbol{V}_i$. A contribuição do $i$-ésimo grupo na função de verossimilhança é dada por

$$l_i(\boldsymbol{\theta}) = f(\boldsymbol{W}_i; \boldsymbol{\theta}). \tag{2.10}$$

Portanto, a função de verossimilhança é dada pelo produto das $n$ contribuições definidas em (2.10), ou seja,

$$\begin{aligned} l(\boldsymbol{\theta}) &= \prod_{i=1}^{n} l_i(\boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} (2\pi)^{-m_i} |\boldsymbol{V}_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^T \boldsymbol{V}_i^{-1} (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})\right]. \end{aligned}$$

Assim, o logaritmo da função de verossimilhança para o vetor $\boldsymbol{\theta}$, para o caso normal, é dado por

$$\begin{aligned} L(\boldsymbol{\theta}) &= \log l(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{n} L_i(\boldsymbol{\theta}), \end{aligned} \tag{2.11}$$

em que $L_i(\boldsymbol{\theta}) = -m_i \log(2\pi) - \frac{1}{2}\log|\boldsymbol{V}_i| - \frac{1}{2}(\boldsymbol{r}_i)^T \boldsymbol{V}_i^{-1} \boldsymbol{r}_i$, com $\boldsymbol{r}_i = \boldsymbol{W}_i - \boldsymbol{\mu}_{iW}$.

Seja $\boldsymbol{D}$, $(q \times q)$, a matriz de variâncias-covariâncias dos efeitos aleatórios, dada por

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

O vetor de parâmetros a ser estimado é $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, \boldsymbol{\tau}^T\right)^T$, em que $\tau_r$ $(r = 1, ..., q+m)$ é o $r$-ésimo elemento do vetor $\boldsymbol{\tau}$ que contém $q$ parâmetros de variância $(\tau_1, ..., \tau_q)^T$ e $m = \frac{q(q-1)}{2}$ $(q \geq 2)$ parâmetros de covariâncias entre os efeitos aleatórios $(\tau_{12}, ..., \tau_{q(q-1)})^T$.

Considere o vetor $\boldsymbol{\eta}$ de componentes de variância, $\boldsymbol{\eta} = (\eta_1, \eta_2, \boldsymbol{\eta}_3^T)^T$, em que $\eta_1 = \sigma_u^2$, $\eta_2 = \sigma^2$ e $\boldsymbol{\eta}_3 = \boldsymbol{\tau}$. Assim, o vetor de parâmetros a ser estimado é $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \boldsymbol{\eta}^T\right)^T$.

A função escore é definida por

$$\boldsymbol{U}(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{2.12}$$

Aplicando resultados matriciais de álgebra e diferenciação (vide, por exemplo, Magnus e Neudecker, 2002) ao logaritmo da função de verossimilhança, pode-se representar a função escore na forma

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

em que

$$\boldsymbol{U}_i(\boldsymbol{\theta}) = \begin{pmatrix} U_i^{\gamma}(\boldsymbol{\theta}) \\ \boldsymbol{U}_i^{\beta}(\boldsymbol{\theta}) \\ U_i^{\mu}(\boldsymbol{\theta}) \\ U_i^{\sigma_u^2}(\boldsymbol{\theta}) \\ U_i^{\sigma^2}(\boldsymbol{\theta}) \\ \boldsymbol{U}_i^{\tau}(\boldsymbol{\theta}) \end{pmatrix}.$$

A função escore parcial associada a $\gamma$ é dada por

$$\begin{aligned} U_i^{\gamma}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \gamma} \\ &= -\frac{1}{2}\frac{\partial \log|\boldsymbol{V}_i|}{\partial \gamma} - \frac{1}{2}\frac{\partial[\boldsymbol{r}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{r}_i]}{\partial \gamma} \\ &= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) + \boldsymbol{r}_i^T \boldsymbol{V}_i^{-1}\left[\begin{pmatrix} \mu \mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix} + \frac{1}{2}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right], \end{aligned}$$

com

$$\frac{\partial \boldsymbol{V}_i}{\partial \gamma} = \begin{pmatrix} 2\gamma\sigma_u^2 \boldsymbol{I}_{m_i} & \sigma_u^2 \boldsymbol{I}_{m_i} \\ \sigma_u^2 \boldsymbol{I}_{m_i} & \mathbf{0} \end{pmatrix}.$$

A função escore parcial para $\boldsymbol{\beta}$, associada aos efeitos fixos do modelo é dada por

$$\begin{aligned} \boldsymbol{U}_i^{\beta}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &= -\frac{1}{2}\frac{\partial\left[\boldsymbol{r}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{r}_i\right]}{\partial \boldsymbol{\beta}} \\ &= \begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i. \end{aligned}$$

A função escore parcial para $\mu$, associada aos efeitos fixos do modelo é dada por

$$\begin{aligned} U_i^{\mu}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \mu} \\ &= -\frac{1}{2}\frac{\partial\left[\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]}{\partial \mu} \\ &= \begin{pmatrix} \gamma\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i. \end{aligned}$$

As funções escore parciais para $\boldsymbol{\eta}$, associadas apenas aos componentes de variância são dadas por

$$\begin{aligned} \boldsymbol{U}_i^{\eta}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_r} \\ &= -\frac{1}{2}\frac{\partial \log|\boldsymbol{V}_i|}{\partial \boldsymbol{\eta}_r} - \frac{1}{2}\frac{\partial\left[\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]}{\partial \boldsymbol{\eta}_r} \\ &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_r}\right) - \frac{1}{2}\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \boldsymbol{\eta}_r}\boldsymbol{r}_i \\ &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_r}\right) + \frac{1}{2}\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_r}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \end{aligned}$$

em que

$$\begin{aligned} U_i^{\eta_1}(\boldsymbol{\theta}) &= U_i^{\sigma_u^2}(\boldsymbol{\theta}) \\ &= -\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) + \frac{1}{2}\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \end{aligned}$$

com

$$\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2} = \begin{pmatrix} \gamma^2\boldsymbol{I}_{m_i} & \gamma\boldsymbol{I}_{m_i} \\ \gamma\boldsymbol{I}_{m_i} & \boldsymbol{I}_{m_i} \end{pmatrix},$$

$$\begin{aligned} U_i^{\eta_2}(\boldsymbol{\theta}) &= U_i^{\sigma^2}(\boldsymbol{\theta}) \\ &= -\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) + \frac{1}{2}\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \end{aligned}$$

com

$$\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2} = \begin{pmatrix} \boldsymbol{I}_{m_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

e

$$\begin{aligned}\boldsymbol{U}_i^{\eta_3}(\boldsymbol{\theta}) &= \boldsymbol{U}_i^{\tau}(\boldsymbol{\theta}) \\ &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}^T}\right) + \frac{1}{2}\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i.\end{aligned}$$

## 2.4 Matriz de informação de Fisher

Vamos considerar um modelo estatístico geral com logaritmo da função de verossimilhança da forma

$$L(\boldsymbol{\theta}) = \log[f(\boldsymbol{W}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta})] = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}),$$

em que $\boldsymbol{W}_i = (\boldsymbol{y}_i^T, \boldsymbol{u}_i^T)^T$, $\boldsymbol{y}_i$, $\boldsymbol{u}_i$, $\boldsymbol{X}_i$ e $\boldsymbol{Z}_i$ são como definidos para os modelos (2.4)-(2.6) e $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)^T$ é um vetor $(k \times 1)$ de parâmetros no espaço $\boldsymbol{\Theta} \subseteq \Re^k$. Assumimos que as distribuições de probabilidade correspondentes para diferentes valores de $\boldsymbol{\theta}$ são distintas, para garantir a identificabilidade do modelo.

Consideremos também

$$\begin{aligned}\boldsymbol{U}(\boldsymbol{\theta}) &= \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= [L_1(\boldsymbol{\theta}), ..., L_k(\boldsymbol{\theta})]^T \\ &= \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\theta}),\end{aligned} \tag{2.13}$$

e

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{\theta}) &= \frac{1}{n}\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T},\end{aligned} \tag{2.14}$$

em que $\boldsymbol{U}_i(\boldsymbol{\theta}) \equiv \partial L_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$.

A matriz de informação de Fisher para o parâmetro $\boldsymbol{\theta}$ fica dada por

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{\theta}) &= -E_\theta[\boldsymbol{H}(\boldsymbol{\theta})] \\ &= E_\theta\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{U}_i(\boldsymbol{\theta})\boldsymbol{U}_i(\boldsymbol{\theta})^T\right],\end{aligned}$$

em que o valor esperado $E_\theta(\cdot)$ é calculado com respeito à função densidade $f(\boldsymbol{W}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta})$.

Vamos supor que as condições usuais de regularidade são satisfeitas (vide, por exemplo, Lehmann, 1983). Assim, o estimador de máxima verossimilhança consistente $\widehat{\boldsymbol{\theta}}$ existe, e $\boldsymbol{U}(\boldsymbol{\theta})$ e $\widehat{\boldsymbol{\theta}}$ seguem, no limite, distribuições normais, respectivamente: $n^{-1/2}\boldsymbol{U}(\boldsymbol{\theta}) \to \boldsymbol{N}[\boldsymbol{0}, \bar{\boldsymbol{F}}(\boldsymbol{\theta})]$ e $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to \boldsymbol{N}[\boldsymbol{0}, \boldsymbol{\Sigma}_\theta]$, com $\boldsymbol{\Sigma}_\theta = \bar{\boldsymbol{F}}^{-1}(\boldsymbol{\theta})$, em que $\bar{\boldsymbol{F}}(\boldsymbol{\theta}) = \lim_{n\to\infty}\boldsymbol{F}(\boldsymbol{\theta})$ (Dagenais e Dufour, 1991).

A seguir vamos apresentar três alternativas de estimadores consistentes de $\bar{\boldsymbol{F}}(\boldsymbol{\theta})$, que são comumente considerados:

$$\begin{aligned}\widehat{\boldsymbol{F}}_1(\tilde{\boldsymbol{\theta}}) &= -H(\tilde{\boldsymbol{\theta}}), \\ \widehat{\boldsymbol{F}}_2(\tilde{\boldsymbol{\theta}}) &= \frac{1}{n}\sum_{i=1}^{n}U_i(\tilde{\boldsymbol{\theta}})U_i(\tilde{\boldsymbol{\theta}})^T \text{ e} \\ \widehat{\boldsymbol{F}}_3(\tilde{\boldsymbol{\theta}}) &= \boldsymbol{F}(\tilde{\boldsymbol{\theta}}).\end{aligned} \tag{2.15}$$

Desde que $\tilde{\boldsymbol{\theta}}$ seja um estimador consistente de $\boldsymbol{\theta}$, cada um dos estimadores em (2.15) converge para $\bar{\boldsymbol{F}}(\boldsymbol{\theta})$.

A fim de obtermos as estimativas dos erros padrão para o estimador do vetor de parâmetros $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, \boldsymbol{\tau}^T\right)^T$ desenvolvemos a matriz de informação de Fisher, a qual assume a forma

$$\boldsymbol{F}(\boldsymbol{\theta}) = \begin{bmatrix} F_{\gamma\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\gamma\beta}(\boldsymbol{\theta}) & F_{\gamma\mu}(\boldsymbol{\theta}) & F_{\gamma\sigma_u^2}(\boldsymbol{\theta}) & F_{\gamma\sigma^2}(\boldsymbol{\theta}) & F_{\gamma\tau}(\boldsymbol{\theta}) \\ \boldsymbol{F}_{\beta\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\beta\beta}(\boldsymbol{\theta}) & \boldsymbol{F}_{\beta\mu}(\boldsymbol{\theta}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ F_{\mu\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\mu\beta}(\boldsymbol{\theta}) & F_{\mu\mu}(\boldsymbol{\theta}) & 0 & 0 & 0 \\ F_{\sigma_u^2\gamma}(\boldsymbol{\theta}) & \boldsymbol{0} & 0 & F_{\sigma_u^2\sigma_u^2}(\boldsymbol{\theta}) & F_{\sigma_u^2\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\sigma_u^2\tau}(\boldsymbol{\theta}) \\ F_{\sigma^2\gamma}(\boldsymbol{\theta}) & \boldsymbol{0} & 0 & F_{\sigma^2\sigma_u^2}(\boldsymbol{\theta}) & F_{\sigma^2\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\sigma^2\tau}(\boldsymbol{\theta}) \\ \boldsymbol{F}_{\tau\gamma}(\boldsymbol{\theta}) & \boldsymbol{0} & 0 & \boldsymbol{F}_{\tau\sigma_u^2}(\boldsymbol{\theta}) & \mathrm{b}F_{\tau\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\tau\tau}(\boldsymbol{\theta}) \end{bmatrix}. \tag{2.16}$$

As expressões para cada elemento da matriz $\boldsymbol{F}(\boldsymbol{\theta})$ são apresentadas no Apêndice C. Para estimar a matriz de informação $\boldsymbol{F}(\boldsymbol{\theta})$, foi usado o estimador $\widehat{\boldsymbol{F}}(\tilde{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{U}_i(\tilde{\boldsymbol{\theta}})\boldsymbol{U}_i(\tilde{\boldsymbol{\theta}})^T$.

É possível mostrar que o vetor de parâmetros associado apenas aos componentes de variância $\boldsymbol{\eta}$ é ortogonal ao vetor de parâmetros $\boldsymbol{\beta}$ e ao parâmetro $\mu$, associados aos efeitos

fixos do modelo, isto é, verifica-se que

$$\begin{aligned} \boldsymbol{F}_{\beta\eta,i}(\boldsymbol{\theta}) &= E_\theta\left[\frac{1}{n}\boldsymbol{U}_i^\beta(\boldsymbol{\theta})\boldsymbol{U}_i^\eta(\boldsymbol{\theta})^T\right] \\ &= \mathbf{0} \end{aligned}$$

e

$$\begin{aligned} \boldsymbol{F}_{\mu\eta,i}(\boldsymbol{\theta}) &= E_\theta\left[\frac{1}{n}U_i^\mu(\boldsymbol{\theta})\boldsymbol{U}_i^\eta(\boldsymbol{\theta})^T\right] \\ &= 0. \end{aligned}$$

## 2.5 Estimação de máxima verossimilhança

O método de máxima verossimilhança (MV) consiste em obter estimativas de parâmetros maximizando uma função de verossimilhança. Para a aplicação deste método, primeiro é obtida a função de verossimilhança como uma função dos parâmetros de um determinado modelo estatístico, baseada em suposições a respeito de distribuições de probabilidade. Como pode ser visto, por exemplo, em Casella e Berger (2002), as estimativas de máxima verossimilhança dos parâmetros são os valores que maximizam a função de verossimilhança, isto é, os valores dos parâmetros que tornam os valores observados da variável dependente mais prováveis, dadas as suposições de distribuições de probabilidade.

O valor do vetor de parâmetros $\boldsymbol{\theta}$ que maximiza o logaritmo da função de verossimilhança $L(\boldsymbol{\theta})$, em todo o espaço paramétrico $\Theta$, ou seja, $\widehat{\boldsymbol{\theta}}$ é chamado de estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ e satisfaz

$$L(\widehat{\boldsymbol{\theta}}) \geq sup_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}).$$

As estimativas de máxima verossimilhança dos parâmetros fixos e dos componentes de variância são obtidas maximizando (2.11) simultaneamente com respeito a $\gamma$, $\boldsymbol{\beta}$, $\mu$ e $\boldsymbol{\eta}$.

O processo iterativo para estimar os parâmetros fixos e os componentes de variância deve

alternar os passos

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

$$\boldsymbol{\beta}^{(k+1)}=\left[\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1(k)}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}\right]^{-1}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1(k)}\begin{pmatrix}\boldsymbol{y}_i-\gamma^{(k)}\mu^{(k)}\mathbf{1}_{m_i}\\ \boldsymbol{u}_i-\mu^{(k)}\mathbf{1}_{m_i}\end{pmatrix},$$

$$\begin{aligned}\mu^{(k+1)} &= \left[\sum_{i=1}^{n}\begin{pmatrix}\gamma^{(k)}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1(k)}\begin{pmatrix}\gamma^{(k)}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}\right]^{-1}\\ &\quad\times\sum_{i=1}^{n}\begin{pmatrix}\gamma^{(k)}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1(k)}\begin{pmatrix}\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta}^{(k)}\\ \boldsymbol{u}_i\end{pmatrix}\end{aligned}$$

e

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

para $k=0,1,2,...$ e $L(\gamma,\boldsymbol{\beta},\mu,\boldsymbol{\eta})$ denota o logaritmo da função de verossimilhança.

Nem sempre os estimadores de máxima verossimilhança podem ser expressos em forma explícita e, portanto, precisamos de um método iterativo para a obtenção das raízes das equações de máxima verossimilhança associadas. Nos casos em que as duas primeiras derivadas do logaritmo da função de verossimilhança existam, com relação aos parâmetros de interesse, os procedimentos usuais para calcular os estimadores de máxima verossimilhança estão baseados em uma expansão em série de Taylor em torno de alguma estimativa inicial. Nesse caso, podemos usar o algoritmo de Newton-Raphson ou algoritmo escore de Fisher.

O algoritmo escore de Fisher é uma versão do algoritmo de Newton-Raphson comumente usada para encontrar estimativas de parâmetros por máxima verossimilhança em modelos mistos (Osborne, 1992).

Longford (1993) apresentou o algoritmo escore de Fisher para estimar o vetor de parâmetros $\boldsymbol{\theta}$ como segue:

$$\boldsymbol{\theta}^{(k+1)}=\boldsymbol{\theta}^{(k)}+\boldsymbol{\Sigma}_\theta(\boldsymbol{\theta}^{(k)})\boldsymbol{U}^\theta(\boldsymbol{\theta}^{(k)}),\ \ k=0,1,2,....$$

Para iniciar o processo iterativo descrito acima, valores iniciais $\boldsymbol{\theta}^{(0)}$ devem ser fornecidos. Para amostras grandes e sob certas condições de regularidade, é razoável admitir que o

estimador de máxima verossimilhança $\widehat{\boldsymbol{\theta}}$ seja assintoticamente normal com média $\boldsymbol{\theta}$ e matriz de variâncias-covariâncias $\widehat{\boldsymbol{\Sigma}}_\theta$.

## 2.6 Predição dos efeitos aleatórios e da covariável longitudinal medida com erros

Em muitas aplicações é preciso predizer os efeitos aleatórios. Harville (1976) obteve este preditor através de uma extensão do Teorema de Gauss-Markov aplicado a modelos com presença de efeitos aleatórios. Laird e Ware (1982) consideram o uso do preditor linear de Bayes empírico no contexto de modelos mistos para dados com estrutura longitudinal. Assim, para a predição das variáveis latentes $\boldsymbol{b}_i$ (efeitos aleatórios) e $\boldsymbol{u}_i^*$ (variável verdadeira) vamos utilizar o método de Bayes empírico, usando o fato de que as médias condicionais de $\boldsymbol{b}_i$ e $\boldsymbol{u}_i^*$ dado o vetor de dados observados $\boldsymbol{W}_i = (\boldsymbol{y}_i^T, \boldsymbol{u}_i^T)^T$ seguem uma distribuição normal.

Assim, para a predição dos efeitos aleatórios $\boldsymbol{b}_i$, vamos partir da distribuição conjunta

$$\begin{pmatrix} \boldsymbol{W}_i \\ \boldsymbol{b}_i \end{pmatrix} \overset{\text{ind}}{\sim} \boldsymbol{N}_{2m_i+q}\left[\begin{pmatrix} \boldsymbol{\mu}_{iW} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_i & \boldsymbol{C}_{Wb} \\ \boldsymbol{C}_{Wb}^T & \boldsymbol{D} \end{pmatrix}\right],$$

em que a matriz de covariâncias entre $\boldsymbol{W}_i$ e $\boldsymbol{b}_i$ é $\boldsymbol{C}_{Wb} = \begin{pmatrix} \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{0} \end{pmatrix}$, e considerar a partição

$\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{i11} & \boldsymbol{\Sigma}_{i12} \\ \boldsymbol{\Sigma}_{i12}^T & \boldsymbol{\Sigma}_{i22} \end{pmatrix}$, em que
$\boldsymbol{\Sigma}_{i11} = \boldsymbol{V}_i$,
$\boldsymbol{\Sigma}_{i12} = \boldsymbol{C}_{Wb}$,
$\boldsymbol{\Sigma}_{i21} = \boldsymbol{C}_{Wb}^T$ e
$\boldsymbol{\Sigma}_{i22} = \boldsymbol{D}$.

De acordo com a propriedade da distribuição normal multivariada, temos que a distribuição condicional de $\boldsymbol{b}_i|\boldsymbol{W}_i$ assume a forma

$$\boldsymbol{b}_i|\boldsymbol{W}_i \sim \boldsymbol{N}_q(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}),$$

em que

$$\begin{aligned} \boldsymbol{\mu}_{2.1} &= \boldsymbol{0} + \boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}) \\ &= \boldsymbol{C}_{Wb}^T\boldsymbol{V}_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}) \end{aligned}$$

e

$$\begin{aligned}\boldsymbol{\Sigma}_{22.1} &= \boldsymbol{\Sigma}_{i22} - \boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}\boldsymbol{\Sigma}_{i12} \\ &= \boldsymbol{D} - \boldsymbol{C}_{Wb}^T\boldsymbol{V}_i^{-1}\boldsymbol{C}_{Wb}.\end{aligned}$$

Logo,

$$\boldsymbol{b}_i|\boldsymbol{W}_i \sim \boldsymbol{N}_q\left[\boldsymbol{C}_{Wb}^T\boldsymbol{V}_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}); \boldsymbol{D} - \boldsymbol{C}_{Wb}^T\boldsymbol{V}_i^{-1}\boldsymbol{C}_{Wb}\right].$$

Para $\boldsymbol{V}_i$ fixa, segue que o estimador de Bayes empírico do efeito aleatório $\boldsymbol{b}_i$ pode ser obtido pela média da distribuição a *posteriori*, ou seja

$$\begin{aligned}\widehat{\boldsymbol{b}}_i &= E(\widehat{\boldsymbol{b}_i|\boldsymbol{W}_i}) && (2.17)\\ &= \hat{\boldsymbol{C}}_{Wb}^T\hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}). && (2.18)\end{aligned}$$

Para uma revisão do processo de predição de $\boldsymbol{b}_i$ nos modelos mistos lineares normais, vide Harville (1976, 1977), Laird e Ware (1982) e Verbeke e Molemberghs (2000), entre outros.

De forma similar, para a predição da variável $\boldsymbol{u}_i^*$, vamos partir da distribuição conjunta

$$\begin{pmatrix}\boldsymbol{W}_i \\ \boldsymbol{u}_i^*\end{pmatrix} \overset{\text{ind}}{\sim} \boldsymbol{N}_{3m_i}\left[\begin{pmatrix}\boldsymbol{\mu}_{iW} \\ \mu\boldsymbol{1}_{m_i}\end{pmatrix}, \begin{pmatrix}\boldsymbol{V}_i & \boldsymbol{C}_{Wu^*} \\ \boldsymbol{C}_{Wu^*}^T & \sigma_u^2\boldsymbol{I}_{m_i}\end{pmatrix}\right],$$

em que a matriz de covariância entre $\boldsymbol{W}_i$ e $\boldsymbol{u}_i^*$ é $\boldsymbol{C}_{Wu^*} = \begin{pmatrix}\gamma\sigma_u^2\boldsymbol{I}_{m_i} \\ \sigma_u^2\boldsymbol{I}_{m_i}\end{pmatrix}$.

Assim,

$$\boldsymbol{u}_i^*|\boldsymbol{W}_i \sim \boldsymbol{N}_q(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}),$$

em que

$$\boldsymbol{\mu}_{2.1} = \mu\boldsymbol{1}_{m_i} + \boldsymbol{C}_{Wu^*}^T\boldsymbol{V}_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})$$

e

$$\boldsymbol{\Sigma}_{22.1} = \sigma_u^2\boldsymbol{I}_{m_i} - \boldsymbol{C}_{Wu^*}^T\boldsymbol{V}_i^{-1}\boldsymbol{C}_{Wu^*}.$$

Logo,

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

Para $\boldsymbol{V}_i$ fixa, segue que o estimador de Bayes empírico da variável $\boldsymbol{u}_i^*$ pode ser obtido

pela média da distribuição a *posteriori*, ou seja,

$$\widehat{\boldsymbol{u}^*}_i = E(\widehat{\boldsymbol{u}_i^*|\boldsymbol{W}_i}) \tag{2.19}$$

$$= \hat{\mu}\mathbf{1}_{m_i} + \hat{\boldsymbol{C}}_{Wu^*}^T \hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}). \tag{2.20}$$

## 2.7 Testes de hipóteses

Uma importante etapa na proposta de MML com erros de medição é a realização de testes de hipóteses acerca da significância de seus parâmetros. Neste contexto, um importante teste a ser desenvolvido é o teste para a significância da variável medida com erros, ou seja, podemos testar

$$H_0 : \gamma = 0 \ \textit{versus} \ H_1 : \gamma \neq 0.$$

Observemos que sob a hipótese $H_0$, o MML (3.3) proposto não se reduz ao MML normal apresentado em Laird e Ware (1982), visto que a covariável que se supunha medida com erros continua sendo aleatória no modelo. Portanto, o modelo (3.3) reduz-se a um modelo misto linear com com uma covariável aleatória.

De forma geral, conforme Degenais e Dufour (1991), consideremos o problema de testar uma hipótese linear geral $H_0 : \boldsymbol{A\theta} - \boldsymbol{a} = \mathbf{0}$ (ou $H_0 : \boldsymbol{A\theta} = \boldsymbol{a}$), em que $\boldsymbol{A\theta} - \boldsymbol{a}$ é uma função vetorial linear $(k_1 \times 1)$, $1 < k_1 < k$, a matriz $\boldsymbol{A}$ $(k_1 \times k)$ é conhecida e de posto $k_1$ e $\boldsymbol{a}$ é um vetor $(k_1 \times 1)$ conhecido.

Vários critérios podem ser utilizados para testar a hipótese linear geral $H_0 : \boldsymbol{A\theta} - \boldsymbol{a} = \mathbf{0}$. Neste trabalho vamos nos concentrar nos seguintes: teste da razão de verossimilhanças ($LR$), teste de Wald ($W$), teste do escore ($S$) de Rao (1948) e teste $C(\alpha)$ de Neyman (1959), como descritos em Degenais e Dufour (1991). As estatísticas $LR$, $W$, $S$ e $C(\alpha)$ dos testes para $H_0 : \boldsymbol{A\theta} - \boldsymbol{a} = \mathbf{0}$ são, respectivamente,

$$LR = 2\left[L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}^0)\right], \tag{2.21}$$

$$W = n(\boldsymbol{A}\hat{\boldsymbol{\theta}} - \boldsymbol{a})^T \left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_\theta \boldsymbol{A}^T\right]^{-1} (\boldsymbol{A}\hat{\boldsymbol{\theta}} - \boldsymbol{a}), \tag{2.22}$$

$$S = \frac{1}{n}\boldsymbol{U}^T(\hat{\boldsymbol{\theta}}^0)\hat{\boldsymbol{\Sigma}}_{\theta^0}\boldsymbol{U}(\hat{\boldsymbol{\theta}}^0) \ \text{ e} \tag{2.23}$$

$$PC(\tilde{\boldsymbol{\theta}}^0) = \frac{1}{n}\boldsymbol{U}^T(\tilde{\boldsymbol{\theta}}^0)\tilde{\boldsymbol{\Sigma}}_{\theta^0}\boldsymbol{A}^T\left[\boldsymbol{A}\tilde{\boldsymbol{\Sigma}}_{\theta^0}\boldsymbol{A}^T\right]^{-1}\boldsymbol{A}\tilde{\boldsymbol{\Sigma}}_{\theta^0}\boldsymbol{U}(\tilde{\boldsymbol{\theta}}^0), \tag{2.24}$$

em que $\hat{\boldsymbol{\theta}}^0$ e $\hat{\boldsymbol{\theta}}$ são os estimadores de MV de $\boldsymbol{\theta}$, respectivamente, sob o modelo reduzido e sob o modelo completo, e $\tilde{\boldsymbol{\theta}}^0$ é um estimador $\sqrt{n}$-consistente de $\boldsymbol{\theta}$ (pelo menos sob $H_0$) que satisfaz $\boldsymbol{A}\tilde{\boldsymbol{\theta}}^0 - \boldsymbol{a} = \boldsymbol{0}$. Estimativas consistentes para $\boldsymbol{\Sigma}_\theta$ podem ser obtidas pela inversa das matrizes de informação em (2.15).

Estamos supondo que $\hat{\boldsymbol{\Sigma}}_\theta$, $\hat{\boldsymbol{\Sigma}}_{\theta^0}$ e $\tilde{\boldsymbol{\Sigma}}_{\theta^0}$ têm posto linha completo e são definidas como em (2.15). Sob $H_0$, a distribuição assintótica de cada uma das estatísticas dos testes é $\chi^2_{k_1}$.

O critério $C(\alpha)$ de Neyman (1959) foi originalmente sugerido para testar hipóteses da forma $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$, em que $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ e $\boldsymbol{\theta}_1$ é um subvetor $(k_1 \times 1)$ de $\boldsymbol{\theta}$. Este critério pode ser visto como uma generalização do teste do escore de Rao, obtido substituindo o estimador de MV reduzido $\hat{\boldsymbol{\theta}}^0$ por $\tilde{\boldsymbol{\theta}}^0 = (\boldsymbol{\theta}_1^{0T}, \tilde{\boldsymbol{\theta}}_2^{0T})^T$, em que $\tilde{\boldsymbol{\theta}}_2^0$ é um estimador $\sqrt{n}$-consistente de $\boldsymbol{\theta}_2$ (veja Neyman, 1959).

## 2.8 Verificação da qualidade do ajuste

Após a realização de um ajuste de um modelo é necessário fazer uma verificação acerca das suposições assumidas, o que é feito por meio de uma análise de diagnóstico. Qualquer análise estatística deve incluir uma análise crítica dos pressupostos do modelo. A análise de resíduos tem sido o primeiro procedimento de diagnóstico sugerido para avaliar a adequação do modelo proposto e detectar uma eventual sensibilidade com relação a observações aberrantes. O processo de análise de resíduos, bem como da análise de diagnóstico em geral, já está bem definido para modelos normais lineares, assim como para algumas outras classes de modelos, como os modelos lineares generalizados (vide, por exemplo, Paula (2013) e trabalhos lá citados). Nobre (2003) e Nobre e Singer (2007) discutiram métodos de diagnóstico para modelos mistos com distribuição normal dos erros,

Para o modelo normal (2.8), uma medida natural da proximidade da $i$-ésima observação para o centro de distribuição é a distância de Mahalanobis

$$\delta_i(\boldsymbol{\theta}) = (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^T\boldsymbol{V}_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}),$$

que sob (2.8) tem distribuição qui-quadrado com $2m_i$ graus de liberdade $\left(\chi^2_{2m_i}\right)$. Substituindo os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ produz $\hat{\delta}_i \equiv \delta_i(\hat{\boldsymbol{\theta}})$, que tem, assintoticamente, a mesma distribuição qui-quadrado de $\delta_i(\boldsymbol{\theta})$.

Uma verificação de normalidade é conseguida através da transformação de cada $\hat{\delta}_i$ em um

desvio assintoticamente normal padrão e então pode-se traçar os valores ordenados contra os esperados das estatísticas de ordem normal; para o caso especial de regressão univariada com mínimos quadrados, este é um gráfico bem conhecido. Desvios da linha de 45º sugerem afastamento da normalidade, em especial, valores de $\hat{\delta}_i$ maiores do que o esperado indicam casos discrepantes (Gnanadesikan, 1977; Hopper e Mathews, 1982; Little 1988a; Little e Smith, 1987). Para realizar a transformação para a normalidade boas aproximações podem ser obtidas utilizando a raiz cúbica ou a raiz quarta de $\hat{\delta}_i$ (Hawkins e Wixley, 1986). Se este gráfico revelar valores extremos, pode-se ajustar um modelo $t$ de Student multivariado como alternativa.

Uma alternativa de aproximação é a de Wilson-Hilferty (vide Johnson *et al.* 1994), com a qual obtém-se

$$d_i^{[N]} = \frac{\left(\frac{\hat{\delta}_i}{2m_i}\right)^{1/3} - \left(1 - \frac{1}{9m_i}\right)}{\left(\frac{1}{9m_i}\right)^{1/2}}, \tag{2.25}$$

que tem, aproximadamente, distribuição normal padrão, $d_i^{[N]} \overset{\text{iid}}{\sim} N(0,1)$, $i = 1, ..., n$. Gráficos normais de probabilidade das distâncias transformadas $d_i^{[N]}$ podem ser utilizados para avaliar a qualidade do ajuste do modelo normal.

Tal resultado também nos permite avaliar a adequação do modelo empregando o envelope simulado proposto por Atkinson (1985). A fim de implementar esta ferramenta gráfica para a verificação do modelo, primeiro devem ser simuladas $J$ amostras de (2.8) usando as estimativas de ML $\hat{\boldsymbol{\theta}}$. Para a $j$-ésima amostra simulada, calcula-se a estimativa ML de $\boldsymbol{\theta}$ e os valores transformados $d_{j1}^{[N]}, ..., d_{jn}^{[N]}$ a partir de (2.25), que são ordenados como $d_{j(1)}^{[N]} \leq ... \leq d_{j(n)}^{[N]}$. Os pares ordenados

$$\left(\Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right), \hat{d}_{(i)}^{[N]}\right), \; i = 1, ..., n,$$

são representados em um gráfico, em que $\Phi^{-1}(\cdot)$ denota a função quantílica da distribuição normal padrão e $\hat{d}_{(i)}^{[N]}$ é calculado a partir de (2.25) com as estimativas de ML $\hat{\boldsymbol{\theta}}$. Os limites do envelope são dados por $\min_{j=1}^{J} d_{j(i)}^{[N]}$ e $\max_{j=1}^{J} d_{j(i)}^{[N]}$ e a linha que conecta os pontos

$$\left(\Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right), \sum_{j=1}^{J} d_{j(i)}^{[N]}/J\right), \; i = 1, ..., n, \; j = 1, ..., J$$

é também desenhada no gráfico. Este gráfico pode ser usado como base para nos guiar na avaliação do modelo postulado.

# Capítulo 3

# Modelo Misto Linear Elíptico com Erros de Medição

## 3.1 Introdução

Uma classe de modelos denominada modelos mistos lineares elípticos foi proposta por Savalli *et al.* (2006), em que o modelo marginal é também elíptico. Essa proposta traz muitas vantagens. Por exemplo, no desenvolvimento de procedimentos de estimação, metodologias de diagnóstico e testes para os componentes de variância e pode ser interpretada como uma generalização do modelo misto linear normal no sentido de flexibilização da curtose da distribuição dos erros. Outra vantagem é que quando os erros têm distribuição com caudas mais pesadas do que a distribuição normal, as estimativas de máxima verossimilhança dos parâmetros envolvidos são mais robustas contra observações aberrantes, no sentido da distância de Mahalanobis. Em Osorio *et al.* (2007) foram derivadas as curvaturas normais de influência local para vários esquemas de perturbação para a classe de modelos mistos lineares elípticos. Russo *et al.* (2009) estenderam a classe proposta por Savalli *et al.* (2006) substituindo o efeito fixo linear por um efeito fixo não linear, criando assim a classe denominada modelos mistos parcialmente não lineares elípticos para os quais desenvolveram procedimentos de estimação e metodologias de diagnóstico. Russo *et al.* (2012) desenvolveram para essa mesma classe testes para os componentes de variância através de uma estatística tipo escore proposta por Silvapulle e Silvapulle (1995). Estudos de sensibilidade da estatística do teste e vários estudos de simulação para avaliar os impactos da classificação incorreta da curtose no tamanho e poder do teste foram apresentados.

Ainda para os modelos mistos parcialmente não lineares elípticos, Russo *et al.* (2012) propuseram uma estrutura geral para as matrizes de variâncias-covariâncias dos erros e efeitos aleatórios incluindo como casos particulares estruturas autoregressivas e heteroscedásticas; procedimentos de estimação e metodologias de diagnóstico foram também desenvolvidos. Ibacache-Pulgar *et al.* (2012) apresentaram recentemente uma outra extensão da classe

proposta por Savalli *et al.* (2006), em que um componente fixo não paramétrico é adicionado aos efeitos fixos e aleatórios criando assim os modelos mistos semiparamétricos elípticos Foi assumido que o componente não paramétrico é do tipo *B-spline* cúbica. Um procedimento de estimação tipo *back-fitting* foi desenvolvido para a estimação dos parâmetros envolvidos e verificou-se que as estimativas, inclusive do componente não paramétrico, são robustas contra observações aberrantes como no caso paramétrico. Curvaturas de influência local foram derivadas para alguns esquemas de perturbação e algumas aplicações foram apresentadas. Ibacache-Pulgar e Paula (2011) apresentaram um estudo sobre a existência e unicidade das estimativas de máxima verossimilhança em modelos semiparamétricos $t$ de Student. Este trabalho foi estendido recentemente para a classe simétrica por Ibacache-Pulgar *et al.* (2013). A proposta deste trabalho é estender o modelo proposto por Savalli *et al.* (2006) no sentido de incluir um componente aleatório para uma variável longitudinal sujeita a erros de medidas. Descrevemos a seguir o modelo a ser estudado neste trabalho.

## 3.2 Modelo misto linear elíptico

Uma extensão do modelo (2.1)-(2.2) para a classe elíptica é apresentada em Savalli *et al.* (2006), em que

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \tag{3.1}$$

$$\boldsymbol{b}_i \overset{\text{iid}}{\sim} El_q(\boldsymbol{0}, \boldsymbol{D}), \quad i = 1, ..., n. \tag{3.2}$$

A distribuição normal é a mais utilizada na modelagem de muitos fenômenos, contudo, tem sido criticada por fornecer estimativas de máxima verossimilhança sensíveis a observações aberrantes. A fim de acomodar tais observações, as quais podem ser influentes nas conclusões de um estudo, diversos autores têm sugerido o uso de distribuições elípticas. Essas distribuições permitem estender os modelos já desenvolvidos com a suposição de erros normais, além de acomodar as observações aberrantes por meio de distribuições com caudas mais leves ou mais pesadas do que as da normal.

Neste sentido, Savalli *et al.* (2006) partiram da formulação hierárquica e assumiram a mesma estrutura obtida para a conjunta no caso normal, apresentada em (2.3), mas considerando uma distribuição elíptica diretamente para esta conjunta, ou seja,

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim El_{m_i+q}\left[\begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_{m_i} & \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{Z}_i^T & \boldsymbol{D} \end{pmatrix}\right].$$

Pelas propriedades das distribuições elípticas, descritas por exemplo em Arellano-Valle (1994), a distribuição marginal de $\boldsymbol{y}_i$ também é elíptica. Dessa forma, assim como no modelo

misto linear normal, as inferências puderam ser baseadas na distribuição marginal $\boldsymbol{y}_i \sim El_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_{m_i})$.

### 3.2.1 Inclusão de uma variável medida com erros

Vamos considerar agora um modelo de erros elípticos, incluindo no modelo (3.1) uma covariável contínua sujeita a erros de medição. Assim, vamos ter um modelo misto linear com erros elípticos em uma variável. O modelo proposto é dado por

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \gamma\boldsymbol{u}_i^* + \boldsymbol{\epsilon}_i, \tag{3.3}$$

$$\boldsymbol{u}_i = \boldsymbol{u}_i^* + \boldsymbol{e}_i,\ i = 1, ..., n, \tag{3.4}$$

em que $\boldsymbol{y}_i$ é um vetor $m_i$-dimensional de respostas observadas para o $i$-ésimo grupo, $\boldsymbol{X}_i$ é uma matriz $(m_i \times p)$ que contém as variáveis explanatórias, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos, $\boldsymbol{Z}_i$ é uma matriz de planejamento $(m_i \times q)$ de efeitos aleatórios $\boldsymbol{b}_i$, $\boldsymbol{u}_i^*$ é a covariável $m_i$-dimensional sujeita a erros de medição com $\boldsymbol{u}_i$ denotando seus valores observados, $\gamma$ é um parâmetro escalar fixo enquanto que $\boldsymbol{\epsilon}_i$ and $\boldsymbol{e}_i$ denotam os erros $m_i$-dimensionais.

Assumimos que os efeitos aleatórios $\boldsymbol{b}_i$, a covariável sujeita a erros de medição $\boldsymbol{u}_i^*$ e os erros do modelo $\boldsymbol{\epsilon}_i$ and $\boldsymbol{e}_i$ são não correlacionados com a seguinte distribuição conjunta:

$$\begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{u}_i^* \\ \boldsymbol{\epsilon}_i \\ \boldsymbol{e}_i \end{bmatrix} \sim \mathrm{El}_{3m_i+q}\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\mu}_i^* \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{D} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_u & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \sigma^2\boldsymbol{I}_{m_i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \sigma_e^2\boldsymbol{I}_{m_i} \end{pmatrix}\right], \tag{3.5}$$

em que $\mathrm{El}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota uma distribuição elíptica $r$-dimensional com parâmetro de posição $\boldsymbol{\mu}$ e matriz de dispersão $\boldsymbol{\Sigma}$ (vide, por exemplo, Fang *et al.*, 1990).

Seja $\boldsymbol{W}_i = [\boldsymbol{y}_i^T, \boldsymbol{u}_i^T]^T$ o vetor de dados observados, para $i = 1, ..., n$. A partir das equações

3.3, 3.4 e 3.5 temos que

$$\boldsymbol{W}_i = \begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_i & \gamma\boldsymbol{I}_{m_i} & \boldsymbol{I}_{m_i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{m_i} & \boldsymbol{0} & \boldsymbol{I}_{m_i} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{u}_i^* \\ \boldsymbol{\epsilon}_i \\ \boldsymbol{e}_i \end{bmatrix}, \tag{3.6}$$

para $i = 1, ..., n$. Isto é, $\boldsymbol{W}_i$ é uma combinação linear dos vetores latentes. Agora, por 3.5 segue que

$$\boldsymbol{W}_i \sim \mathrm{El}_{2\mathrm{m_i}} \left( \begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} + \gamma\boldsymbol{\mu}_i^* \\ \boldsymbol{\mu}_i^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_i & \gamma\boldsymbol{\Sigma}_u \\ \gamma\boldsymbol{\Sigma}_u & \boldsymbol{\Sigma}_u + \sigma_e^2\boldsymbol{I}_{m_i} \end{bmatrix} \right), \tag{3.7}$$

em que $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \gamma^2\boldsymbol{\Sigma}_u + \sigma^2\boldsymbol{I}_{m_i}$, para $i = 1, ..., n$. Geralmente, a análise inferencial clássica é baseada nesta distribuição marginal.

Assim como no caso normal, a verificação da identificabilidade em MML elípticos com erros de medição é uma etapa fundamental. Porém, pelas análises de identificabilidade apresentadas na Seção 2.2.1 e tendo em vista que as distribuições elípticas também são totalmente especificadas pelos dois primeiros momentos, temos que as condições de identificabilidade para o modelo elíptico proposto são uma extensão natural do caso normal, quando a curtose é conhecida. Assim, para evitar problemas de identificabilidade vamos assumir que o parâmetro de escala associado com os erros de medição, $\sigma_e^2$, é conhecido e que a matriz de planejamento $\boldsymbol{X}_i^*$ seja de posto completo (vide Seção 2.2.1). Para isso, devemos supor que $\beta_0$ também seja conhecido. Além disso, como no caso normal, assumimos que $\boldsymbol{\mu}_i^* = \mu\boldsymbol{1}_{m_i}$ e $\boldsymbol{\Sigma}_u = \sigma_u^2\boldsymbol{I}_{m_i}$.

Assim, o vetor de parâmetros a ser estimado é $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, vech(\boldsymbol{D})\right)$.

O logaritmo da função de verossimilhança para o vetor $\boldsymbol{\theta}$ fica dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}), \tag{3.8}$$

em que

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

$\boldsymbol{\delta}_i = (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^T \boldsymbol{V}_i^{-1} (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})$ $(i = 1, ..., n)$ é a distância de Mahalanobis, com

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

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

Podemos enumerar algumas vantagens do uso da abordagem conjunta trabalhando com a distribuição marginal:

i) forma fechada para a distribuição marginal que também é elíptica, sem a necessidade do uso de métodos de aproximação de integrais;

ii) preservação das médias e estrutura de variâncias-covariâncias das variáveis $\boldsymbol{y}_i$ e $\boldsymbol{u}_i$ no modelo marginal;

iii) possibilidade de definição de curtoses diferentes para cada grupo;

iv) derivação de funções escore, informação de Fisher e procedimentos iterativos em forma fechada, facilitando a aplicação de métodos de estimação, inferência e diagnóstico como no caso normal;

v) possibilidade de predição das variáveis latentes $\boldsymbol{b}_i$ e $\boldsymbol{u}_i^*$ através do método Bayes empírico.

Uma desvantagem dessa metodologia é que a curtose deve ser a mesma em cada grupo para as distribuições marginais dos $\boldsymbol{y}_i$, $\boldsymbol{b}_i$, $\boldsymbol{u}_i$ e $\boldsymbol{u}_i^*$.

## 3.3 Função escore

Seja $\boldsymbol{D}$, $(q \times q)$, a matriz de variâncias-covariâncias dos efeitos aleatórios, dada por

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

e seja $\tau_r$ $(r = 1, ..., q+s)$ o $r$-ésimo elemento do vetor $\boldsymbol{\tau} = vech(\boldsymbol{D})$, que contém $q$ parâmetros de variância $(\tau_1, ..., \tau_q)^T$ e $s = \frac{q(q-1)}{2}$ $(q \geq 2)$ parâmetros de covariâncias entre os efeitos aleatórios $(\tau_{12}, ..., \tau_{q(q-1)})^T$. Assim, o vetor de parâmetros a ser estimado fica dado por $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, \boldsymbol{\tau}^T\right)^T$.

Além disso, assumindo que a função geradora de densidades $g(\cdot)$ apresentada na Seção 1.5, é contínua e diferenciável, podemos definir as seguintes quantidades:

$$v(\delta_i) = -2W_g(\delta_i),$$

em que

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

Aplicando resultados matriciais de álgebra e diferenciação no logaritmo da função de verossimilhança (vide, por exemplo, Magnus e Neudecker, 2002), podemos representar a função escore na forma

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

em que $\boldsymbol{U}_i(\boldsymbol{\theta}) = \left(U_i^{\gamma}(\boldsymbol{\theta}), \boldsymbol{U}_i^{\beta}(\boldsymbol{\theta})^T, U_i^{\mu}(\boldsymbol{\theta}), U_i^{\sigma_u^2}(\boldsymbol{\theta}), U_i^{\sigma^2}(\boldsymbol{\theta}), \boldsymbol{U}_i^{\tau}(\boldsymbol{\theta})^T\right)^T$.

Os cálculos algébricos dos resultados que se seguem são apresentados no Apêndice A.1.

A função escore parcial associada a $\gamma$, que é um parâmetro que está presente tanto da média quanto da estrutura de variância-covariância do modelo, é dada por

$$\begin{aligned} U_i^{\gamma}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \gamma} \\ &= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) + \frac{1}{2}v(\delta_i)\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix} \mu\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix} + \boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right], \end{aligned}$$

em que $\delta_i = \boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i$, com $\boldsymbol{r}_i = (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})$.

As funções escore parciais para $\boldsymbol{\beta}$ e $\mu$, associadas apenas aos efeitos fixos do modelo são dadas por

$$\begin{aligned} \boldsymbol{U}_i^{\beta}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &= v(\delta_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i \end{aligned}$$

e

$$\begin{aligned} U_i^{\mu}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \mu} \\ &= v(\delta_i)\begin{pmatrix} \gamma\mathbf{1}_i \\ \mathbf{1}_i \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i. \end{aligned}$$

Considere o vetor $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)^T$ com os parâmetros apenas de componentes de variância, em que $\eta_1 = \sigma_u^2$, $\eta_2 = \sigma^2$ e $\eta_3 = \boldsymbol{\tau}$. As funções escore parciais para $\boldsymbol{\eta}$ são dadas por

$$\begin{aligned} U_i^{\eta_k}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_k} \\ &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_k}\right) + \frac{1}{2}v(\delta_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_k}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \end{aligned}$$

em que

$$\begin{aligned} U_i^{\eta_1}(\boldsymbol{\theta}) &= U_i^{\sigma_u^2}(\boldsymbol{\theta}) \\ &= -\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) + \frac{1}{2}v(\delta_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \end{aligned}$$

$$\begin{aligned} U_i^{\eta_2}(\boldsymbol{\theta}) &= U_i^{\sigma^2}(\boldsymbol{\theta}) \\ &= -\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) + \frac{1}{2}v(\delta_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \end{aligned}$$

e

$$\begin{aligned} \boldsymbol{U}_i^{\eta_3}(\boldsymbol{\theta}) &= \boldsymbol{U}_i^{\tau}(\boldsymbol{\theta}) \\ &= -\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}}\right) + \frac{1}{2}v(\delta_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i. \end{aligned}$$

A quantidade $v(\delta_i)$ que aparece nas funções escore pode ser interpretada como um peso, e como $g(\delta_i)$ é em geral uma função decrescente, segue que $v(\delta_i) > 0$. Para algumas distribuições para os erros, tais como $t$ de Student, $v(\delta_i)$ é inversamente proporcional à distância de Mahalanobis. Assim, observações aberrantes no sentido dessa distância receberão um peso menor na solução das equações de estimação (funções escore) dos parâmetros envolvidos, em particular com os parâmetros relacionados com a covariável medida com erros. A Tabela 3.1 apresenta as expressões de $v(\delta_i)$ para algumas distribuições elípticas. No caso da distribuição exponencial potência, o parâmetro $\zeta$ é uma medida de curtose; se $-1 < \zeta < 0$, a distribuição tem caudas mais leves do que a normal e se $0 < \zeta < 1$, a distribuição tem caudas mais pesadas. Quando $\zeta = 0$ recaímos na distribuição normal e portanto, este pode ser visto como um parâmetro de afastamento da normalidade.

Na seção seguinte calculamos a matriz de informação de Fisher associada ao vetor de parâmetros $\boldsymbol{\theta}$. Essa matriz será utilizada na obtenção da matriz de variâncias-covariâncias assintótica do estimador de máxima verossimilhança $\widehat{\boldsymbol{\theta}}$, ou seja, calcular o erro padrão das

Tabela 3.1: Expressões das quantidades $v(\delta_i)$ para algumas distribuições elípticas.

| Distribuição | $v(\delta_i) = -2W_g(\delta_i)$ |
|---|---|
| Normal | $1$ |
| $t$ de Student | $\frac{\nu+2m_i}{\nu+\delta_i}$ |
| 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 | $\delta_i^{1/2}\tanh\left(\frac{\delta_i^{1/2}}{2}\right)$ |

estimativas, baseado na inversa da matriz de informação de Fisher. A prova dos resultados que se seguem e os cálculos algébricos são apresentados no Apêndice C. Para mais detalhes referentes a estes resultados, no caso dos modelos mistos lineares elípticos, vide Mitchell (1989) e Savalli (2006), e no caso dos modelos mistos semiparamétricos elípticos, vide Ibacache-Pulgar (2009).

## 3.4 Matriz de informação de Fisher

A fim de obter as estimativas dos erros padrão para o vetor de parâmetros $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, \boldsymbol{\tau}^T\right)^T$ foi calculada a matriz de informação de Fisher, a qual assume a forma

$$\boldsymbol{F}(\boldsymbol{\theta}) = \begin{bmatrix} F_{\gamma\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\gamma\beta}(\boldsymbol{\theta}) & F_{\gamma\mu}(\boldsymbol{\theta}) & F_{\gamma\sigma_u^2}(\boldsymbol{\theta}) & F_{\gamma\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\gamma\tau}(\boldsymbol{\theta}) \\ \boldsymbol{F}_{\beta\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\beta\beta}(\boldsymbol{\theta}) & \boldsymbol{F}_{\beta\mu}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ F_{\mu\gamma}(\boldsymbol{\theta}) & \boldsymbol{F}_{\mu\beta}(\boldsymbol{\theta}) & F_{\mu\mu}(\boldsymbol{\theta}) & 0 & 0 & \mathbf{0} \\ F_{\sigma_u^2\gamma}(\boldsymbol{\theta}) & \mathbf{0} & 0 & F_{\sigma_u^2\sigma_u^2}(\boldsymbol{\theta}) & F_{\sigma_u^2\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\sigma_u^2\tau}(\boldsymbol{\theta}) \\ F_{\sigma^2\gamma}(\boldsymbol{\theta}) & \mathbf{0} & 0 & F_{\sigma^2\sigma_u^2}(\boldsymbol{\theta}) & F_{\sigma^2\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\sigma^2\tau}(\boldsymbol{\theta}) \\ \boldsymbol{F}_{\tau\gamma}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} & \boldsymbol{F}_{\tau\sigma_u^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\tau\sigma^2}(\boldsymbol{\theta}) & \boldsymbol{F}_{\tau\tau}(\boldsymbol{\theta}) \end{bmatrix}. \tag{3.9}$$

Conforme foi discutido para o caso normal, a matriz de informação $\boldsymbol{F}(\boldsymbol{\theta})$ será estimada usando o estimador da matriz de variâncias-covariâncias assintótica dos estimadores de máxima verossimilhança $\widehat{\boldsymbol{I}}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{U}_i(\boldsymbol{\theta})\boldsymbol{U}_i(\boldsymbol{\theta})^T$. Desta forma, a matriz de informação de Fisher será dada por

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{\theta}) &= \mathrm{E}\left[\widehat{\boldsymbol{I}}(\boldsymbol{\theta})\right] \\ &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{U}_i(\boldsymbol{\theta})\boldsymbol{U}_i(\boldsymbol{\theta})^T\right]. \end{aligned} \tag{3.10}$$

As expressões para cada elemento da matriz $\boldsymbol{F}(\boldsymbol{\theta})$ são apresentadas no Apêndice C.

## 3.5 Estimação de máxima verossimilhança

Para o processo de otimização precisamos supor que o logaritmo da função de verossimilhança $L(\boldsymbol{\theta})$, definido em (3.8), é uma função que satisfaz certas condições de regularidade no contexto de regressão paramétrica.

O valor do vetor de parâmetros $\boldsymbol{\theta}$ que maximiza $L(\boldsymbol{\theta})$ em todo o espaço paramétrico $\Theta$, ou seja, $\widehat{\boldsymbol{\theta}}$ é chamado de estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ e satisfaz

$$L(\widehat{\boldsymbol{\theta}}) \geq sup_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}).$$

As estimativas de máxima verossimilhança dos parâmetros fixos e dos componentes de variância são obtidas maximizando (3.8) simultaneamente com respeito a $\gamma$, $\boldsymbol{\beta}$, $\mu$ e $\boldsymbol{\eta}$.

A solução da equação $U^{\gamma}(\boldsymbol{\theta}) = \mathbf{0}$ fica dada por

$$\sum_{i=1}^{n}\left\{\widehat{\boldsymbol{c}}_i + v(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{A}}_i\left[\begin{pmatrix}\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}\\ \boldsymbol{u}_i - \widehat{\mu}\mathbf{1}_{m_i}\end{pmatrix} - \begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\widehat{\gamma}\right]\right\} = \mathbf{0}$$

e

$$\begin{aligned}\widehat{\gamma} \;=\; & \left[-\sum_{i=1}^{n}\frac{1}{2}v(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{A}}_i\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right]^{-1}\\ & \sum_{i=1}^{n}\left[\widehat{\boldsymbol{c}}_i + v(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \frac{1}{2}v(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{A}}_i\begin{pmatrix}\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}\\ \boldsymbol{u}_i - \widehat{\mu}\mathbf{1}_{m_i}\end{pmatrix}\right],\end{aligned}$$

em que $\boldsymbol{A} = \boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}$.

A solução da equação $\boldsymbol{U}^{\beta}(\boldsymbol{\theta}) = \mathbf{0}$ é dada por

$$\sum_{i=1}^{n} v(\widehat{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\widehat{\boldsymbol{V}}_i^{-1}\left[\begin{pmatrix}\boldsymbol{y}_i - \widehat{\gamma}\widehat{\mu}\mathbf{1}_{m_i}\\ \boldsymbol{u}_i - \widehat{\mu}\mathbf{1}_{m_i}\end{pmatrix} - \begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}\widehat{\boldsymbol{\beta}}\right] = \mathbf{0};$$

$$\sum_{i=1}^{n} v(\widehat{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\boldsymbol{y}_i - \widehat{\gamma}\widehat{\mu}\mathbf{1}_{m_i}\\ \boldsymbol{u}_i - \widehat{\mu}\mathbf{1}_{m_i}\end{pmatrix} - \sum_{i=1}^{n} v(\widehat{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}\widehat{\boldsymbol{\beta}} = \mathbf{0}$$

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} v(\widehat{\delta}_i) \begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1} \begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}\right]^{-1} \sum_{i=1}^{n} v(\widehat{\delta}_i) \begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1} \begin{pmatrix} \boldsymbol{y}_i - \widehat{\gamma}\widehat{\mu}\mathbf{1}_{m_i} \\ \boldsymbol{u}_i - \widehat{\mu}\mathbf{1}_{m_i} \end{pmatrix}.$$

A solução da equação $U^{\mu}(\boldsymbol{\theta}) = \mathbf{0}$ é dada por

$$\sum_{i=1}^{n} v(\widehat{\delta}_i) \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1} \left[\begin{pmatrix} \boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} \\ \boldsymbol{u}_i \end{pmatrix} - \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix} \widehat{\mu}\mathbf{1}_{m_i}\right] = 0;$$

$$\sum_{i=1}^{n} v(\widehat{\delta}_i) \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1} \begin{pmatrix} \boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} \\ \boldsymbol{u}_i \end{pmatrix} - \sum_{i=1}^{n} v(\widehat{\delta}_i) \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1} \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix} \widehat{\mu} = 0$$

$$\widehat{\mu} = \left[\sum_{i=1}^{n} v(\widehat{\delta}_i) \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1} \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}\right]^{-1} \sum_{i=1}^{n} v(\widehat{\delta}_i) \begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1} \begin{pmatrix} \boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} \\ \boldsymbol{u}_i \end{pmatrix}.$$

Assim, o processo iterativo para estimar os parâmetros fixos e os componentes de variância deve alternar os passos

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

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

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

e

$$\boldsymbol{\eta}^{(k+1)} = \text{argmax}_{\boldsymbol{\eta}} \left\{ L\left(\gamma^{(k)}, \boldsymbol{\beta}^{(k)}, \mu^{(k)}, \boldsymbol{\eta}\right)\right\}, \tag{3.11}$$

para $k = 0, 1, 2, ...$ e $L(\gamma, \boldsymbol{\beta}, \mu, \boldsymbol{\eta})$ denota o logaritmo da função de verossimilhança. Valores iniciais devem ser fornecidos para inicializar o processo iterativo. Em particular, para os modelos elípticos não gaussianos, podemos considerar as estimativas obtidas a partir do modelo normal para iniciar o processo.

Nota-se, pelas equações do processo iterativo para obter $\hat{\gamma}$, $\hat{\boldsymbol{\beta}}$ e $\hat{\mu}$, a presença das quantidades $v(\delta_i)$ $(i = 1, ..., n)$ que fazem o papel de pesos. Em particular, em modelos com distribuição para os erros com caudas mais pesadas do que a normal $v(\delta_i)$ decresce à medida que $\delta_i$ cresce. Logo, as estimativas de máxima verossimilhança serão robustas no sentido da distância de Mahalanobis.

Devido à similaridade entre a inferência nos modelos elípticos e normal, sendo que o modelo normal é uma caso particular dos modelos elípticos, é razoável admitir que, para amostras grandes e sob certas condições de regularidade, o estimador de máxima verossimilhança de $\widehat{\boldsymbol{\theta}}$ seja assintoticamente normal com média $\boldsymbol{\theta}$ e matriz de variâncias-covariâncias $\boldsymbol{F}^{-1}(\hat{\theta})$.

## 3.6 Predição dos efeitos aleatórios e da covariável longitudinal medida com erros

Nesta seção seguiremos as mesmas ideias usadas no caso normal para a predição das variáveis latentes $\boldsymbol{b}_i$ (efeitos aleatórios) e $\boldsymbol{u}_i^*$ (variável verdadeira), ou seja, vamos utilizar o método de Bayes empírico.

Assim, para a predição dos efeitos aleatórios $\boldsymbol{b}_i$ do modelo (3.3) vamos usar o fato que a média condicional de $\boldsymbol{b}_i$ dado o vetor de dados observados $\boldsymbol{W}_i = (\boldsymbol{y}_i^T, \boldsymbol{u}_i^T)^T$ segue uma

distribuição elíptica. Partindo da distribuição conjunta

$$\begin{pmatrix} \boldsymbol{W}_i \\ \boldsymbol{b}_i \end{pmatrix} \overset{\text{ind}}{\sim} El_{2m_i+q}\left[\begin{pmatrix} \boldsymbol{\mu}_{iW} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_i & \boldsymbol{C}_{Wb} \\ \boldsymbol{C}_{Wb}^T & \boldsymbol{D} \end{pmatrix}\right],$$

em que $\boldsymbol{C}_{Wb} = \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{0} \end{bmatrix}$ é a matriz de covariância entre $\boldsymbol{W}_i$ e $\boldsymbol{b}_i$, temos que

$$\boldsymbol{b}_i|\boldsymbol{W}_i \sim El_q\left[\boldsymbol{C}_{Wb}^T\boldsymbol{V}_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}); \boldsymbol{D} - \boldsymbol{C}_{Wb}^T\boldsymbol{V}_i^{-1}\boldsymbol{C}_{Wb}\right].$$

Para $\boldsymbol{V}_i$ fixa, segue que o estimador de Bayes empírico dos efeitos aleatórios $\boldsymbol{b}_i's$ pode ser obtido pela média da distribuição a *posteriori*, ou seja,

$$\begin{aligned} \widehat{\boldsymbol{b}}_i &= E(\widehat{\boldsymbol{b}_i|\boldsymbol{W}}_i) && (3.12) \\ &= \hat{\boldsymbol{C}}_{Wb}^T\hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}). && (3.13) \end{aligned}$$

Assim, o vetor de efeitos aleatórios predito é dado por $\hat{\boldsymbol{b}} = (\hat{\boldsymbol{b}}_1^T, ..., \hat{\boldsymbol{b}}_n^T)^T$.

Para uma revisão no processo de estimação de $\boldsymbol{b}_i$ nos modelos mistos lineares elípticos, vide Savalli *et al.* (2006).

De forma similar, para a predição da variável $\boldsymbol{u}_i^*$, e partindo da distribuição conjunta

$$\begin{pmatrix} \boldsymbol{W}_i \\ \boldsymbol{u}_i^* \end{pmatrix} \overset{\text{ind}}{\sim} El_{3m_i}\left[\begin{pmatrix} \boldsymbol{\mu}_{iW} \\ \mu\boldsymbol{1}_{m_i} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_i & \boldsymbol{C}_{Wu^*} \\ \boldsymbol{C}_{Wu^*}^T & \sigma_u^2\boldsymbol{I}_{m_i} \end{pmatrix}\right],$$

em que $\boldsymbol{C}_{Wu^*} = \begin{pmatrix} \gamma\sigma_u^2\boldsymbol{I}_{m_i} \\ \sigma_u^2\boldsymbol{I}_{m_i} \end{pmatrix}$ é a matriz de covariâncias entre $\boldsymbol{W}_i$ e $\boldsymbol{u}_i^*$, concluímos que a distribuição de $\boldsymbol{u}_i^*$ dado $\boldsymbol{W}_i$ é dada por

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

Para $\boldsymbol{V}_i$ fixa, segue que o estimador de Bayes empírico da variável $\boldsymbol{u}_i^*$ pode ser obtido pela média da distribuição a *posteriori*, ou seja

$$\begin{aligned} \widehat{\boldsymbol{u}^*}_i &= E(\widehat{\boldsymbol{u}_i^*|\boldsymbol{W}}_i) && (3.14) \\ &= \hat{\mu}\boldsymbol{1}_{m_i} + \hat{\boldsymbol{C}}_{Wu^*}^T\hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}). && (3.15) \end{aligned}$$

Assim, a variável verdadeira predita é dada por $\widehat{\boldsymbol{u}^*} = \left(\widehat{\boldsymbol{u}^*}_1^T, ..., \widehat{\boldsymbol{u}^*}_{m_n}^T\right)^T$.

É interessante notar que as esperanças condicionais, ou os preditores de Bayes empírico,

obtidas em (3.12) e (3.14), para o caso elíptico apresentam as mesmas expressões que as obtidas para o caso normal em (2.17) e (2.19). No entanto, as estimativas de máxima verossimilhança devem ser diferentes. Ou seja, as mesmas expressões simbólicas, porém os valores estimados devem variar de acordo com a distribuição utilizada.

## 3.7 Distribuição $t$ de Student

Uma das distribuições mais conhecidas da classe elíptica é a distribuição $t$ de Student, cuja curtose é controlada pelos graus de liberdade, $\nu$. A distribuição $t$ de Student tem a propriedade de apresentar caudas mais pesadas do que as caudas da distribuição normal e a aproximação para a normal ocorre à medida que $\nu$ aumenta. Assim, a distribuição $t$ de Student é recomendada para a análise de dados que apresentam indícios de caudas pesadas (através, por exemplo, da análise de resíduos) quando analisados sob erros normais. Como as estimativas de máxima verossimilhança sob erros $t$ de Student são robustas contra observações aberrantes (vide, por exemplo, discussão em Lucas, 1997) a necessidade de aplicação de métodos robustos para a estimação dos parâmetros torna-se menos atrativa do que sob erros normais com pontos aberrantes. A distribuição $t$ de Student pode ser, portanto, recomendada para os erros do modelo (2.1)-(2.2) nos casos de indícios de caudas mais pesadas do que a distribuição normal, cuja indicação pode ser obtida por análises de resíduos.

A função densidade de probabilidade da distribuição $t$ de Student padrão com $\nu$ graus de liberdade, pode ser escrita como

$$f(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \ -\infty < y < \infty,$$

em que $\Gamma(y)$ é a função gamma.

Para a distribuição $t$ de Student padrão com $\nu$ graus de liberdade, a média é 0 e a variância é $\nu/(\nu-2)$ para $\nu > 2$. Uma vez que a distribuição $t$ de Student é simétrica, sua média e mediana são iguais. Quando $\nu = 1$, a distribuição $t$ de Student padrão reduz-se à distribuição Cauchy. Uma distribuição $t$ de Student multivariada é uma generalização da sua versão univariada. A função densidade de probabilidade de um vetor aleatório $n$-dimensional $\boldsymbol{y} = (y_1, ..., y_n)^T$ seguindo uma distribuição $t$ de Student multivariada com $\nu$

graus de liberdade e parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, denotada por $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, é dada por

$$f(\boldsymbol{y}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma(\nu/2)\nu^{n/2}\pi^{n/2}|\boldsymbol{\Sigma}|^{1/2}\left[1+\frac{1}{\nu}(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]^{(\nu+n)/2}}, \ \boldsymbol{y} \in \Re^n,$$

em que $\boldsymbol{\mu}$ é um vetor $(n \times 1)$ e $\boldsymbol{\Sigma}$ é uma matriz $(n \times n)$. Para $\nu > 1$, o vetor de médias e a matriz de variâncias-covariâncias de $\mathbf{x}$ são dados, respectivamente, por

$$E(\boldsymbol{y}) = \boldsymbol{\mu} \ \text{ e } \ Var(\boldsymbol{y}) = \left(\frac{\nu}{\nu-2}\right)\boldsymbol{\Sigma}, \ \text{para } \nu > 2.$$

Para uma inferência robusta podemos substituir a distribuição normal multivariada geralmente assumida em um modelo de regressão pela distribuição $t$ multivariada correspondente. Por exemplo, para um modelo MML, podemos assumir que os efeitos aleatórios ou os erros aleatórios intraindivíduos, ou ambos, seguem uma distribuição $t$ de Student. Tais abordagens robustas podem ser encontradas em Lange *et al.* (1989), Pinheiro *et al.* (2001) e Song *et al.* (2007).

Existem outras formas para a obtenção de estimativas robustas. Por exemplo, podemos substituir uma distribuição normal assumida em um modelo por uma *mistura* de duas ou mais distribuições normais. Também, podemos substituir uma distribuição paramétrica assumida por uma distribuição não paramétrica, como fizeram Lai e Shih (2003) para modelos com efeitos mistos não lineares.

### 3.7.1 Modelos mistos lineares $t$ de Student

Em MMLs é comum assumir que os efeitos aleatórios e os erros intraindivíduos seguem distribuições normais multivariadas. Assim, a inferência por máxima verossimilhança para MMLs é sensível a pontos aberrantes. Em inferência robusta, Wu (2010) apresenta uma abordagem que consiste em substituir as distribuições normais multivariadas pelas correspondentes distribuições $t$ de Student com as mesmas médias e matrizes de variâncias-covariâncias. Uma vez que as distribuições $t$ de Student têm caudas mais pesadas do que a normal, é esperado que essas distribuições acomodem melhor pontos aberrantes.

A seguir, apresentamos uma abordagem em MMLs em que é assumida a distribuição $t$ de Student tanto para os efeitos aleatórios quanto para os erros intraindivíduos.

Seja $\boldsymbol{y}_i = (y_{i1}, ..., y_{im_i})^T$ as $m_i$ respostas medidas no indivíduo $i$, $i = 1, ..., n$. Um MML usual é dado por

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{3.16}$$

$$\boldsymbol{b}_i \overset{\text{iid}}{\sim} \boldsymbol{N}_q(\boldsymbol{0}, \boldsymbol{D}),\ \boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} \boldsymbol{N}_{m_i}(\boldsymbol{0}, \boldsymbol{R}_i), \tag{3.17}$$

em que $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ são efeitos fixos, $\boldsymbol{b}_i = (b_{i1}, ..., b_{iq})^T$ são efeitos aleatórios, $\boldsymbol{X}_i$ e $\boldsymbol{Z}_i$ são matrizes de planejamento conhecidas, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{im_i})^T$ são erros intraindivíduos, $\boldsymbol{R}_i$ é a matriz de variâncias-covariâncias para os erros intraindivíduos e $\boldsymbol{D}$ é a matriz de variâncias-covariâncias dos efeitos aleatórios.

A versão do modelo (3.16)-(3.17) sob erros $t$ de Student (Wu, 2010) é dada por

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{3.18}$$

$$\boldsymbol{b}_i \overset{\text{iid}}{\sim} \boldsymbol{t}_q(\boldsymbol{0}, \boldsymbol{D}, \nu_i),\ \boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} \boldsymbol{t}_{m_i}(\boldsymbol{0}, \boldsymbol{R}_i, \nu_i), \tag{3.19}$$

em que $\nu_i$ denota os graus de liberdade.

Em estudos longitudinais, dados atípicos podem ocorrer no nível da população, o que sugere uma distribuição $t$ de Student para os efeitos aleatórios para acomodar estes dados, e/ou podem ocorrem no nível do indivíduo, podendo ser sugerida uma distribuição $t$ de Student para os erros intraindivíduos para acomodar valores discrepantes. Em outras palavras, em modelos lineares de efeitos mistos robustos, podemos considerar distribuições $t$ de Student tanto para os efeitos aleatórios quanto para os erros de cada indivíduo.

### 3.7.2 Verificação da qualidade do ajuste

Em modelos com erros de medição, a qualidade do ajuste tem recebido muito menos atenção na literatura do que a inferência. Como em de Castro e Galea (2010), e similarmente ao caso normal, podemos utilizar a distância de Mahalanobis transformada para avaliar a adequação do modelo $t$ de Student multivariado ajustado. Temos que a quantidade $\vartheta_i = \delta_i/2m_i$, sendo $\delta_i$ a distância de Mahalanobis, segue distribuição $F_{(2m_i,\nu)}$. Além disso, $\hat{\vartheta}_i = \hat{\delta}_i/2m_i$ tem a mesma distribuição assintótica de $\vartheta_i$ (Box e Tiao, 1973).

De forma análoga ao caso normal, após a aplicação da transformação de Wilson-Hilferty (Johnson *et al.*, 1994) obtemos

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

que tem, aproximadamente, distribuição normal padrão, $d_i^{[t]} \overset{\text{iid}}{\sim} \boldsymbol{N}(0, 1)$, $i = 1, ..., n$. Gráficos normais de probabilidade das distâncias transformadas $d_i^{[t]}$ podem ser utilizados para avaliar a qualidade do ajuste do modelo $t$ de Student multivariado.

# Capítulo 4

# Diagnóstico de influência

## 4.1 Introdução

A detecção de dados atípicos (aberrantes, alavanca ou influentes) e a verificação de possíveis afastamentos das suposições estabelecidas sobre o modelo são etapas importantes em qualquer análise estatística. Isto é essencial para avaliar a sensibilidade dos resultados obtidos com o conjunto de dados disponível, já que observações atípicas podem distorcer as estimativas dos parâmetros, conduzindo em alguns casos a decisões errôneas.

Existem várias alternativas para avaliar a influência de perturbações nos dados e/ou nos pressupostos do modelo sobre as estimativas dos parâmetros de interesse (vide, por exemplo, Cook e Weisberg (1982) e Galea *et al.* (2000)). A eliminação de casos é uma técnica de diagnóstico comum para avaliar o efeito de uma observação sobre o processo de estimação e teste de hipóteses. Esta é uma análise de influência global, já que o efeito da observação é quantificado eliminando-a do conjunto de dados (Cook, 1977).

Alternativamente, Cook (1986) propôs um interessante método, denominado influência local, para avaliar o efeito de pequenas perturbações nos dados e/ou nos pressupostos do modelo estatístico, sobre as estimativas de máxima verossimilhança, sem eliminar observações. Cook propôs usar a curvatura normal da superfície do afastamento pela verossimilhança que é essencialmente equivalente a usar a segunda derivada do afastamento pela verossimilhança. O método foi aplicado por Galea *et al.* (1997) em modelos lineares elípticos. Resultados adicionais sobre influência local e aplicações podem ser encontrados em Escobar e Meeker (1992), Zhao e Lee (1998), Lesaffre e Verbeke (1998), Osorio *et al.* (2007) e Ibacache-Pulgar *et al.* (2012), entre outros.

O desenvolvimento do método de influência local no contexto de modelos com efeitos mistos e dados com estrutura longitudinal pode ser encontrado nos trabalhos de Osorio

(2006), que estudou o modelo linear com efeito misto elíptico, e Osorio *et al.* (2007) que estudaram modelos lineares elípticos com estrutura longitudinal, entre outros.

Já no contexto de modelos com erros nas variáveis o método de influência local tem sido estudado por diversos autores, entre eles Zhao e Lee (1995), que derivaram funções de influência para modelos lineares e não lineares generalizados com erros de medição; e Zhong *et al.* (2000), que desenvolveram diagnósticos de influência local e global para modelos lineares com erros nas variáveis baseados na função de verossimilhança corrigida proposta por Nakamura (1990).

No estudo de diagnósticos de influência, um enfoque corresponde à acomodação das observações discrepantes ou influentes utilizando distribuições simétricas com caudas mais pesadas do que a distribuição normal. Neste sentido, uma escolha interesante corresponde à classe de distribuições de contornos elípticos. O principal atrativo desta classe é que permite estender os modelos desenvolvidos sob suposição de erro normal considerando distribuições simétricas com caudas mais leves ou mais pesadas do que a normal (Osorio, 2006).

## 4.2 Influência local

Vamos considerar o logaritmo da função de verossimilhança de um modelo elíptico, dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}), \tag{4.1}$$

em que $L_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log \mathrm{g}(\delta_i)$ é a contribuição da $i$-ésima observação.

Suponhamos que $L_i(\boldsymbol{\theta}|\boldsymbol{\omega})$ seja o logaritmo da função de verossimilhança perturbada, que depende do vetor de perturbações $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^T$, restrito ao subconjunto euclidiano aberto $\boldsymbol{\Omega} \in \Re^n$, e assumimos que exista um vetor $\boldsymbol{\omega}_0$ de não perturbação que satisfaça $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$. Vamos supor também que $\widehat{\boldsymbol{\theta}}$ seja a estimativa de máxima verossimilhança obtida ao maximizar $L(\boldsymbol{\theta})$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ a estimativa de máxima verossimilhança obtida ao maximizar $L(\boldsymbol{\theta}|\boldsymbol{\omega})$. Como alternativa para comparar $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$, Cook (1986) propõe medir a distância entre as estimativas, relativas aos contornos do logaritmo da função de verossimilhança não perturbada $L(\boldsymbol{\theta})$, por meio da função de afastamento da verossimilhança, definida como

$$LD(\boldsymbol{\omega}) = 2\left[L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}|\boldsymbol{\omega})\right] \geq 0.$$

A ideia da influência local é estudar o comportamento de $LD(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0$. Deve-se escolher uma direção unitária arbitrária, $\boldsymbol{\ell}$ ($||\boldsymbol{\ell}|| = 1$), e considerar o gráfico da linha

projetada $LD(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ *versus* $a$, para $a \in \Re$. Note que $LD(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ tem um mínimo local em $a = 0$, visto que $LD(\boldsymbol{\omega}_0) = 0$. Cada linha projetada pode ser caracterizada por meio da curvatura normal, chamada $C_\ell(\boldsymbol{\theta})$ em torno de $a = 0$. Valores grandes da curvatura $C_\ell(\boldsymbol{\theta})$ indicam sensibilidade ao esquema de perturbação na direção $\boldsymbol{\ell}$. Também $C_\ell(\boldsymbol{\theta})$ é chamada influência local sobre a estimativa de $\boldsymbol{\theta}$, do esquema de perturbação, na direção $\boldsymbol{\ell}$. Cook (1986) mostra que a curvatura normal na direção $\boldsymbol{\ell}$ é dada por

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

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

$$\ddot{\mathbf{L}} = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\bigg|_{\theta=\widehat{\theta}} \quad \text{e} \quad \boldsymbol{\Delta} = \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T}\bigg|_{\theta=\widehat{\theta},\ \omega=\omega_0}, \tag{4.2}$$

em que $-\ddot{\mathbf{L}} = -\ddot{\mathbf{L}}(\boldsymbol{\theta})$ é a matriz de informação observada e $\boldsymbol{\Delta}$ é a matriz de perturbação, sendo avaliada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. $C_\ell(\boldsymbol{\theta})$ representa a curvatura normal sob a estimativa de $\boldsymbol{\theta}$ após perturbar o modelo $L(\boldsymbol{\theta})$. É possível que valores elevados da curvatura $C_\ell(\boldsymbol{\theta})$ indiquem a presença de alta sensibilidade na estimativa induzida pelas perturbações na direção $\boldsymbol{\ell}$.

Poon e Poon (1999) propuseram usar a curvatura normal conformal (curvatura invariante sob transformações uniformes de escala) definida por

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

em que $||\cdot||_F$ denota a norma de Frobenius definida por $||\boldsymbol{A}||_F = \{\text{tr}(\boldsymbol{A}^T\boldsymbol{A})\}^{1/2}$ para uma matriz $\boldsymbol{A}$. A característica dessa curvatura é permitir que, para qualquer direção $\boldsymbol{\ell}$ verifica-se $0 \leq B_\ell(\boldsymbol{\theta}) \leq 1$. Isto permite, por exemplo, a comparação da curvatura entre diferentes modelos.

A partir das equações em (4.2) podemos avaliar a influência que pequenas perturbações podem exercer sobre as estimativas dos parâmetros e sobre os resultados inferenciais, por exemplo. A direção $\boldsymbol{\ell}_{max}$, chamada de direção de máxima curvatura normal, $C_{max}$, é o autovetor normalizado correspondente ao maior autovalor $C_{max}$ da matriz $\boldsymbol{B} = \boldsymbol{\Delta}^T \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta}$. Existe interesse em avaliar a direção que produz a maior influência local e utilizando $\boldsymbol{\ell}_{max}$ podemos identificar as maiores mudanças locais no afastamento da verossimilhança para o esquema de perturbação utilizado.

Uma forma de identificar alguma influência substancial nos resultados é considerar o gráfico de índices da direção $\boldsymbol{\ell}_{max}$. Se, por exemplo, o $i$-ésimo componente de $\boldsymbol{\ell}_{max}$ é relativamente grande, isso indica que modificações do peso $\omega_i$ podem levar a mudanças substanciais

nos resultados da análise.

Escobar e Meeker (1992) propuseram estudar a curvatura normal na direção $\boldsymbol{\ell} = \boldsymbol{e}_{i,n}$, em que $\boldsymbol{e}_{i,n}$ é um vetor $(n \times 1)$ cujo $i$-ésimo elemento é igual a 1 e os demais elementos iguais a zero. De acordo com os autores, essa curvatura é dada por

$$C_i = 2|\boldsymbol{e}_{i,n}^T \boldsymbol{B} \boldsymbol{e}_{i,n}| = 2|b_{ii}|,$$

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

## 4.3 Derivação da curvatura

A seguir calculamos a matriz de informação observada, $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, e a matriz de perturbação, $\boldsymbol{\Delta}$, para diferentes esquemas de perturbação, para os casos normal e elíptico.

### 4.3.1 Matriz de informação observada

As matrizes de informação observada associadas aos MMLs normal e elíptico com erros nas variáveis assumem, ambas, a forma

$$-\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \ddot{\mathbf{L}}_i(\boldsymbol{\theta}), \tag{4.3}$$

em que

$$\begin{aligned}
\ddot{\mathbf{L}}_i(\boldsymbol{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\theta=\widehat{\theta}} \\
&= \begin{pmatrix}
\ddot{\mathbf{L}}_i^{\gamma\gamma} & \ddot{\mathbf{L}}_i^{\gamma\beta} & \ddot{\mathbf{L}}_i^{\gamma\mu} & \ddot{\mathbf{L}}_i^{\gamma\sigma_u^2} & \ddot{\mathbf{L}}_i^{\gamma\sigma^2} & \ddot{\mathbf{L}}_i^{\gamma\tau} \\
\ddot{\mathbf{L}}_i^{\beta\gamma} & \ddot{\mathbf{L}}_i^{\beta\beta} & \ddot{\mathbf{L}}_i^{\beta\mu} & \ddot{\mathbf{L}}_i^{\beta\sigma_u^2} & \ddot{\mathbf{L}}_i^{\beta\sigma^2} & \ddot{\mathbf{L}}_i^{\beta\tau} \\
\ddot{\mathbf{L}}_i^{\mu\gamma} & \ddot{\mathbf{L}}_i^{\mu\beta} & \ddot{\mathbf{L}}_i^{\mu\mu} & \ddot{\mathbf{L}}_i^{\mu\sigma_u^2} & \ddot{\mathbf{L}}_i^{\mu\sigma^2} & \ddot{\mathbf{L}}_i^{\mu\tau} \\
\ddot{\mathbf{L}}_i^{\sigma_u^2\gamma} & \ddot{\mathbf{L}}_i^{\sigma_u^2\beta} & \ddot{\mathbf{L}}_i^{\sigma_u^2\mu} & \ddot{\mathbf{L}}_i^{\sigma_u^2\sigma_u^2} & \ddot{\mathbf{L}}_i^{\sigma_u^2\sigma^2} & \ddot{\mathbf{L}}_i^{\sigma_u^2\tau} \\
\ddot{\mathbf{L}}_i^{\sigma^2\gamma} & \ddot{\mathbf{L}}_i^{\sigma^2\beta} & \ddot{\mathbf{L}}_i^{\sigma^2\mu} & \ddot{\mathbf{L}}_i^{\sigma^2\sigma_u^2} & \ddot{\mathbf{L}}_i^{\sigma^2\sigma^2} & \ddot{\mathbf{L}}_i^{\sigma^2\tau} \\
\ddot{\mathbf{L}}_i^{\tau\gamma} & \ddot{\mathbf{L}}_i^{\tau\beta} & \ddot{\mathbf{L}}_i^{\tau\mu} & \ddot{\mathbf{L}}_i^{\tau\sigma_u^2} & \ddot{\mathbf{L}}_i^{\tau\sigma^2} & \ddot{\mathbf{L}}_i^{\tau\tau}
\end{pmatrix}.
\end{aligned}$$

Devido à grande quantidade de elementos na matriz de informação, as expressões algébricas são apresentadas no Apêndice B.

### 4.3.2 Matriz de perturbação

O objetivo nesta seção é estudar três esquemas de perturbação bem conhecidas na literatura de influência local: ponderação de casos, perturbação na matriz de escala e perturbação na variável resposta, considerando a metodologia proposta por Zhu *et al.* (2007). Uma revisão sobre esses esquemas de perturbação pode ser encontrada em Zhu e Lee (2003), Osorio *et al.* (2007) e Zhu *et al.* (2007), entre outros.

A matriz de perturbação associada ao modelo misto linear elíptico com erros de medição assume a forma

$$\boldsymbol{\Delta} = \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \mid_{\theta=\widehat{\theta},\ \omega=\omega_0} = \begin{pmatrix} \Delta^{\gamma} \\ \boldsymbol{\Delta}^{\beta} \\ \Delta^{\mu} \\ \Delta^{\sigma_u^2} \\ \Delta^{\sigma^2} \\ \boldsymbol{\Delta}^{\tau} \end{pmatrix},$$

em que $\widehat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança e $\boldsymbol{\omega}_0$ é o vetor de não perturbação. A seguir são discutidos e apresentados os esquemas de perturbação e as expressões da matriz $\boldsymbol{\Delta}$.

#### Esquemas de perturbação

Para o desenvolvimento da abordagem de influência local é fundamental selecionar o esquema de perturbação de forma adequada, pois perturbar arbitrariamente um modelo pode levar à inferência enganosa sobre a causa de um efeito influente. Para a seleção da perturbação adequada será considerada a metodologia recentemente proposta por Zhu *et al.* (2007). O método desenvolvido estende a abordagem de Cook (1986) em vários aspectos. Por exemplo, é mostrado que o tensor métrico do espaço de perturbação fornece informações importantes sobre a seleção de uma perturbação adequada para um modelo. Recentemente, a abordagem de Zhu *et al.* (2007) para a avaliação da influência local foi aplicada por Shi *et al.* (2009) para os modelos lineares generalizados com covariáveis com dados incompletos, Chen *et al.* (2009) para modelos de equações estruturais não lineares e Chen *et al.* (2011) para

modelos lineares generalizados mistos. A aplicação desta metodologia em modelos de erros de medição tem sido pouco considerada na literatura. Uma recente aplicação foi apresentada por Giménez e Galea (2013) em modelos funcionais heteroscedásticos com erros de medição.

Suponha $\boldsymbol{\omega}$ um vetor de perturbações de dimensão $(n \times 1)$, que pertence a $\boldsymbol{\Omega}$, subconjunto de $\Re^n$, e o seguinte modelo estatístico perturbado:

$$M = \{f(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\omega}) : \boldsymbol{\omega} \in \boldsymbol{\Omega}\},$$

em que $f(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\omega})$ é a densidade de $\boldsymbol{W} = (\boldsymbol{W}_1, ..., \boldsymbol{W}_n)$, dada por

$$f(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\omega}) = \prod_{i=1}^{n} f_{W_i}(\boldsymbol{W}_i, \boldsymbol{\theta}, \omega_i), \tag{4.4}$$

perturbada por $\boldsymbol{\omega}$, e $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}|\omega_i)$ é sua correspondente função de log-verossimilhança. Denotando o vetor de não perturbação por $\boldsymbol{\omega}_0$, supomos que $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta})$, sendo que, no nosso caso, $L_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{V}_i| + \log \mathrm{g}(\delta_i)$, com $\delta_i = (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^T \boldsymbol{V}^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})$, $i = 1, ..., n$. Em geral, a função densidade do modelo estatístico perturbado $M$ é dada por

$$f(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\omega}) = \exp\{L(\boldsymbol{\theta}|\boldsymbol{\omega})\}[\mathrm{c}(\boldsymbol{\omega}; \boldsymbol{\theta})]^{-1}, \tag{4.5}$$

em que $\mathrm{c}(\boldsymbol{\omega}; \boldsymbol{\theta}) = \int \exp\{L(\boldsymbol{\theta}|\boldsymbol{\omega})\}\mathrm{d}\boldsymbol{W}$, e $f(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\omega}_0) = f(\boldsymbol{W}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{W_i}(\boldsymbol{W}_i, \boldsymbol{\theta})$.

Seja $\boldsymbol{U}(\boldsymbol{\omega}) = \frac{\partial L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}$ a função escore para $\boldsymbol{\omega}$ no modelo perturbado e $\mathrm{G}(\boldsymbol{\omega}) = \mathbf{E}_\omega\{\boldsymbol{U}(\boldsymbol{\omega}) \boldsymbol{U}^T(\boldsymbol{\omega})\}$ uma matriz do tensor métrico, que é a matriz de informação de Fisher com respeito ao vetor de perturbação $\boldsymbol{\omega}$, em que $\mathbf{E}_\omega$ denota a esperança com respeito a $f(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\omega})$, dada em (4.4). Segundo Zhu *et al.* (2007), uma perturbação $\boldsymbol{\omega}$ é apropriada se satisfaz

$$\mathrm{G}(\boldsymbol{\omega}_0) = \mathrm{c}\boldsymbol{I}_n, \tag{4.6}$$

em que $\mathrm{c} > 0$.

Se $\mathrm{G}(\boldsymbol{\omega}_0) \neq \mathrm{c}\boldsymbol{I}_n$, podemos escolher uma nova parametrização definida por

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

tal que $\mathrm{G}(\tilde{\boldsymbol{\omega}})$, avaliada em $\boldsymbol{\omega}_0$ seja igual a $\mathrm{c}\boldsymbol{I}_n$.

**Ponderação de casos**

Este esquema de perturbação permite avaliar a contribuição individual de cada observação sobre o processo de estimação. Neste caso, as contribuições individuais recebem ponderações diferentes. Seja $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^T$, com $0 \leq \omega_i \leq 1$, sendo o vetor de perturbação, e $\boldsymbol{\omega}_0 = (1, ..., 1)^T$ sendo o vetor de não perturbação. Entrão, o logaritmo da função de verossimilhança do modelo perturbado fica dado por

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

em que $L_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{V}_i| + \log \mathrm{g}(\delta_i)$ denota a contribuição individual da $i$-ésima observação no logaritmo da função de verossimilhança.

Neste caso, Zhu *et al.* (2007) mostraram que $\mathrm{G}(\boldsymbol{\omega}) = \mathrm{Diag}\{Var_\omega(L_1(\boldsymbol{\theta})), ..., Var_\omega(L_n(\boldsymbol{\theta}))\}$, em geral, não é da forma $\mathrm{G}(\boldsymbol{\omega}_0) = \mathrm{c}\boldsymbol{I}_n$, a menos que $m_1 = m_2 = \cdots = m_n$. Assim, os mesmos autores propõem

$$\tilde{\omega}_i = 1 + \sqrt{Var_{\omega_0}(L_i(\boldsymbol{\theta}))}(\omega_i - 1) \tag{4.7}$$

como perturbação apropriada para o esquema de ponderação de casos.

Portanto, o logaritmo da função de verossimilhança perturbado fica dado por

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

Como o esquema apropriado depende de $Var_{\omega_0}(L_i(\boldsymbol{\theta}))$, que depende da função de verossimilhança de cada modelo, apresentamos na sequência os esquemas de perturbação apropriados para os modelos normal e $t$ de Student.

**Caso normal**

Para a ponderação de casos no caso normal, temos que

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

pois, $\delta_i$ tem distribuição qui-quadrado com $2m_i$ graus de liberdade. Assim, o esquema de perturbação fica dado por

$$\tilde{\omega}_i = 1 + \sqrt{m_i}(\omega_i - 1),$$

e os elementos da matriz de perturbação ficam dados por

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

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

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

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

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

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

em que $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ e $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$.

### Caso $t$ de Student

Para a ponderação de casos no caso $t$ de Student, temos que

$$Var_{\omega_0}(L_i(\boldsymbol{\theta})) \;=\; \left(\frac{2m_i+\nu}{2}\right)^2 Var\left[\log(1+\nu^{-1}\delta_i)\right].$$

Considerando o Lema 6 apresentado em Arellano-Valle (2010), segue que

$$Var\left[\log(1+\nu^{-1}\delta_i)\right] \;=\; \Phi'\left(\frac{\nu}{2}\right) - \Phi'\left(\frac{\nu+2m_i}{2}\right),$$

em que $\Phi'(\cdot)$ é a função trigama.

Portanto, o esquema de perturbação fica dado por

$$\tilde{\omega}_i = 1 + \frac{2m_i + \nu}{2}\sqrt{\Phi'\left(\frac{\nu}{2}\right) - \Phi'\left(\frac{\nu + 2m_i}{2}\right)}(\omega_i - 1),$$

e os elementos da matriz de perturbação ficam dados por

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

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

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

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

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

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

em que $p_i = \frac{2m_i+\nu}{2}\sqrt{\Phi'\left(\frac{\nu}{2}\right) - \Phi'\left(\frac{\nu+2m_i}{2}\right)}$, $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ e $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$.

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

### Perturbação na matriz de escala

Seja $\boldsymbol{V}_i$ a matriz de escala do modelo misto linear elíptico com erros de medição. Para este esquema de perturbação vamos assumir $\omega_i^{-1}\boldsymbol{V}_i$ no lugar de $\boldsymbol{V}_i$, sendo $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^T$, com $\omega_i > 0$, o vetor de perturbação, e $\boldsymbol{\omega}_0 = (1, ..., 1)^T$ sendo o vetor de não perturbação. Assim, o logaritmo da função de verossimilhança fica dado por

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

em que

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

com $\delta_{i\omega} = \omega_i\delta_i = (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^T\omega_i\boldsymbol{V}_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})$.

Seguindo a metodologia de Zhu *et al.* (2007), temos que

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

e

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

Logo,

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

e, portanto, $G(\boldsymbol{\omega}) = \mathrm{Diag}\{\mathrm{g}_{11}(\boldsymbol{\omega}), ..., \mathrm{g}_{nm}(\boldsymbol{\omega})\}$.

Agora, temos

$$\begin{aligned} \mathrm{g}_{ij}(\boldsymbol{\omega}) &= \mathbf{E}_\omega\left\{\left(\frac{m_i}{\omega_i}\right)^2 + \frac{2m_i}{\omega_i}W_g(\delta_{i\omega})\delta_i + W_g^2(\delta_{i\omega})\delta_i^2\right\} \\ &= \left(\frac{m_i}{\omega_i}\right)^2 + \frac{2m_i}{\omega_i}\,\mathbf{E}_\omega\left[W_g(\delta_{i\omega})\delta_i\right] + \mathbf{E}_\omega\left[W_g^2(\delta_{i\omega})\delta_i^2\right] \\ &= \left(\frac{m_i}{\omega_i}\right)^2 + \frac{2m_i}{\omega_i}\left(-\frac{m_i}{2\omega_i}\right) + \frac{f_{g_i}}{\omega^2} \\ &= \frac{f_{g_i}}{\omega_i^2}, \end{aligned}$$

pois

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

e

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

Para $i = 1, ..., n$, temos que

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

que não é da forma (4.6).

Como $\mathrm{G}(\boldsymbol{\omega}_0) \neq c\boldsymbol{I}_n$, usamos a parametrização proposta por Zhu *et al.* (2007), em que $\tilde{\boldsymbol{\omega}} = \mathbf{1}_{m_i} + \mathrm{Diag}\{\sqrt{f_{g_1}}, ..., \sqrt{f_{g_n}}\}(\boldsymbol{\omega} - \mathbf{1}_{m_i})$.

Portanto, uma perturbação adequada para o esquema de perturbação na matriz de escala fica dada por

$$\tilde{\omega_i} = 1 + \sqrt{f_{g_i}}(\omega_i - 1).$$

Para o modelo normal a quantidade $f_{g_i}$ assume a forma $f_{g_i} = m_i(m_i+1)$ e para o modelo $t$ de Student fica dada por $f_{g_i} = m_i(m_i+1)\frac{\nu+2m_i}{\nu+2m_i+2}$.

Diferenciando $L_i(\boldsymbol{\theta}|\boldsymbol{\omega})$ em relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}$, restrito a $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, obtemos

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\gamma\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = \sqrt{f_{g_i}}\left[W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right]\left[2\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T - \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right]\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -2\sqrt{f_{g_i}}\left[W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right]\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\mu\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -2\sqrt{f_{g_i}}\left[W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right]\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma_u^2\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\sqrt{f_{g_i}}\left[W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right]\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i;$$

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma^2\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\sqrt{f_{g_i}}\left[W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right]\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i \text{ e}$$

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\tau_j\partial\omega_i}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\sqrt{f_{g_i}}\left[W_g'(\hat{\delta}_i)\hat{\delta}_i + W_g(\hat{\delta}_i)\right]\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\tau_j}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,$$

em que $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ e $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$.

**Perturbação no vetor de respostas e na covariável medida com erros**

Este esquema de perturbação pode ser utilizado se o objetivo for avaliar a sensibilidade das estimativas quando são introduzidas pequenas perturbações nos componentes de cada vetor de respostas e na covariável longitudinal medida com erros: $\boldsymbol{W}_i = (\boldsymbol{y}_i^T, \boldsymbol{u}_i^T)^T$.

Sejam $\boldsymbol{\omega}_i = (\omega_{i1}, ..., \omega_{i2m_i})^T \in \Re^{2m_i}$ o vetor de perturbação e $\boldsymbol{\omega}_0 = \mathbf{0}$ o vetor $(2m_i \times 1)$ de não perturbação. Similarmente ao caso usual de perturbação na variável resposta (vide, por exemplo, Osorio et al., 2007), vamos considerar a seguinte perturbação no vetor de respostas observadas:

$$\boldsymbol{W}_{i\omega} = \boldsymbol{W}_i + \boldsymbol{V}_i^{1/2}\omega_i.$$

Assim, o logaritmo da função de verossimilhança do modelo (**??**) perturbado fica dado por

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

em que $\delta_{i\omega} = (\boldsymbol{W}_{i\omega} - \boldsymbol{\mu}_{iW})^T\boldsymbol{V}_i^{-1}(\boldsymbol{W}_{i\omega} - \boldsymbol{\mu}_{iW})$.

De acordo com a metodologia de Zhu *et al.* (2007), temos que

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

Logo,

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

em que $c_i = \frac{2d_{g_i}}{m_i}$, $i = 1, ..., n$.

Assim, temos que $\mathrm{G}(\boldsymbol{\omega}) = \mathrm{Diag}(c_1\boldsymbol{I}_{2m_1}, ..., c_n\boldsymbol{I}_{2m_n})$, que não é da forma (4.6).

Como $\mathrm{G}(\boldsymbol{\omega}) \neq c\boldsymbol{I}_n$, usamos a parametrização proposta por Zhu *et al.* (2007), em que

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

Portanto, uma perturbação adequada para o esquema de perturvação no vetor de respostas observadas é

$$\tilde{\omega_i} = \boldsymbol{V}_i^{1/2}\frac{\omega_i}{\sqrt{c_i}},$$

e o vetor de respostas observadas perturbado fica dado por

$$\boldsymbol{W}_{i\omega} = \boldsymbol{W}_i + \boldsymbol{V}_i^{1/2}\frac{\omega_i}{\sqrt{c_i}}, \quad i = 1, ..., n.$$

Para o modelo normal a quantidade $d_{g_i}$ assume a forma $d_{g_i} = \frac{m_i}{2}$ e para o modelo $t$ de Student fica dada por $d_{g_i} = \frac{m_i}{2}\frac{\nu+2m_i}{\nu+2m_i+2}$.

Diferenciando $L_i(\boldsymbol{\theta}|\boldsymbol{\omega})$ em relação a $\boldsymbol{\theta}$ e $\boldsymbol{\omega}$, restrito a $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, obtemos

$$\begin{aligned}
\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\gamma\partial\boldsymbol{\omega}_i^T}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= -\frac{2}{\sqrt{2d_{g_i}/m_i}}\left\{W_g(\hat{\delta}_i)\left[\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1/2}\right.\right.\\
&\quad\left.+\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\left(\frac{\partial\boldsymbol{V}_i^{1/2}}{\partial\gamma}-\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\right)\hat{\boldsymbol{V}}_i^{1/2}\right]+W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\\
&\quad\left.\times\left[\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right]\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1/2}\right\},
\end{aligned}$$

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}_i^T}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\left(\frac{2}{\sqrt{2d_{g_i}/m_i}}\right)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\left[W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}+2W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\right]\hat{\boldsymbol{V}}_i^{1/2}\hat{\boldsymbol{V}}_i^{-1},$$

$$\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\mu\partial\boldsymbol{\omega}_i^T}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = -\left(\frac{2}{\sqrt{2d_{g_i}/m_i}}\right)\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\left[W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}+2W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\right]\hat{\boldsymbol{V}}_i^{1/2}\hat{\boldsymbol{V}}_i^{-1},$$

$$\begin{aligned}
\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma_u^2\partial\boldsymbol{\omega}_i^T}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \frac{2}{\sqrt{2d_{g_i}/m_i}}\left\{W_g(\hat{\delta}_i)\left[\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i^{1/2}}{\partial\sigma_u^2}-\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1/2}\right]\right.\\
&\quad\left.-W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1/2}\right\},
\end{aligned}$$

$$\begin{aligned}
\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\sigma^2\partial\boldsymbol{\omega}_i^T}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \frac{2}{\sqrt{2d_{g_i}/m_i}}\left\{W_g(\hat{\delta}_i)\left[\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i^{1/2}}{\partial\sigma^2}-\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1/2}\right]\right.\\
&\quad\left.-W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1/2}\right\}
\end{aligned}$$

e

$$\begin{aligned}
\left.\frac{\partial^2 L_i(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\tau_j\partial\boldsymbol{\omega}_i^T}\right|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \frac{2}{\sqrt{2d_{g_i}/m_i}}\left\{W_g(\hat{\delta}_i)\left[\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i^{1/2}}{\partial\tau_j}-\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\tau_j}\hat{\boldsymbol{V}}_i^{-1/2}\right]\right.\\
&\quad\left.-W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\tau_j}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1/2}\right\},
\end{aligned}$$

em que $\hat{\delta}_i = \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i$ e $\hat{\boldsymbol{r}}_i = \boldsymbol{W}_i - \hat{\boldsymbol{\mu}}_{iW}$.

Podemos observar que as expressões acima dependem das matrizes $\boldsymbol{V}_i^{1/2}$ e $\frac{\partial\boldsymbol{V}_i^{1/2}}{\partial\boldsymbol{\theta}}$, que não são obtidas de imediato. Portanto, apresentamos a seguir os procedimentos para obtermos estas matrizes.

Para qualquer matriz $\boldsymbol{V}_i$ $(2m_i \times 2m_i)$ simétrica e não negativa definida, existe uma matriz simétrica não negativa definida $\boldsymbol{V}_i^{1/2} = \boldsymbol{T}_i$, tal que $\boldsymbol{V}_i = \boldsymbol{V}_i^{1/2}\boldsymbol{V}_i^{1/2} = \boldsymbol{T}_i^2$. Além disso, $\boldsymbol{T}_i$ é única e pode ser expressa por

$$\boldsymbol{T}_i = \boldsymbol{S}_i \boldsymbol{A}_i^{1/2} \boldsymbol{S}_i^T,$$

em que $\boldsymbol{A}_i^{1/2} = diag(\sqrt{\alpha_1}, ..., \sqrt{\alpha_{2m_i}})$, com $\alpha_1, ..., \alpha_{2m_i}$ sendo os autovalores de $\boldsymbol{V}_i$ e $\boldsymbol{S}_i$ é uma matriz $(2m_i \times 2m_i)$ ortogonal $(\boldsymbol{S}_i\boldsymbol{S}_i^T = \boldsymbol{I}_{2m_i})$ tal que $\boldsymbol{S}_i\boldsymbol{V}_i\boldsymbol{S}_i^T = \boldsymbol{A}_i$, com $\boldsymbol{A}_i = diag(\alpha_1, ..., \alpha_{2m_i})$. Então, a derivada de $\boldsymbol{V}_i$ com respeito ao escalar $\eta_j$ fica dada por

$$\frac{\partial \boldsymbol{V}_i}{\partial \eta_j} = \boldsymbol{T}_i \frac{\partial \boldsymbol{T}_i}{\partial \eta_j} + \frac{\partial \boldsymbol{T}_i}{\partial \eta_j} \boldsymbol{T}_i, \quad \text{para } j = 1, ..., q. \tag{4.8}$$

A equação acima pode ser escrita como $\dot{\boldsymbol{C}}_i = \boldsymbol{T}_i\dot{\boldsymbol{T}}_i + \dot{\boldsymbol{T}}_i\boldsymbol{T}_i$, em que $\dot{\boldsymbol{C}}_i = \frac{\partial \boldsymbol{V}_i}{\partial \eta_j}$ e $\dot{\boldsymbol{T}}_i = \frac{\partial \boldsymbol{T}_i}{\partial \eta_j}$, a qual tem sido extensivamente estudada na literatura (vide, por exemplo, Jameson, 1968). Note que $\dot{\boldsymbol{C}}_i$, $\dot{\boldsymbol{T}}_i$ e $\boldsymbol{T}_i$ são todas matrizes simétricas. Assim, sejam $\boldsymbol{G}_i = \boldsymbol{S}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{S}_i$ e $\boldsymbol{Q} = [(q_{rs})]$, matrizes simétricas $(2m_i \times 2m_i)$, com $q_{rs} = (\sqrt{\alpha_r}, ..., \sqrt{\alpha_s})^{-1}$, para $r, s = 1, ..., 2m_i$. Então, a solução para a equação (4.8) fica dada por

$$\frac{\partial \boldsymbol{T}_i}{\partial \eta_j} = \frac{\partial \boldsymbol{V}_i^{1/2}}{\partial \eta_j} = \boldsymbol{S}_i(\boldsymbol{G}_i \oplus \boldsymbol{Q})\boldsymbol{S}_i^T,$$

em que $\oplus$ denota o produto de Hadamard para $i = 1, ..., q$.

# Capítulo 5

# Aplicação

## 5.1 Introdução

Neste capítulo apresentamos uma aplicação dos modelos apresentados e discutidos nos capítulos anteriores. A aplicação é apresentada de forma a comparar aspectos de estimação e influência local nos modelos normal e $t$ de Student na abordagem proposta neste trabalho.

## 5.2 Aplicação: dados reduzidos de Boston analisados por Zhong *et al.* (2002)

Conforme descrito na Seção 1.2, Zhong *et al.* (2002) selecionaram dados de apenas 132 setores censitários de 15 distritos da cidade de Boston (ao todo são 506 setores censitários em 92 distritos). Os setores censitários dos distritos são tomados como medidas repetidas e, por isso, os autores ajustaram um modelo linear de efeitos mistos. Além disso, neste conjunto de dados todas a variáveis independentes (Tabela 1.1) podem ser medidas precisamente, com exceção da variável que mede a poluição (NOXSQ), a qual foi considerada com erros de medição.

Zhong *et al.* (2002) consideraram erros com distribuição normal. Em nosso caso, utilizaremos o mesmo subconjunto de dados, mas a abordagem será uma extensão do modelo (1.1) ajustado por esses autores, ajustando um modelo misto linear, com uma variável explicativa sujeita a erros de medição e supondo uma distribuição elíptica tanto para os efeitos aleatórios quanto para os erros aleatórios. Para efeito de comparação, utilizaremos as mesmas variáveis do modelo (1.1), proposto por esses autores.

### 5.2.1 Modelo proposto

Consideremos o seguinte modelo misto linear elíptico com erros de medição:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{b}_i\mathbf{1}_{m_i} + \gamma\boldsymbol{u}_i^* + \boldsymbol{\epsilon}_i, \tag{5.1}$$

$$\boldsymbol{u}_i = \boldsymbol{u}_i^* + \boldsymbol{e}_i, \tag{5.2}$$

em que $\boldsymbol{y}_i$ denota os valores observados do logaritmo do valor mediano das casas ocupadas pelos proprietários (LMV) no $i$-ésimo distrito, $\boldsymbol{X}_i$ denota a matriz de covariáveis fixas, $\boldsymbol{\beta}$ é o vetor de parâmetros associados aos coeficientes de regressão que determinam o incremento no valor do LMV, $\boldsymbol{b}_i$ denota o efeito aleatório do $i$-ésimo distrito, $\boldsymbol{u}_i$ é a variável observada NOXSQ sujeita a erros de medição, $\gamma$ é o parâmetro associado à variável verdadeira $\boldsymbol{u}_i^*$, $\boldsymbol{\epsilon}_i$ é o vetor de erros aleatórios do setor censitário, $\boldsymbol{e}_i$ é o vetor de erros associados à variável medida com erros e $\mathbf{1}_{m_i}$ denota o vetor de uns $(m_i \times 1)$, para $i = 1, ..., 15$.

Como em Zhong *et al.* (2002), é usual assumir que tanto os erros aleatórios, que são não correlacionados através dos setores, quanto os efeitos aleatórios e os erros de medição seguem distribuição normal. Entretanto, é sabido que as estimativas de máxima verossimilhança derivadas do modelo normal são sensíveis à observações aberrantes. Nesse caso, uma alternativa é assumir um modelo com caudas mais pesadas do que a normal para acomodar essas possíveis observações. Neste sentido, vamos supor que o vetor da resposta observada (LMV) e a variável observada NOXSQ, sujeita a erros de medição, seguem uma distribuição da forma

$$\boldsymbol{W}_i = \begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{u}_i \end{pmatrix} \overset{\text{ind}}{\sim} El_{2m_i}\left(\boldsymbol{\mu}_{iW}, \boldsymbol{V}_i\right), \tag{5.3}$$

em que:
$\boldsymbol{\mu}_{iW} = \begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} + \gamma\mu\mathbf{1}_{m_i} \\ \mu\mathbf{1}_{m_i} \end{pmatrix}$ e
$\boldsymbol{V}_i = \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + (\gamma^2\sigma_u^2 + \sigma^2)\boldsymbol{I}_{m_i} & \gamma\sigma_u^2\boldsymbol{I}_{m_i} \\ \gamma\sigma_u^2\boldsymbol{I}_{m_i} & (\sigma_u^2 + \sigma_e^2)\boldsymbol{I}_{m_i} \end{bmatrix}$, com $\boldsymbol{Z}_i$ sendo um vetor de $1's$ de ordem $(m_i \times 1)$ e $\boldsymbol{D} = \tau$, escalar.

Devido aos problemas de identificabilidade, discutidos nas Seções 2.2.1 e 3.2.1, vamos assumir que o parâmetro de escala associado aos erros de medição, $\sigma_e^2$, é conhecido e que a matriz de planejamento $\boldsymbol{X}_i^*$ possui posto completo (vide Seção 2.2.1), o que implica que devemos supor que o intercepto $\beta_0$ também é conhecido. Assim, o vetor de parâmetros a ser estimado é $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, \tau\right)^T$.

### 5.2.2 Ajustando os modelos normal e $t$ de Student

Os modelos foram ajustados usando o método de máxima verossimilhança sob erros normal e $t$ de Student. Para obter as estimativas dos parâmetros foi utilizado o método *BFGS* presente na função *optim* do *software* R (R Development Core Team, 2011). Como sugerido por Lange *et al.* (1989), para escolher os graus de liberdade da distribuição $t$ de Student usamos o critério de Akaike (AIC) (Akaike, 1974), segundo o qual devemos escolher, dentre os modelos considerados, aquele que apresente o menor valor de AIC, visto que, maximizar o logaritmo da função de verossimilhança equivale a maximizar o critério de Akaike. Portanto, segundo a Tabela 5.1, o número de graus de liberdade escolhido é $\nu = 5$. Foi assumido que o parâmetro de escala associado aos erros de medição é $\sigma_e^2 = 0,2$. Já a escolha do coeficiente $\beta_0$ foi de acordo com valores obtidos a partir de ajustes do modelo normal usual, indicando $\beta_0 = 9,0$. Além disso, para encontrarmos as estimativas do vetor de parâmetros $\boldsymbol{\theta}$ precisamos fornecer valores iniciais para o processo iterativo. Tais valores também foram obtidos a partir de ajustes do modelo normal usual.

Tabela 5.1: Valores do critério de informação de Akaike (AIC) sob o modelo $t$ de Student para diferente graus de liberdade $\nu$.

| $\nu$ | *AIC* |
|---|---|
| 1 | 890,73 |
| 2 | 883,31 |
| 3 | 880,88 |
| 4 | 880,03 |
| **5** | **879,83** |
| 6 | 879,95 |
| Normal | 891,10 |

Os erros padrão dos estimadores dos coeficientes de regressão e dos parâmetros de escala foram calculados a partir da matriz de informação de Fisher, tanto para o modelo normal quanto para o modelo $t$ de Student, e os resultados dos ajustes são apresentados na Tabela 5.3. Para podermos comparar os resultados obtidos por Zhong *et al.* (2002) por meio da função de escore corrigida (CSFE), na Tabela 5.2 são apresentadas as estimativas por eles obtidas.

Tabela 5.2: Estimativas obtidas por Zhong *et al.* (2002) (CSFE).

| Parâmetro | Estimativa (CSFE) | Erro padrão | Valor Z |
|---|---|---|---|
| Intercepto ($\beta_0$) | 9,1400 | 0,5897 | 15,5000 |
| Room ($\beta_1$) | -0,0023 | 0,0047 | -0,4900 |
| Age ($\beta_2$) | 0,0011 | 0,0035 | 0,3100 |
| Dist ($\beta_3$) | 0,0014 | 2,8000 | 0,0005 |
| Black ($\beta_4$) | 0,3460 | 0,2932 | 1,1800 |
| Lstat ($\beta_5$) | -0,5750 | 0,1132 | -5,0800 |
| Crim ($\beta_6$) | -0,0076 | 0,0024 | -3,1200 |
| Chas ($\beta_7$) | 0,0022 | 0,1100 | 0,0200 |
| Noxsq ($\gamma$) | -0,0118 | 0,0084 | -1,4000 |
| $\sigma^2$ | 0,0650 | | |
| $\tau$ | 0,0392 | | |

Tabela 5.3: Estimativas de máxima verossimilhança, erros padrão aproximados e valores $Z$ para os modelos normal e $t$ de Student ajustados aos dados dos setores censitários de Boston reduzidos.

| | Normal | | | $t$ de Student | | |
|---|---|---|---|---|---|---|
| Parâmetro | Estimativa | Erro padrão | Valor-Z | Estimativa | Erro padrão | Valor-Z |
| Room ($\beta_1$) | -0,0010 | 0,0099 | -0,1023 | -0,0021 | 0,0101 | -0,2105 |
| Age ($\beta_2$) | 0,0009 | 0,0043 | 0,2113 | 0,0013 | 0,0043 | 0,2988 |
| Dist ($\beta_3$) | 0,0786 | 0,4647 | 0,1692 | 0,1712 | 0,4517 | 0,3791 |
| Black ($\beta_4$) | 0,4503 | 0,5947 | 0,7572 | 0,4513 | 0,5993 | 0,7531 |
| Lstat ($\beta_5$) | -0,5427 | 0,2211 | -2,4547 | -0,5668 | 0,2229 | -2,5432 |
| Crim ($\beta_6$) | -0,0072 | 0,0053 | -1,3602 | -0,0065 | 0,0054 | -1,2061 |
| Chas ($\beta_7$) | -0,0352 | 0,3469 | -0,1015 | -0,0225 | 0,3572 | -0,0631 |
| Noxsq ($\gamma$) | -0,0097 | 0,0017 | -5,7188 | -0,0124 | 0,0014 | -8,6419 |
| $\mu$ | 45,5838 | 2,6904 | 16,9430 | 47,4301 | 2,4409 | 19,4315 |
| $\sigma_u^2$ | 63,4978 | 30,3667 | | 48,4985 | 41,4285 | |
| $\sigma^2$ | 0,0290 | 0,0147 | | 0,0284 | 0,0249 | |
| $\tau$ | 0,0464 | 0,0723 | | 0,0319 | 0,0580 | |
| $L(\widehat{\boldsymbol{\theta}})$ | -434,1991 | | | -427,9173 | | |
| $AIC$ | 892,3982 | | | 879,8346 | | |

Pelas estimativas da Tabela 5.2 nota-se que apenas as variáveis explicativas LSTAT e CRIM são marginalmente significativas. Isso não quer dizer que as demais variáveis explicativas devam ser removidas do modelo. Procedimentos de seleção de modelos devem ser aplicados a fim de avaliar quais variáveis explicativas devem ser mantidas no modelo. Contudo, olhando apenas as variáveis marginalmente significativas, nota-se que o valor esperado para o logaritmo do valor mediano das casas ocupadas deve crescer com a diminuição da taxa de criminalidade (mantendo-se as demais variáveis fixas). Mesma tendência deve ocorrer à medida que o logaritmo da proporção da população de baixa renda diminuir. Nota-se também que a concentração de óxido de nitrogênio (variável medida com erros) não é significativa marginalmente. Testes para avaliar $H_0 : \tau = 0$ contra $H_1 : \tau > 0$ não foram aplicados, assim não podemos afirmar se há necessidade de incorporar efeitos aleatórios no modelo proposto por Zhong *et al.* (2002) para contemplar no modelo a correlação intraunidades

experimentais (intradistritos).

Pelas estimativas apresentadas na Tabela 5.3 nota-se que para ambos os modelos (com erros normais e com erros $t$ de Student com $\nu = 5$ graus de liberdade) apenas as variáveis LSTAT e NOXSQ são marginalmente significativas. Com relação à variável explicativa LSTAT, tem-se a mesma interpretação do modelo ajustado por Zhong *et al.* (2002). As estimativas pontuais são muito parecidas, porém os erros padrão aproximados obtidos pelo método da função escore corrigido são menores do que pelo método de máxima verossimilhança. Para a variável NOXSQ nota-se tanto sob erros normais como também $t$ de Student que à medida que aumenta a concentração de óxido de nitrogênio, diminui o logaritmo do valor mediano das casas ocupadas.

Um importante teste a ser realizado é o teste para a significância da variável medida com erros, ou seja, pode-se testar $H_0 : \gamma = 0$ *versus* $H_1 : \gamma \neq 0$. Sob a hipótese nula, o modelo (5.1) reduz-se a um modelo misto linear com com uma covariável aleatória, visto que a covariável observada que se supunha medida com erros continua sendo aleatória no modelo. Para testar as hipóteses acima foi aplicado o teste do escore ($S$) discutido na Seção 2.7, resultando em $S = 83,73$ (valor-$p < 0,001$) para o modelo normal e $S = 97,02$ (valor-$p < 0,001$) para o modelo $t$ de Student. Assim, conclui-se que foi significância a inclusão da variável medida com erros em ambos os modelos, sendo que a estatística do teste teve maior valor para o modelo $t$ de Student.

Com relação ao parâmetro $\tau$, que mede a variância do efeito aleatório, podemos aplicar um teste apropriado para testar $H_0 : \tau = 0$ contra $H_1 : \tau > 0$. Conforme discutido em Savalli *et al.* (2006) a distribuição nula assintótica de estatísticas apropriadas para hipóteses do tipo acima em modelos mistos lineares elípticos segue assintoticamente uma distribuição

$$\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2,$$

em que $\chi_0^2$ denota a distribuição degenerada na origem. Apresentamos abaixo um resumo da aplicação do teste da razão de verossimilhanças para testar as hipóteses com relação a $\tau$.

Tabela 5.4: Razão de verossimilhanças (RV) e valor-$p$ para testar hipóteses sobre o parâmetro $\tau$ sob os modelos normal e $t$ de Student.

| Hipóteses | RV-Normal | Valor-$p$ | RV-$t$ de Student | Valor-$p$ |
|---|---|---|---|---|
| $H_0 : \tau = 0$ | | | | |
| $H_1 : \tau > 0$ | 6,1922 | $< 0,001$ | 4,5342 | $< 0,001$ |

Pelos resultados apresentados na Tabela 5.4 nota-se que o parâmetro $\tau$ é significativo, indicando que é importante considerar o efeito aleatório tanto no ajuste do modelo normal quanto no $t$ de Student.

Para avaliar os ajustes dos modelos normal e $t$ de Student foram construídos os gráficos das distâncias transformadas, apresentadas e discutidas nas Seções 2.8 e 3.7.2. Observando os valores de $L(\widehat{\boldsymbol{\theta}})$ e $AIC$ na Tabela 5.3 e os gráficos normais de probabilidade das distâncias transformadas da Figura 5.1 temos a indicação de que o modelo $t$ de Student com $\nu = 5$ graus de liberdade apresenta um ajuste mais adequado em relação ao normal. Nota-se uma excelente concordância entre os valores observados sob o modelo $t$ de Student, enquanto que sob o modelo normal observa-se alguns afastamentos para valores baixos da distância transformada.

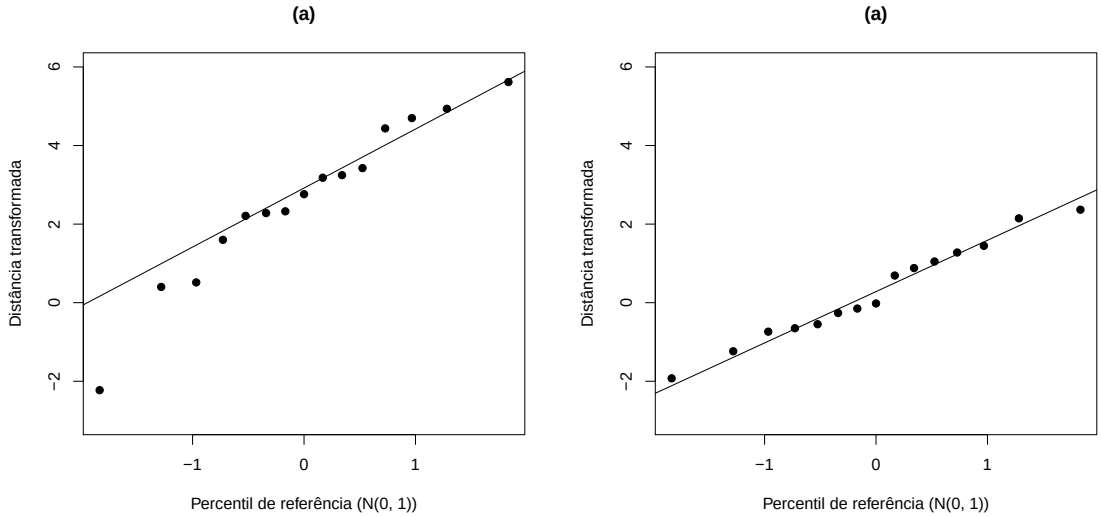


Figura 5.1: Gráficos normais de probabilidades para as distâncias transformadas sob os modelos normal (a) e $t$ de Student (b) ajustados aos dados dos setores censitários de Boston reduzidos.

### 5.2.3 Diagnóstico de influência

Nesta seção apresentamos alguns gráficos de medidas de influência local para os esquemas de perturbação ponderação de casos, perturbação na matriz de escala e perturbação na resposta observada, utilizando a metodologia de Zhu *et al.* (2007). O objetivo é detectar observações influentes e avaliar a sensibilidade das estimativas de máxima verossimilhança para os modelos ajustados.

### Ponderação de casos

Utilizamos a metodologia proposta por Zhu *et al.* (2007) para encontrar o esquema de ponderação de casos adequado para o modelo proposto. O esquema encontrado apresenta diferenças quando comparado com o esquema de ponderação de casos usual para casos desbalanceados (vide Seção 4.3.2), visto que o tamanho do grupo interfere no esquema de perturbação.

Para verificarmos na prática como essa diferença afeta a análise de sensibilidade, apresentamos na Figura 5.2 os gráficos de índices de $|\ell_{max}|$ para os modelos normal, considerando a perturbação usual (Figura 5.2a) e a proposta por Zhu *et al.* (2007) (Figura 5.2b), e $t$ de Student com 5 graus de liberdade, em que foram atribuídas diferentes ponderações às observações. Observamos que, além da configuração dos índices $|\ell_{max}|$ ser diferente, sob o modelo normal e considerando a perturbação usual, o gráfico de influência indica que os distritos #1 e #10 são possivelmente influentes nas estimativas de máxima verossimilhança; considerando a perturbação de Zhu *et al.* (2007), indica os distritos #1, #9 e #10. Já sob o modelo $t$ de Student, o gráfico de influência indica que os distritos #9 e #10 aparecem com menos destaque que no caso normal e há a indicação de que os distritos #6 e #7 são possivelmente influentes.

### Perturbação na matriz de escala

Na Figura 5.3 são apresentados os gráficos de índices de $|\ell_{max}|$ para os modelos normal e $t$ de Student com 5 graus de liberdade. Sob o modelo normal, o gráfico de influência indica que os distritos #1, #9 e #10 são possivelmente influentes nas estimativas de máxima verossimilhança. Já sob o modelo $t$ de Student, o gráfico de influência indica que os distritos #6, #7, #10 e #11 são possivelmente influentes. Em ambos os modelos, o distrito #10 aparece com destaque.

### Perturbação no vetor de respostas observadas

Os gráficos de índices de $|\ell_{max}|$, apresentados na Figura 5.4, indicam que os setores censitários #148, #148 e #152 podem ser influentes nas estimativas de máxima verossimilhança para o modelo normal. Quando ajustado o modelo $t$ de Student com 5 graus de liberdade, estes mesmo setores não são indicados como influentes, porém, os índices de $|\ell_{max}|$ identificam como possivelmente influentes os setores censitários #186, #190 e #200.

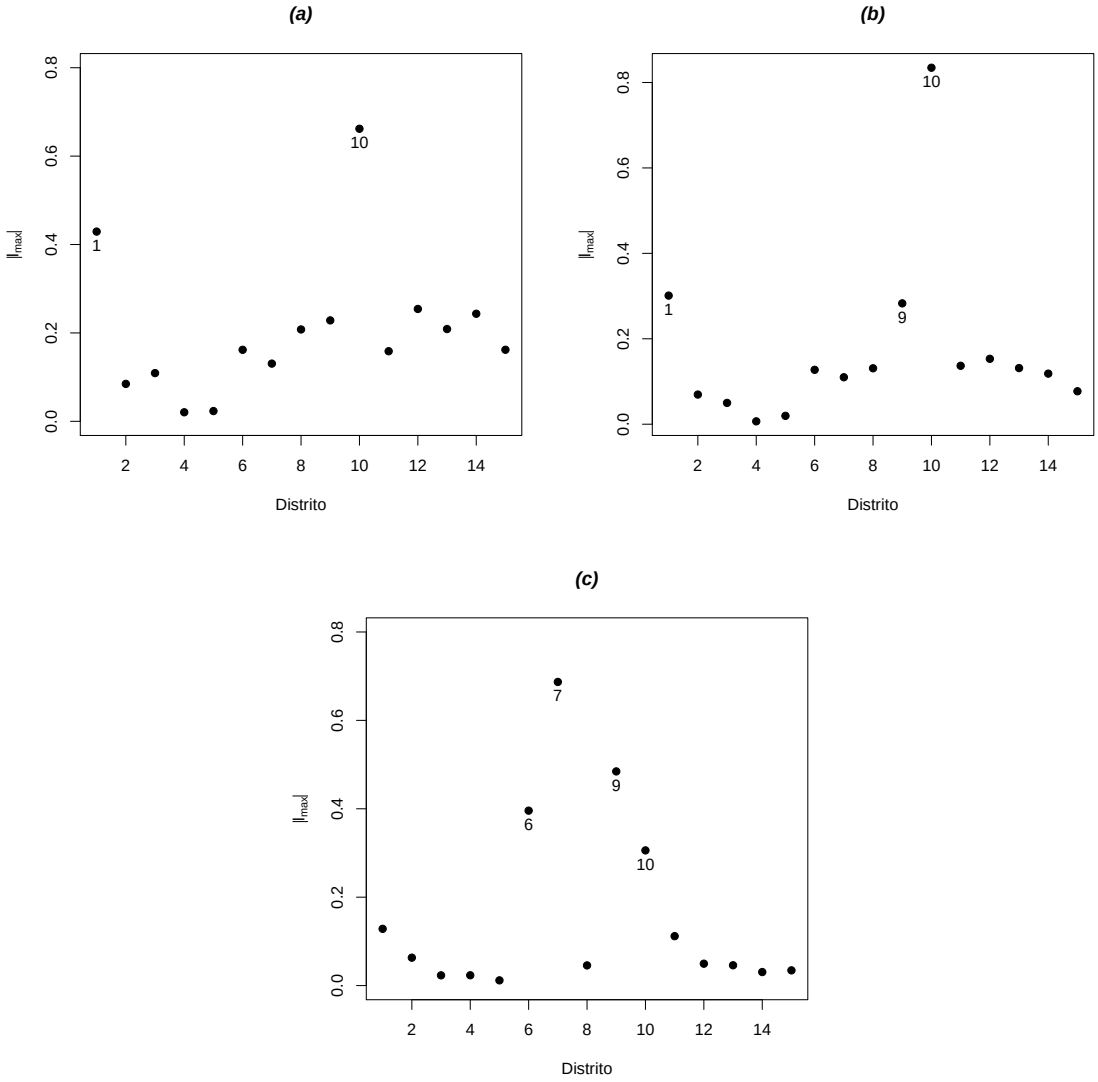


Figura 5.2: Gráficos dos índices de $|\ell_{max}|$ sob os modelos normal com perturbação usual (a) e perturbação de Zhu *et al.* (2007) (b) e $t$ de Student (c) ajustados aos dados dos setores censitários de Boston reduzidos, sob ponderação de casos.

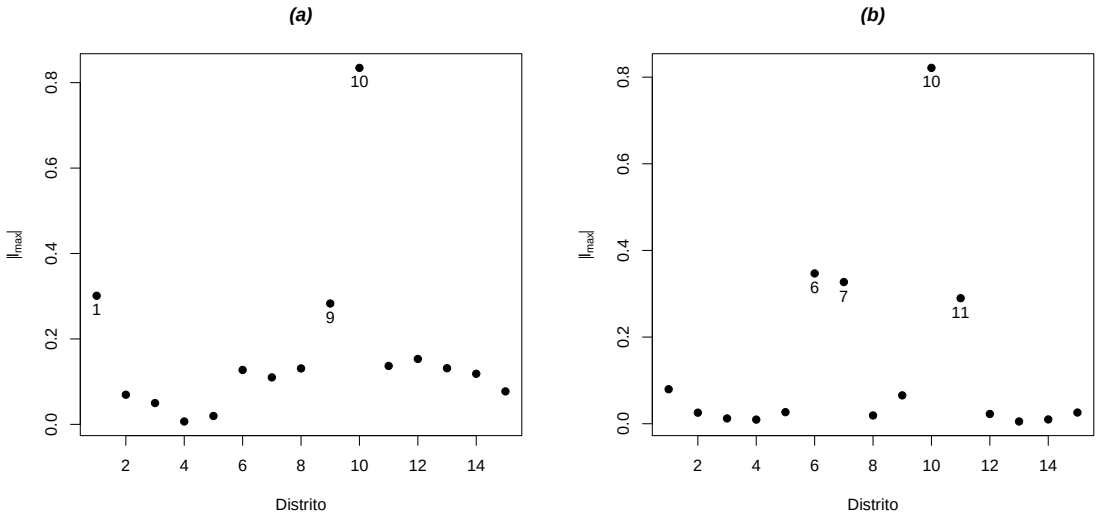


Figura 5.3: Gráficos dos índices de $|\ell_{max}|$ sob os modelos normal (a) e $t$ de Student (b) ajustados aos dados dos setores censitários de Boston reduzidos, sob perturbação na matriz de escala.

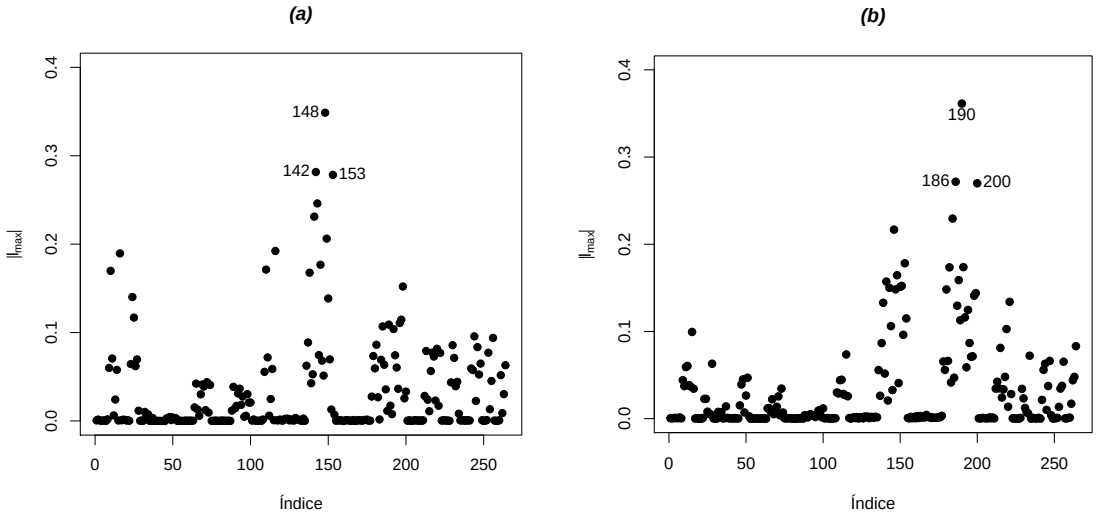


Figura 5.4: Gráficos dos índices de $|\ell_{max}|$ para perturbação no vetor de respostas sob os modelos normal (a) e $t$ de Student (b) ajustados aos dados dos setores censitários de Boston reduzidos.

### 5.2.4 Influência nas estimativas de máxima verossimilhança

Com o intuito de avaliar o impacto de alguns distritos nas estimativas de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$, são apresentadas na Tabela 5.5 as estivativas com todos os distritos (Est), ao lado as novas estimativas, obtidas reajustando o modelo (5.1)-(5.2) sem as observações #9 e #10 e as respectivas mudanças relativas (MR-percentual). Os distritos #9 e #10 foram retirados conforme a indicação da análise de influência local. Observando com mais cuidado o perfil destes distritos notou-se que não apresentam valores muito distintos dos demais, no entanto, são os grupos que possuem um número elevado de setores censitários, muito assima dos demais grupos.

Tabela 5.5: Mudanças relativas percentuais (MR) nas estimativas de máxima verossimilhança para os modelos normal e $t$ de Student com $\nu = 5$.

| | Normal | | | | | $t$ de Student | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Est | #9 | MR | #10 | MR | Est | #9 | MR | #10 | MR |
| $\beta_1$ | -0,0010 | 0,0005 | 153,73 | -0,0023 | 129,95 | -0,0021 | 0,0004 | 116,72 | -0,0039 | 85,27 |
| $\beta_2$ | 0,0009 | -0,0004 | 138,94 | 0,0009 | 2,49 | 0,0013 | -0,0003 | 120,69 | 0,0013 | 5,07 |
| $\beta_3$ | 0,0786 | 0,0630 | 19,85 | 0,0520 | 33,91 | 0,1712 | 0,1809 | 5,67 | 0,0961 | 43,87 |
| $\beta_4$ | 0,4503 | 0,3081 | 31,59 | 0,4023 | 10,67 | 0,4513 | 0,3730 | 17,35 | 0,4545 | 0,70 |
| $\beta_5$ | -0,5427 | -0,5433 | 0,12 | -0,5713 | 5,27 | -0,5668 | -0,5706 | 0,67 | -0,6043 | 6,61 |
| $\beta_6$ | -0,0072 | -0,0087 | 21,25 | -0,0071 | 0,60 | -0,0065 | -0,0079 | 22,11 | -0,0060 | 7,11 |
| $\beta_7$ | -0,0352 | -0,0659 | 87,19 | -0,0159 | 54,95 | -0,0225 | -0,0499 | 121,20 | -0,0128 | 43,42 |
| $\gamma$ | -0,0097 | -0,0066 | 31,74 | -0,0092 | 5,48 | -0,0124 | -0,0108 | 12,86 | -0,0118 | 4,83 |
| $\mu$ | 45,5838 | 46,0141 | 0,94 | 44,0217 | 3,43 | 47,4301 | 48,6194 | 2,51 | 44,8717 | 5,39 |
| $\sigma_u^2$ | 63,4978 | 66,0817 | 4,07 | 61,8890 | 2,53 | 48,4985 | 45,1938 | 6,81 | 46,7913 | 3,52 |
| $\sigma^2$ | 0,0290 | 0,0286 | 1,18 | 0,0326 | 12,32 | 0,0284 | 0,0305 | 7,13 | 0,0310 | 8,96 |
| $\tau$ | 0,0464 | 0,0503 | 8,30 | 0,0436 | 6,16 | 0,0319 | 0,0357 | 11,80 | 0,0307 | 3,84 |
| $L(\boldsymbol{\theta})$ | -434,20 | -374,86 | | -363,87 | | -427,92 | -365,16 | | -360,64 | |

Pelos resultados da Tabela 5.5 nota-se que as observações retiradas exercem grande impacto percentual sobre as estimativas e também sobre o valor de máximo da função de verossimilhança, os quais foram maiores tanto no modelo normal quanto no $t$ de Student. Na maioreia dos casos as maiores variações percentuais ocorreram para o modelo normal, sendo a variação máxima obtida ao eliminar a obsevação #9. Esta mesma observação causou uma variação menor no modelo $t$ de Student.

### 5.2.5 Ajuste do modelo proposto sem erros de medição

Para efeito de comparação foram ajustados os modelos normal e $t$ de Student sem considerar erros de medição, ou seja, considerou-se o parâmetro de escala associado aos erros de medição $\sigma_e^2 = 0$. Na Tabela 5.6 são apresentadas as estimativas dos parâmetros e na Figura 5.5 são apresentados os gráficos das distâncias transformadas para avaliar a qualidade dos

ajustes.

Tabela 5.6: Estimativas de máxima verossimilhança, erros padrão aproximados e valores $Z$ para os modelos normal e $t$ de Student ajustados aos dados dos setores censitários de Boston reduzidos, sem considerar erros de medição.

| | Normal | | | $t$ de Student | | |
|---|---|---|---|---|---|---|
| Parâmetro | Estimativa | Erro padrão | Valor-Z | Estimativa | Erro padrão | Valor-Z |
| Room ($\beta_1$) | -0,0013 | 0,0010 | -0,1344 | -0,0022 | 0,0107 | -0,2052 |
| Age ($\beta_2$) | 0,0026 | 0,0044 | 0,5915 | 0,0039 | 0,0046 | 0,8481 |
| Dist ($\beta_3$) | 0,2295 | 0,4784 | 0,4797 | 0,4393 | 0,4840 | 0,9076 |
| Black ($\beta_4$) | 0,4409 | 0,6005 | 0,7342 | 0,4263 | 0,6369 | 0,6693 |
| Lstat ($\beta_5$) | -0,5758 | 0,2244 | -2,5655 | -0,6480 | 0,2374 | -2,7292 |
| Crim ($\beta_6$) | -0,0069 | 0,0053 | -1,2931 | -0,0060 | 0,0057 | -1,0398 |
| Chas ($\beta_7$) | -0,0396 | 0,3503 | -0,1132 | -0,0285 | 0,3797 | -0,0750 |
| Noxsq ($\gamma$) | -0,0055 | 0,0018 | -3,0573 | -0,0036 | 0,0016 | -2,3321 |
| $\mu$ | 45,5836 | 2,6946 | 16,9163 | 47,5087 | 2,4849 | 19,1187 |
| $\sigma_u^2$ | 63,6980 | 30,4622 | | 50,2649 | 42,9376 | |
| $\sigma^2$ | 0,0293 | 0,0149 | | 0,0320 | 0,0280 | |
| $\tau$ | 0,0528 | 0,0813 | | 0,0377 | 0,0682 | |
| $L(\widehat{\boldsymbol{\theta}})$ | -435,7699 | | | -431,9526 | | |
| $AIC$ | 895,5398 | | | 887,9052 | | |

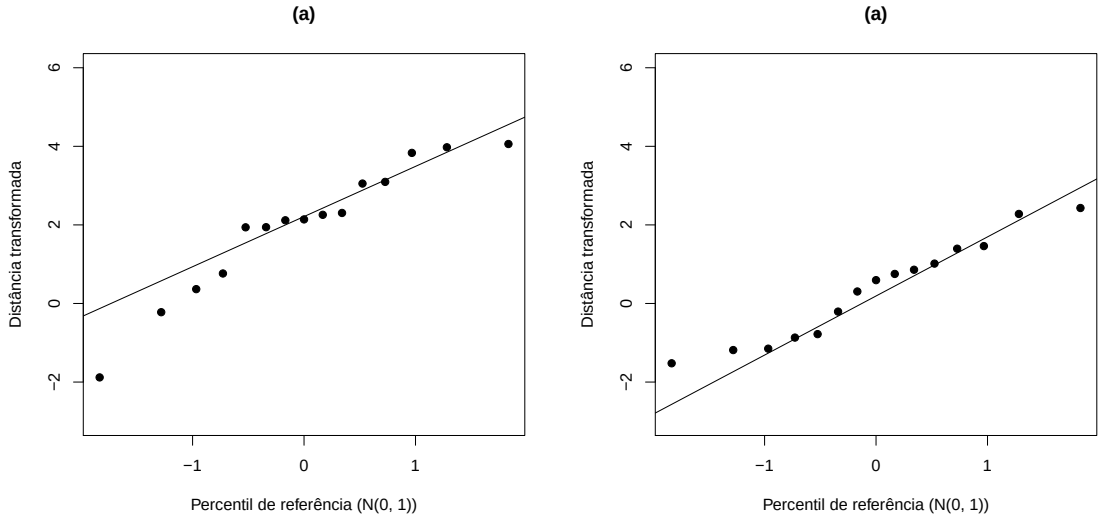


Figura 5.5: Gráficos normais de probabilidades para as distâncias transformadas sob os modelos normal (a) e $t$ de Student (b) ajustados aos dados dos setores censitários de Boston reduzidos, sem considerar erros de medição.

Pelas estimativas apresentadas na Tabela 5.6 nota-se que para ambos os modelos apenas as variáveis LSTAT e NOXSQ são marginalmente significativas, assim como observado quando incorporados os erros de medição (Tabela 5.3) e as estimativas pontuais também são

muito parecidas. Ainda, os valores de $AIC$ indicam que o modelo $t$ de Student se ajusta melhor aos dados. Por outro lado, observando os resultados da Tabela 5.3 para o critério de Akaike, quando considerou-se os erros de medição tanto o modelo normal quanto o $t$ de Student obtiveram melhores ajustes.

Observando os gráficos normais de probabilidades das distâncias transformadas da Figura 5.5, tem-se a indicação de que o modelo $t$ de Student apresenta um ajuste mais adequado, com melhor concordância entre os valores observados, em relação ao normal. O mesmo comportamento foi observado quando os modelos foram ajustados considerando erros de medição (Figura 5.1). Porém, comparando os ajustes com e sem erros de medição para ambos os modelos, nota-se uma melhor concordância entre os valores observados quando os erros de medição são incorporados aos modelos.

# Capítulo 6

# Considerações finais

Neste trabalho nós estendemos os modelos lineares mistos com erros elípticos, adicionando uma covariável sujeita a erros de medição no preditor linear. Além de modelar os efeitos das covariáveis que contribuem de maneira paramétrica, a dependência das medidas intraunidades amostrais sobre a variável resposta e estender a modelagem estatística além da distribuição normal, esta nova classe possibilita a modelagem de fenômenos que envolvem uma variável que pode estar sujeita a erros de medição, o que a torna mais flexível. Esta classe é definida de forma apropriada para que a distribuição marginal comum da resposta observada e da covariável observada e medida com erros também seja elíptica. Assim, os métodos de integração numérica não são necessários para obter o modelo marginal, e a média e a estrutura de variâncias-covariâncias do modelo hierárquico são preservadas. Além disso, a flexibilidade da curtose é permitida para cada distribuição marginal comum e desde que as distribuições condicionais também sejam elípticas. Outra vantagem diz respeito às previsões dos efeitos aleatórios bem como da covariável sujeita a erros de medição, que podem ser realizadas de maneira semelhante à do caso normal, por meio do método de Bayes empírico. Considerando que a verificação da identificabilidade em modelos que consideram efeitos mistos e erros de medição é uma etapa fundamental para a definição de modelos, foram analisadas e apresentadas as condições para que a classe proposta seja identificável. Outras contribuições importantes desta tese são o desenvolvimento de um processo iterativo baseado no método de máxima verossimilhança, derivado para a obtenção das estimativas dos parâmetros, as quais parecem ser robustas contra observações discrepantes no sentido da distância de Mahalanobis. O desenvolvimento de métodos de diagnóstico para estudar a sensibilidade das estimativas dos parâmetros, em que as curvaturas de influência local foram derivadas sob alguns esquemas de perturbação usuais, selecionados apropriadamente segundo a recente metodologia proposta por Zhu *et al.* (2007). Um exemplo de motivação analisado sob erros normais foi novamente analisado, considerando os erros com caudas pesadas para mostrar a aplicabilidade da teoria desenvolvida.

## 6.1 Perspectivas futuras

O processo iterativo, baseado no método de máxima verossimilhança, desenvolvido para estimar os coeficientes de regressão e os componentes de variância, sob o modelo misto linear elíptico com erros de medição proposto, bem como a análise de diagnóstico de influência local, foi implementado no *software* R (R Development Core Team, 2011). Uma primeira perspectiva de trabalho futuro é melhorar a implementação por meio de códigos mais eficientes para que possam ser utilizados por outros usuários interessados.

Outra perspectiva de trabalho futuro é estender o modelo proposto nesta tese no sentido de adicionar um componente fixo não paramétrico aos efeitos fixos e aleatórios. Neste sentido, Ibacache-Pulgar *et al.* (2012) apresentaram uma extensão da classe proposta por Savalli *et al.* (2006), criando assim os modelos mistos semiparamétricos elípticos.

# Referências


Akaike, H. (1974). A new look at statistical models identification. IEEE Transactions on Automatic Control AU-19, 716-722.

Arellano-Valle, R. B. (1994). Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão. Tese de doutorado. IME-USP, São Paulo.

Arellano-Valle, R. B. (2010). On the information matrix of the multivariate skew-$t$ model. METRON - International Journal of Statistics, 68, 371-386.

Atkinson, C. A. (1985). Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Oxford University Press, Oxford.

Belsley, D. A.; Kuh, E. e Welsch, R. E. (1980). Regression Diagnostics: Identifying Influence Data and Sources of Collinearity. Wiley, New York.

Box, G. E. P. e Tiao, G. C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, Massachusetts.

Casella, G. e Berger, R. L. (2002). Statistical Inference. Duxbury/Thomson, Pacific Grove.

de Castro, M. e Galea, M. (2010). Robust inference in an heteroscedastic measurement error model. Journal of the Korean Statistical Society, 39, 439-447.

Chen, F.; Zhu, H. e Lee, S. (2009). Perturbation selection and local influence analysis for nonlinear structural equation model. Psychometrika, 74, 493-516.

Chen, F.; Zhu, H.; Song, X. e Lee, S. (2011). Perturbation selection and local influence analysis for generalized linear mixed models. Journal of Computational and Graphical Statistics, 19, 826-842.

Cook, R. D. (1977). Detection of influential observations in linear regression. Technometrics, 19, 15-18.

Cook, R. D. (1986). Assessment of local influence (with discussion). Journal of the Royal Statistical Society, B, 48, 133-169.

Cook, R. D. e Weisberg, S. (1982). Residuals and Influence in Regression. Chapman and Hall, London.

Copt, S. e Victoria-Ferrer, M. P. (2006). High-breakdown inference for mixed linear models. Journal of the American Statistical Association, 101, 292-300.

Cui, H.; Ng, K. W. e Zhu, L. (2004). Estimation in mixed effects model with errors in variables. Journal of Multivatiate Analysis, 91, 53-73.

Davidian, M. e Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: an overview and update. Journal of Agricultural, Biological, and Environmental Statistics, 8, 387-419.

Dagenais, M. G. e Dufour, J. M. (1991). Invariance, nonlinear models and asymptotic tests. Econometrica, 59, 1601-1615.

Demidenko, E. (2004). Mixed Models: Theory and Applications. John Wiley & Sons, New Jersey.

Diggle, P. J.; Liang, K. T. e Zeger, S. L. (1994). Analysis of Longitudinal Data. Clarendon Press, Oxford.

Escobar, E. e Meeker, W. (1992). Assessing influence in regression analysis with censored data. Biometrics, 48, 507-528.

Fang, K. T. e Anderson, T. W. (1990). Statistical Inference in Elliptically Contoured and Related Distributions. Allerton Press Inc., New York.

Fang, K. T.; Kotz, S. e Ng, K. W. (1990). Symmetric Multivariate and Related Distributions. Chapman & Hall, London.

Fuller, W. A. (1987). Mensurement Error Models. Wiley, New York.

Galea, M.; Paula, G. A. e Bolfarine, H. (1997). Local influence in elliptical linear regression models. The Statistician, 46, 71-79.

Galea, M.; Riquelme, M. e Paula, G. A. (2000). Diagnostics methods in elliptical linear regression models. Brazilian Journal of Probability and Statistics, 14, 167-184.

Graybill, F. A. (1983). Matrices with Applications in Statistics. 2nd Edição. Wadsworth Publishing Company, California.

Giménez, P. e Galea, M. (2013). Influence measures on corrected score estimators in functional heteroscedastic measurement error models. Journal of Multivariate Analysis, 114, 1-15.

Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. John Wiley, New York.

Harrison, D. e Rubinfeld, D.L. (1978). Hedonic prices and the demand for clean air, J. Environ. Economics & Management, 5, 81-102.

Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. The Annals of Statistics, 4, 384-395.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72, 320-342.

Hawkins, D. M. e Wixley, R. A. J. (1986). A note on the transformation of chi-squared variables to normality. The American Statistician, 40, 296-298.

Hopper, J. L. e Matthews, J. D. (1982). Extensions to multivariate normal models for pedigree analysis. Annals of Human Genetics, 46, 373-383.

Ibacache-Pulgar, G. (2009). Modelos Mistos Aditivos Semiparamétricos de Contornos Elípticos. Tese de Doutorado. IME-USP, São Paulo.

Ibacache-Pulgar, G. e Paula, G. A. (2011). Local influence for Student-t partially linear models. Computacional Statistics and Data Analysis, 55, 1462-1478.

Ibacache-Pulgar, G.; Paula, G. A. e Galea, M. (2012). Influence diagnostics for elliptical semiparametric mixed models. Statistical Modelling, 12, 165-170.

Jameson, A. (1968). Solution of the Equation $AX + XB = C$ by inversion of an $M \times M$ or $N \times N$ matrix. SIAM J. Appl. Math., 16, 1020-1023.

Johnson, N. L.; Kotz, S. e Balakrishnan, N. (1994). Continuous Univariate Distributions. 2nd ediction. John Wiley, New York.

Lai, T. L. e Shih, M. C. (2003). Nonparametric estimation in nonlinear mixed effects models. Biometrika, 90, 1-13.

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

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.

Lehmann, E. L. (1983). Theory of Point Estimation. John Wiley and Sons, New York.

Lesaffre, E. e Verbeke, G. (1998). Local influence in linear mixed models. Biometrics, 54, 570-582.

Little, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. Applied Statistics, 37, 23-38.

Little, R. e Smith, P. (1987). Editing and imputation for quantitative survey data. Journal of the American Statistical Association, 82, 58-68.

Longford, N. T. (1993). Random Coeficient Models. Oxford University press, New York.

Lucas, A. (1997). Asymptotic robustness of least median of squares for autoregressions with additive outliers. Communications in Statistics - Theory and Methods, 26, 2363-80.

Magnus, J. R. e Neudecker, H. (2002). Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, Chichester.

Mitchell, A. F. S. (1989). The information matrix, skewness tensor and $\alpha$-connections for the general multivariate elliptic distribution. Annals of the Institute of Statistical Mathematics, 41, 289-304.

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

Neyman, J. (1959). Optimal Asymptotic Tests of Composite Statistical Hypotheses. In Probability and Statistics, the Harald Cramer Volume, ed. by U. Grenander. Uppsala: Almqvist and Wiksell, 213-234.

Nobre, J. S. (2003). Métodos de Diagnóstico para Modelos Lineares Mistos. Dissertação de Mestrado. São Paulo, IME-USP.

Nobre, J. S. e Singer, J. M. (2007). Residual analysis for linear mixed models. Biometrical Journal, 49, 863-875.

Osborne, M.R. (1992). Fisher's method of scoring. Int. Stat. Rev, 86, 271-286.

Osorio, F. (2006). Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos. Tese de Doutorado. IME-USP, São Paulo.

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

Paula, G.A. (2013). Modelos de Regressão com Apoio Computacional. Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.

Pinheiro, J. C.; 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.

Poon, W. e Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. Journal of the Royal Statistical Society, Series B, 61, 51-61.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.Rproject.org.

Rao, C. R. (1948). Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation. Proceedings of the Cambridge Philosophical Society, 44, 50-57.

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

Russo, C.; Aoki, R. e Paula, G. A. (2012). Assessment of variance components in nonlinear mixed-effect elliptical models. TEST, 21, 519-545.

Russo, C.; Paula, G. A.; Aoki, R. e Cysneiros, F. J. A . (2012). Influence dianostics in heteroscedastic and/or autoregressive nonlinear elliptical models for correlated data. Journal of Applied Statistics, 39, 1049-1067.

Savalli, C. (2006). Testes do Tipo Escore para Componentes de Variância em Modelos Elípticos Lineares Mistos. Tese de Doutorado. IME/USP, São Paulo.

Savalli, C.; Paula, G. A. e Cysneiros, F. J. A . (2006). Assessment of variance components in elliptical linear mixed models. Statistical Modelling, 6, 59-76.

Shi, X.; Zhu, H. e Ibrahim, J.G. (2009). Local influence for generalized linear models with missing covariates. Biometrics, 65, 1164-1174.

Silvapulle, M. e Silvapulle, P. (1995). A score test against one-sided alternatives, Journal of the American Statistical Association, 429, 1-8.

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

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

Wu, L. (2010). Mixed Effects Models for Complex Data. Monografhs on Statistics and Applied Probability 113. A Chapman & Hall Book.

Verbeke, G. e Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data. Springer, New York.

Zhao, Y. e Lee, A. H. (1995). Assesssment of influence in non-linear measurement error models. Journal of Applied Statistics, 22, 215-225.

Zhao, Y. e Lee, A. H. (1998). Influence diagnostics for simultaneous equations models. Australian and New Zealand Journal of Statistics, 40, 345-357.

Zhong, X. P.; Fung, W. K. e Wei, B. C. (2002). Estimation in linear models with random effects and errors-in-variables. Annals of the Institute of Statistical Mathematics, 54, 595-606.

Zhong, X. P.; Wei, B. C. e Fung, W. K. (2000). Influence analysis for linear measurement error models. Annals of the Institute of Statistical Mathematics, 52, 367-379.

Zhu, H. T. e Lee, S. Y. (2003). Local influence for generalized linear mixed models. The Canadian Journal of Statistics, 31, 293-309.

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.

# Apêndice A

# Derivadas do logaritmo da função de verossimilhança

Neste apêndice são apresentados os cálculos das derivadas de primeira e segunda ordens do logaritmo da função de verossimilhança do modelo misto linear elíptico com erros de medição. Essas derivadas foram utilizadas para a especificação da função escore do vetor de parâmetros $\boldsymbol{\theta}$ e para os métodos de influência local.

## A.1 Derivadas de primeira ordem

No modelo misto linear elíptico com erros de medição (3.3)-(3.4) o logaritmo da função de verossimilhança é dado por

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

em que
$L_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{V}_i| + \log g(\boldsymbol{\delta}_i)$ e $\boldsymbol{\theta} = \left(\gamma, \boldsymbol{\beta}^T, \mu, \sigma_u^2, \sigma^2, \boldsymbol{\tau}^T\right)^T$. Usando resultados de diferenciação de matrizes temos que,

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{A.2}$$

Deste modo, derivando (A.1) com relação a cada componente do vetor $\boldsymbol{\theta}$, temos que

$$
\begin{aligned}
\frac{\partial L_i(\boldsymbol{\theta})}{\partial \gamma} &= -\frac{1}{2}\frac{\partial \log |\boldsymbol{V}_i|}{\partial \gamma} + \frac{\partial \log g(\boldsymbol{\delta}_i)}{\partial \gamma} \\
&= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) + \frac{g'(\boldsymbol{\delta}_i)}{g(\boldsymbol{\delta}_i)} \times \frac{\partial \boldsymbol{\delta}_i}{\partial \gamma} \\
&= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) + W_g(\boldsymbol{\delta}_i)\left[-\begin{pmatrix} \mu\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i + \boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}\boldsymbol{r}_i}{\partial \gamma}\right] \\
&= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) + W_g(\boldsymbol{\delta}_i)\left\{-\begin{pmatrix} \mu\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right. \\
&\quad \left. +\boldsymbol{r}_i^T\left[\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \gamma}\boldsymbol{r}_i - \boldsymbol{V}_i^{-1}\begin{pmatrix} \mu\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix}\right]\right\} \\
&= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) + W_g(\boldsymbol{\delta}_i)\left\{-\begin{pmatrix} \mu\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right. \\
&\quad \left. - \boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i - \boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix} \mu\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix}\right\} \qquad \text{(A.3)} \\
&= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right) - \frac{1}{2}v(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix} \mu\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix} + \boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right],
\end{aligned}
$$

em que $v(\boldsymbol{\delta}_i) = -2W_g(\boldsymbol{\delta}_i)$ e $W_g(\boldsymbol{\delta}_i) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\delta}_i}\log g(\boldsymbol{\delta}_i) = \frac{g'(\boldsymbol{\delta}_i)}{g(\boldsymbol{\delta}_i)}$;

$$
\begin{aligned}
\frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \frac{g'(\boldsymbol{\delta}_i)}{g(\boldsymbol{\delta}_i)} \times \frac{\partial \boldsymbol{r}_i^T}{\partial \boldsymbol{\beta}} \times \frac{\partial\left[\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]}{\partial \boldsymbol{r}_i} \\
&= -W_g(\boldsymbol{\delta}_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T 2\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \\
&= v(\boldsymbol{\delta}_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \qquad \text{(A.4)}
\end{aligned}
$$

com

$$
\boldsymbol{r}_i = (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}) = \begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{u}_i \end{pmatrix} - \begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} + \gamma\mu\mathbf{1}_{m_i} \\ \mu\mathbf{1}_{m_i} \end{pmatrix};
$$

$$\begin{aligned}
\frac{\partial L_i(\boldsymbol{\theta})}{\partial \mu} &= \frac{g'(\boldsymbol{\delta}_i)}{g(\boldsymbol{\delta}_i)} \times \frac{\partial \boldsymbol{r}_i^T}{\partial \mu} \times \frac{\partial \left[\boldsymbol{r}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{r}_i\right]}{\partial \boldsymbol{r}_i} \\
&= -W_g(\boldsymbol{\delta}_i) \begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T 2\boldsymbol{V}_i^{-1} \boldsymbol{r}_i \\
&= v(\boldsymbol{\delta}_i) \begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \boldsymbol{V}_i^{-1} \boldsymbol{r}_i. \qquad \text{(A.5)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_r} &= -\frac{1}{2} \frac{\partial \log |\boldsymbol{V}_i|}{\partial \boldsymbol{\eta}_r} + \frac{\partial g(\boldsymbol{\delta}_i)}{\partial \boldsymbol{\eta}_r} \\
&= -\frac{1}{2} \mathrm{tr} \left( \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_r} \right) + \frac{g'(\boldsymbol{\delta}_i)}{g(\boldsymbol{\delta}_i)} \times \frac{\partial \boldsymbol{\delta}_i}{\partial \boldsymbol{\eta}_r} \\
&= -\frac{1}{2} \mathrm{tr} \left( \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_r} \right) - \frac{1}{2} v(\boldsymbol{\delta}_i) \boldsymbol{r}_i^T \frac{\partial \boldsymbol{V}_i^{-1}}{\partial \boldsymbol{\eta}_r} \boldsymbol{r}_i \\
&= -\frac{1}{2} \mathrm{tr} \left( \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_r} \right) + \frac{1}{2} v(\boldsymbol{\delta}_i) \boldsymbol{r}_i^T \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\eta}_r} \boldsymbol{V}_i^{-1} \boldsymbol{r}_i,
\end{aligned}$$

em que

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \sigma_u^2} = -\frac{1}{2} \mathrm{tr} \left( \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2} \right) + \frac{1}{2} v(\boldsymbol{\delta}_i) \boldsymbol{r}_i^T \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2} \boldsymbol{V}_i^{-1} \boldsymbol{r}_i, \qquad \text{(A.6)}$$

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{1}{2} \mathrm{tr} \left( \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma^2} \right) + \frac{1}{2} v(\boldsymbol{\delta}_i) \boldsymbol{r}_i^T \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \sigma^2} \boldsymbol{V}_i^{-1} \boldsymbol{r}_i \qquad \text{(A.7)}$$

e

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \tau_j} = -\frac{1}{2} \mathrm{tr} \left( \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \tau_j} \right) + \frac{1}{2} v(\boldsymbol{\delta}_i) \boldsymbol{r}_i^T \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \tau_j} \boldsymbol{V}_i^{-1} \boldsymbol{r}_i. \qquad \text{(A.8)}$$

## A.2 Derivadas de segunda ordem

Usando novamente resultados de diferenciação de matrizes temos que a matriz de segundas derivadas em relação a $\boldsymbol{\theta}$ é dada por

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \sum_{i=1}^{n} \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}. \tag{A.9}$$

Derivando (A.3), (A.4), (A.5), (A.6), (A.7) e (A.8), com relação a $\gamma$, $\boldsymbol{\beta}^T$, $\mu$, $\sigma_u^2$, $\sigma^2$ e $\boldsymbol{\tau}^T$, respectivamente, temos que as matrizes de segundas derivadas parciais são dadas por

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\gamma} &= -\frac{1}{2}\frac{\partial\left[\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)\right]}{\partial\gamma} + \frac{\partial\left[W_g(\boldsymbol{\delta}_i)\left(-2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right)\right]}{\partial\gamma} \\
&= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\frac{\partial\boldsymbol{V}_i}{\partial\gamma} + \boldsymbol{V}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma^2}\right) \\
&\quad + W_g'(\boldsymbol{\delta}_i)\left[-2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right]\left[-2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right. \\
&\quad \left. +\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right] + W_g(\boldsymbol{\delta}_i)\left[\frac{\partial\left(-2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right)}{\partial\gamma} + \frac{\partial\left(\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right)}{\partial\gamma}\right] \\
&= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right] - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma^2}\right) \\
&\quad + W_g'(\boldsymbol{\delta}_i)\left[-2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right]^2 \\
&\quad + W_g(\boldsymbol{\delta}_i)\left\{2\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + 4\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right\} \\
&\quad + W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\left[2\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1} - \boldsymbol{V}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma^2}\boldsymbol{V}_i^{-1}\right]\boldsymbol{r}_i,
\end{aligned}$$

sendo que para estes cálculos usamos os resultados:

$$\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\theta}} = -\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\theta}}\boldsymbol{V}_i^{-1} \quad \text{(Graybill, 1983)}$$

e

$$\frac{\partial^2 \boldsymbol{V}_i^{-1}}{\partial\gamma\partial\gamma} = 2\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1} - \boldsymbol{V}_i^{-1}\frac{\partial^2 \boldsymbol{V}_i}{\partial\gamma\partial\gamma}\boldsymbol{V}_i^{-1};$$

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} &= -2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\frac{\partial[\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i)]}{\partial\boldsymbol{\beta}^T}\\
&= -2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[-\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}W_g(\boldsymbol{\delta}_i) + \boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\left(-2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}\right)\right]\\
&= 2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[2W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix} + W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}\right].
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\mu\partial\mu} &= -2\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\frac{\partial[\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i)]}{\partial\boldsymbol{\beta}}\\
&= -2\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[-\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix}W_g(\boldsymbol{\delta}_i) + \boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\left(-2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix}\right)\right]\\
&= 2\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[2W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix} + W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix}\right].
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\sigma_u^2} &= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2} + \boldsymbol{V}_i^{-1}\frac{\partial^2 \boldsymbol{V}_i}{\partial\sigma_u^2\partial\sigma_u^2}\right) + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\boldsymbol{r}_i\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\boldsymbol{r}_i\\
&\quad + W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial^2 \boldsymbol{V}_i^{-1}}{\partial\sigma_u^2\partial\sigma_u^2}\boldsymbol{r}_i\\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\right) + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\\
&\quad + 2W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma^2\partial\sigma^2} &= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\right) + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\\
&\quad + 2W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i.
\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^T} &= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\right) + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \\ &\quad + 2W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i,\end{aligned}$$

sendo que para estes cálculos usamos o resultado:

$$\frac{\partial^2\boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T} = 2\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\eta}^T}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\eta}}\boldsymbol{V}_i^{-1};$$

$$\begin{aligned}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\boldsymbol{\beta}^T} &= \frac{\partial}{\partial\boldsymbol{\beta}^T}\left(-2W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right) + \frac{\partial}{\partial\boldsymbol{\beta}^T}\left(W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right) \\ &= -2W_g'(\boldsymbol{\delta}_i)(-2)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} \\ &\quad +2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} \\ &\quad -2W_g'(\boldsymbol{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i + W_g(\boldsymbol{\delta}_i)(-2)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i \\ &= \left[2W_g'(\boldsymbol{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T + W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\right]\boldsymbol{V}_i^{-1}\left[2\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right. \\ &\quad \left.+\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\mu} &= \frac{\partial}{\partial\mu}\left\{-W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right]\right\}\\
&= 2W_g'(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right]\\
&\quad -2W_g(\boldsymbol{\delta}_i)\left[-\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right.\\
&\quad \left.+\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right]\\
&= 2W_g'(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\left[2\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]\\
&\quad +2W_g(\boldsymbol{\delta}_i)\left[\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right.\\
&\quad \left.+\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right].
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\sigma_u^2} &= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}+\boldsymbol{V}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma\partial\sigma_u^2}\right)\\
&\quad +W_g'(\boldsymbol{\delta}_i)\left[\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\boldsymbol{r}_i\right]\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right]\\
&\quad +W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\boldsymbol{r}_i^T\frac{\partial^2\boldsymbol{V}_i^{-1}}{\partial\gamma\partial\sigma_u^2}\boldsymbol{r}_i\right]\\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma\partial\sigma_u^2}\right)\\
&\quad -W_g'(\boldsymbol{\delta}_i)\left[\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]\\
&\quad -W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\boldsymbol{r}_i^T\boldsymbol{G}_i^{\sigma_u^2}\boldsymbol{r}_i\right],
\end{aligned}
$$

em que

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

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\sigma^2} &= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma^2}\frac{\partial \boldsymbol{V}_i}{\partial\gamma} + \boldsymbol{V}_i^{-1}\frac{\partial^2 \boldsymbol{V}_i}{\partial\gamma\partial\sigma^2}\right) \\
&\quad + W_g'(\boldsymbol{\delta}_i)\left[\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma^2}\boldsymbol{r}_i\right]\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - \boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right] \\
&\quad + W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma^2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - \boldsymbol{r}_i^T\frac{\partial^2 \boldsymbol{V}_i^{-1}}{\partial\gamma\partial\sigma^2}\boldsymbol{r}_i\right] \\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\right) \\
&\quad - W_g'(\boldsymbol{\delta}_i)\left[\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right] \\
&\quad - W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^T\boldsymbol{G}_i^{\sigma^2}\boldsymbol{r}_i\right],
\end{aligned}$$

em que

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

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\boldsymbol{\tau}^T} &= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\frac{\partial \boldsymbol{V}_i}{\partial\gamma} + \boldsymbol{V}_i^{-1}\frac{\partial^2 \boldsymbol{V}_i}{\partial\gamma\partial\boldsymbol{\tau}^T}\right) \\
&\quad + W_g'(\boldsymbol{\delta}_i)\left[\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i\right]\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - \boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\gamma}\boldsymbol{r}_i\right] \\
&\quad + W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - \boldsymbol{r}_i^T\frac{\partial^2 \boldsymbol{V}_i^{-1}}{\partial\gamma\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i\right] \\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\right) \\
&\quad - W_g'(\boldsymbol{\delta}_i)\left[\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right] \\
&\quad - W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{r}_i^T\boldsymbol{G}_i^{\tau^T}\boldsymbol{r}_i\right],
\end{aligned}$$

em que

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

Podemos escrever

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} = -2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i).$$

Assim,

$$\begin{aligned}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\mu} &= -2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\frac{\partial[\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i)]}{\partial\mu}\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix} + 2W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\begin{pmatrix}\gamma\boldsymbol{1}_{m_i}\\ \boldsymbol{1}_{m_i}\end{pmatrix}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma_u^2} &= -2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \left[\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i) + \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\boldsymbol{r}_i\right]\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \left[W_g(\boldsymbol{\delta}_i)\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i + W_g'(\boldsymbol{\delta}_i)\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right]\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{I}_{m_i} + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i,\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma^2} &= -2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \left[\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma^2}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i) + \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\sigma^2}\boldsymbol{r}_i\right]\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{I}_{m_i} + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\end{aligned}$$

e

$$\begin{aligned}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\tau}^T} &= -2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \left[\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i) + \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i\right]\\ &= 2\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{I}_{m_i} + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i.\end{aligned}$$

Podemos escrever

$$\frac{\partial L_i(\boldsymbol{\theta})}{\partial \mu} = -2W_g(\boldsymbol{\delta}_i)\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i = -2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i).$$

Assim,

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \mu \partial \sigma_u^2} &= -2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \left[\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \sigma_u^2}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i) + \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T \frac{\partial \boldsymbol{V}_i^{-1}}{\partial \sigma_u^2}\boldsymbol{r}_i\right] \\
&= 2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \left[W_g(\boldsymbol{\delta}_i)\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i + W_g'(\boldsymbol{\delta}_i)\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right] \\
&= 2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{I}_{m_i} + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \mu \partial \sigma^2} &= -2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \left[\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \sigma^2}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i) + \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T \frac{\partial \boldsymbol{V}_i^{-1}}{\partial \sigma^2}\boldsymbol{r}_i\right] \\
&= 2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{I}_{m_i} + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \mu \partial \boldsymbol{\tau}^T} &= -2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \left[\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \boldsymbol{\tau}^T}\boldsymbol{r}_i W_g(\boldsymbol{\delta}_i) + \boldsymbol{V}_i^{-1}\boldsymbol{r}_i W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T \frac{\partial \boldsymbol{V}_i^{-1}}{\partial \boldsymbol{\tau}^T}\boldsymbol{r}_i\right] \\
&= 2\begin{pmatrix} \gamma \mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \boldsymbol{V}_i^{-1}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{I}_{m_i} + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial \sigma_u^2 \partial \sigma^2} &= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \sigma^2}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2} + \boldsymbol{V}_i^{-1}\frac{\partial^2 \boldsymbol{V}_i}{\partial \sigma_u^2 \partial \sigma^2}\right) + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \sigma_u^2}\boldsymbol{r}_i\boldsymbol{r}_i^T\frac{\partial \boldsymbol{V}_i^{-1}}{\partial \sigma^2}\boldsymbol{r}_i \\
&\quad + W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial^2 \boldsymbol{V}_i^{-1}}{\partial \sigma_u^2 \partial \sigma^2}\boldsymbol{r}_i \\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) + W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \\
&\quad + 2W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i,
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\boldsymbol{\tau}^T} &= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}+\boldsymbol{V}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\sigma_u^2\partial\boldsymbol{\tau}^T}\right)+W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\sigma_u^2}\boldsymbol{r}_i\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i \\
&\quad +W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial^2\boldsymbol{V}_i^{-1}}{\partial\sigma_u^2\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i \\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\right)+W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \\
&\quad +\,2W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i
\end{aligned}
$$

e

$$
\begin{aligned}
\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma^2\partial\boldsymbol{\tau}^T} &= -\frac{1}{2}\operatorname{tr}\left(\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}+\boldsymbol{V}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\sigma^2\partial\boldsymbol{\tau}^T}\right)+W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\sigma^2}\boldsymbol{r}_i\boldsymbol{r}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i \\
&\quad +W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\frac{\partial^2\boldsymbol{V}_i^{-1}}{\partial\sigma^2\partial\boldsymbol{\tau}^T}\boldsymbol{r}_i \\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\right)+W_g'(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \\
&\quad +\,2W_g(\boldsymbol{\delta}_i)\boldsymbol{r}_i^T\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i.
\end{aligned}
$$

# Apêndice B

# Matriz de informação observada

Neste apêndice são apresentadas as expressões de cada elemento da matriz de informação observada, apresentada em (4.3), tanto para o caso normal quanto para o caso elíptico.

## B.1 Elementos da matriz de informação observada - caso normal

A partir dos cálculos das primeiras derivadas obtidas para a função escore apresentadas na Seção 2.3, calculamos as segundas derivadas, avaliadas em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, ou seja,

$$-\ddot{\mathbf{L}}_i(\boldsymbol{\theta}) = -\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\bigg|_{\theta=\widehat{\theta}},$$

. Assim, chegamos às expressões a seguir.

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\gamma\gamma}(\hat{\theta}) &= \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\gamma}\bigg|_{\theta=\widehat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left[\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right] - \frac{1}{2}\operatorname{tr}\left[\hat{\boldsymbol{V}}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma\partial\gamma}\right] - \begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} \\
&\quad - \left[2\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T + \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma} - \frac{1}{2}\hat{\boldsymbol{r}}_i^T\right]\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\beta\beta}(\hat{\theta}) &= \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\bigg|_{\theta=\widehat{\theta}} \\
&= -\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix},
\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma_u^2\sigma^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\sigma^2}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right) - \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,
\end{aligned}$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma_u^2\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\boldsymbol{\tau}^T}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\right) - \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i,
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma^2\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma^2\partial\boldsymbol{\tau}^T}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\right) - \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}
$$

## B.2 Elementos da matriz de informação observada - caso elíptico

A partir dos cálculos das segundas derivadas obtidas no Apêndice A.2 e avaliando em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, chegamos às expressões a seguir.

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\gamma\gamma}(\hat{\theta}) &= \left.\frac{\partial^2\,\mathbf{L}_i(\boldsymbol{\theta})}{\partial\gamma\partial\gamma}\right|_{\theta=\widehat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left[\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right] - \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma^2}\right) \\
&\quad + W_g'(\widehat{\delta}_i)\left[-2\widehat{\boldsymbol{r}}_i^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \widehat{\boldsymbol{r}}_i^T\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\gamma}\widehat{\boldsymbol{r}}_i\right]^2 \\
&\quad + W_g(\widehat{\delta}_i)\left\{2\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + 4\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\widehat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right\} \\
&\quad + W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\left[2\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma} - \frac{\partial^2\boldsymbol{V}_i}{\partial\gamma^2}\right]\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i,
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\beta\beta}(\hat{\theta}) &= \left.\frac{\partial^2\,\mathbf{L}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right|_{\theta=\widehat{\theta}} \\
&= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\widehat{\boldsymbol{V}}_i^{-1}\left[2W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix} + W_g(\widehat{\delta}_i)\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}\right],
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\mu\mu}(\hat{\theta}) &= \left.\frac{\partial^2 \mathbf{L}_i(\boldsymbol{\theta})}{\partial\mu\partial\mu}\right|_{\theta=\widehat{\theta}} \\
&= 2\begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \widehat{\boldsymbol{V}}_i^{-1}\left[2W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix} + W_g(\widehat{\delta}_i)\begin{pmatrix} \widehat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}\right],
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma_u^2\sigma_u^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\sigma_u^2}\right|_{\theta=\widehat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right) + W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i \\
&\quad + 2W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i.
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma^2\sigma^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma^2\partial\sigma^2}\right|_{\theta=\widehat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right) + W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i \\
&\quad + 2W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i,
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\tau\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^T}\right|_{\theta=\widehat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\right) + W_g'(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i \\
&\quad + 2W_g(\widehat{\delta}_i)\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\widehat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i.
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\gamma\beta}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\boldsymbol{\beta}^T}\right|_{\theta=\hat{\theta}} \\
&= \left[2W_g'(\widehat{\boldsymbol{\delta}}_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T + W_g(\widehat{\boldsymbol{\delta}}_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T\right]\widehat{\boldsymbol{V}}_i^{-1}\left[2\begin{pmatrix} \widehat{\mu}\mathbf{1}_{m_i} \\ \mathbf{0} \end{pmatrix}\right. \\
&\quad \left. + \frac{\partial\boldsymbol{V}_i}{\partial\gamma}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i\right].
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\gamma\mu}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\mu}\right|_{\theta=\hat{\theta}} \\
&= 2W_g'(\hat{\boldsymbol{\delta}}_i)\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\left[2\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right] \\
&\quad +2W_g(\hat{\boldsymbol{\delta}}_i)\left[\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T \hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}-\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\right. \\
&\quad \left.+\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T \hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right].
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\gamma\sigma_u^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\sigma_u^2}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)-\frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial^2\boldsymbol{V}_i}{\partial\gamma\partial\sigma_u^2}\right) \\
&\quad -W_g'(\hat{\delta}_i)\left[\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right]\left[2\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right] \\
&\quad -W_g(\hat{\delta}_i)\left[2\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{G}}_i^{\sigma_u^2}\hat{\boldsymbol{r}}_i\right],
\end{aligned}
$$

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

$$
\begin{aligned}
\ddot{\mathbf{L}}_i^{\gamma\sigma^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\sigma^2}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)-W_g(\hat{\delta}_i)\left[2\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{G}}_i^{\sigma^2}\hat{\boldsymbol{r}}_i\right] \\
&\quad -W_g'(\hat{\delta}_i)\left[\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right]\left[2\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}+\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right],
\end{aligned}
$$

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

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\gamma\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\gamma\partial\boldsymbol{\tau}}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\right) - W_g(\hat{\delta}_i)\left[2\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{G}}_i^{\tau}\hat{\boldsymbol{r}}_i\right] \\
&\quad -W_g'(\hat{\delta}_i)\left[\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right]\left[2\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\begin{pmatrix}\hat{\mu}\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial\gamma}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right],
\end{aligned}$$

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

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\beta\mu}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\mu}\right|_{\theta=\hat{\theta}} \\
&= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1}\left[W_g(\hat{\boldsymbol{\delta}}_i)\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix} + 2W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\begin{pmatrix}\hat{\gamma}\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i\right].
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\beta\sigma_u^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma_u^2}\right|_{\theta=\hat{\theta}} \\
&= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1}\left[W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\beta\sigma^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma^2}\right|_{\theta=\hat{\theta}} \\
&= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1}\left[W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\beta\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\tau}^T}\right|_{\theta=\hat{\theta}} \\
&= 2\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\hat{\boldsymbol{V}}_i^{-1}\left[W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\right]\frac{\partial \boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\mu\sigma_u^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\mu\partial\sigma_u^2}\right|_{\theta=\hat{\theta}} \\
&= 2\begin{pmatrix} \hat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \hat{\boldsymbol{V}}_i^{-1}\left[W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\right]\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\mu\sigma^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\mu\partial\sigma^2}\right|_{\theta=\hat{\theta}} \\
&= 2\begin{pmatrix} \hat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \hat{\boldsymbol{V}}_i^{-1}\left[W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\right]\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\mu\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\mu\partial\boldsymbol{\tau}^T}\right|_{\theta=\hat{\theta}} \\
&= 2\begin{pmatrix} \hat{\gamma}\mathbf{1}_{m_i} \\ \mathbf{1}_{m_i} \end{pmatrix}^T \hat{\boldsymbol{V}}_i^{-1}\left[W_g(\hat{\boldsymbol{\delta}}_i)\boldsymbol{I}_{2m_i} + W_g'(\hat{\boldsymbol{\delta}}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\right]\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma_u^2\sigma^2}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\sigma^2}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right) \\
&\quad +\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\left[W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1} + 2W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}\right]\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma_u^2\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma_u^2\partial\boldsymbol{\tau}^T}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\right) \\
&\quad +\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\hat{\boldsymbol{V}}_i^{-1}\left[W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1} + 2W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}\right]\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

$$\begin{aligned}
\ddot{\mathbf{L}}_i^{\sigma^2\tau}(\hat{\theta}) &= \left.\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\sigma^2\partial\boldsymbol{\tau}^T}\right|_{\theta=\hat{\theta}} \\
&= \frac{1}{2}\operatorname{tr}\left(\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\right) \\
&\quad +\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\hat{\boldsymbol{V}}_i^{-1}\left[W_g'(\hat{\delta}_i)\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_i^T\hat{\boldsymbol{V}}_i^{-1} + 2W_g(\hat{\delta}_i)\boldsymbol{I}_{2m_i}\right]\frac{\partial\boldsymbol{V}_i}{\partial\boldsymbol{\tau}^T}\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{r}}_i.
\end{aligned}$$

# Apêndice C

# Matriz de informação de Fisher

Nesta seção apresentamos os cálculos realizados para obter a matriz de informação de Fisher sob o modelo misto linear elíptico com erros de medição, apresentado na Seção 3.2.1. Os resultados seguintes são necessários para obter a matriz de informação e podem ser encontrados, por exemplo, em Graybill (1983), Mitchell (1989) e Fang et al. (1990).

Considere a distância de Mahalanobis escrita da seguinte forma:

$$\begin{aligned}
\delta_i &= (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW})^T \boldsymbol{V}_i^{-1/2} \boldsymbol{V}_i^{-1/2} (\boldsymbol{W}_i - \boldsymbol{\mu}_{iW}) \\
&= \boldsymbol{r}_i^T \boldsymbol{V}_i^{-1/2} \boldsymbol{V}_i^{-1/2} \boldsymbol{r}_i \\
&= \boldsymbol{z}_i^T \boldsymbol{z}_i \\
&= \|\boldsymbol{z}_i\|^2, \qquad \text{(C.1)}
\end{aligned}$$

em que $\|\boldsymbol{z}_i\|$ é a norma do vetor $\boldsymbol{z}_i = \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{r}_i$ e $\boldsymbol{z}_i \overset{\text{ind}}{\sim} El_{m_i}(\boldsymbol{0}, \boldsymbol{I}_{m_i}, g)$, $i = 1, ..., n$.

Além disso, vamos considerar os resultados

$$\begin{aligned}
\mathrm{E}\left\{W_g(\boldsymbol{\delta}_i)\|\boldsymbol{z}_i\|^2\right\} &= -\frac{m_i}{2}, && \text{(C.2)} \\
\mathrm{E}\left\{W_g^2(\boldsymbol{\delta}_i)\|\boldsymbol{z}_i\|^2\right\} &= d_{g_i} \text{ e} && \text{(C.3)} \\
\mathrm{E}\left\{W_g^2(\boldsymbol{\delta}_i)\|\boldsymbol{z}_i\|^4\right\} &= f_{g_i}. && \text{(C.4)}
\end{aligned}$$

Para o modelo normal, temos que $d_{g_i} = \frac{m_i}{2}$ e $f_{g_i} = m_i(m_i + 1)$ e para o modelo $t$ de Student estas quantidades ficam dadas por $d_{g_i} = \frac{m_i}{2} \frac{\nu + 2m_i}{\nu + 2m_i + 2}$ e $f_{g_i} = m_i(m_i + 1) \frac{\nu + 2m_i}{\nu + 2m_i + 2}$.

**Informação de Fisher para $\gamma$**

Considere a função escore para $\gamma$, com $\boldsymbol{z}_i = \boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{r}_i$:

$$U_i^{\gamma}(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{a}_i - W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right],$$

em que

$$\boldsymbol{a}_i = \operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right) \text{ e } \dot{\boldsymbol{A}}_i = \boldsymbol{V}_i^{-1/2}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1/2}.$$

Assim, a informação de Fisher para $\gamma$ é dada por

$$F_{\gamma\gamma,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\gamma}(\boldsymbol{\theta})U_i^{\gamma}(\boldsymbol{\theta})\right],$$

em que

$$\begin{aligned}
U_i^{\gamma}(\boldsymbol{\theta})U_i^{\gamma}(\boldsymbol{\theta}) &= \left\{-\frac{1}{2}\boldsymbol{a}_i - W_g(\boldsymbol{\delta}_i)\left[2\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right]\right\}^2\\
&= \frac{1}{4}\boldsymbol{a}_i^2 + 2\boldsymbol{a}_iW_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{a}_iW_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\\
&\quad +4W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\\
&\quad +4W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i + W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i,
\end{aligned}$$

e

$$\begin{aligned}
\mathbf{E}\left[U_i^{\gamma}(\boldsymbol{\theta})U_i^{\gamma}(\boldsymbol{\theta})\right] &= \frac{1}{4}\boldsymbol{a}_i^2 + 2\boldsymbol{a}_i\,\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\right]\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} + \boldsymbol{a}_i\,\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right]\\
&\quad +4\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&\quad +4\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right] + \mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right]\\
&= \frac{\boldsymbol{a}_i^2}{4}\left(\frac{f_{g_i}}{m_i(m_i+1)} - 1\right) + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)\\
&\quad +2d_{g_i}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}.
\end{aligned}$$

**Matriz de Informação de Fisher para $\boldsymbol{\beta}$**

Temos que

$$\begin{aligned} U_i^{\beta}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &= -2W_g(\boldsymbol{\delta}_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{V}_i^{-1/2}\boldsymbol{r}_i \\ &= -2W_g(\boldsymbol{\delta}_i)\begin{pmatrix} \boldsymbol{X}_i \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i. \end{aligned}$$

Logo,

$$\begin{aligned} U_i^{\beta_j}(\boldsymbol{\theta}) &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \beta_j} \\ &= -2W_g(\boldsymbol{\delta}_i)\begin{pmatrix} \mathbf{x}_{ij}^* \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i, \end{aligned}$$

em que $\mathbf{x}_{ij}^*$ denota a $j$-ésima coluna da matriz de planejamento $\boldsymbol{X}_i$ e $\mathbf{0}$ denota um vetor $m_i \times 1$ de zeros. Assim, a partição da matriz de informação de Fisher referente a $\beta_j$ e $\beta_l$ $(j, l = 1, ..., p)$ para o $i$-ésimo grupo é dada por

$$F_{\beta\beta,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\beta_j}(\boldsymbol{\theta})U_i^{\beta_l}(\boldsymbol{\theta})\right],$$

em que

$$\begin{aligned} U_i^{\beta_j}(\boldsymbol{\theta})U_i^{\beta_l}(\boldsymbol{\theta}) &= \left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix} \mathbf{x}_{ij}^* \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix} \mathbf{x}_{il}^* \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right] \\ &= 4W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix} \mathbf{x}_{ij}^* \\ \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{x}_{il}^* \\ \mathbf{0} \end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i, \end{aligned}$$

e

$$\begin{aligned}
\mathbf{E}\left[U_i^{\beta_j}(\boldsymbol{\theta})U_i^{\beta_l}(\boldsymbol{\theta})\right] &= 4\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{x}_{il}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&= 4\frac{d_{g_i}}{2m_i}\,\mathrm{tr}\left[\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{x}_{il}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\right]\\
&= \frac{2d_{g_i}}{m_i}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}.
\end{aligned}$$

**Informação de Fisher para $\mu$**

Temos que

$$\begin{aligned}
U_i^{\mu}(\boldsymbol{\theta}) &= -2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\\
&= -2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i.
\end{aligned}$$

A informação de Fisher para $\mu$ é dada por

$$F_{\mu\mu,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\mu}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})\right],$$

em que

$$\begin{aligned}
U_i^{\mu}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta}) &= \left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&= 4W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i,
\end{aligned}$$

e

$$\begin{aligned}
\mathbf{E}\left[U_i^{\mu}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})\right] &= 4\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&= \frac{2d_{g_i}}{m_i}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}.
\end{aligned}$$

**Informação de Fisher para $\sigma_u^2$**

Consideremos

$$U_i^{\sigma_u^2}(\boldsymbol{\theta}) = -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i,$$

em que: $\dot{\boldsymbol{B}}_i = \boldsymbol{V}_i^{-1/2}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1/2}$.

A partição da informação de Fisher para $\sigma_u^2$ é dada por

$$F_{\sigma_u^2\sigma_u^2,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})^T\right],$$

em que

$$\begin{aligned}
U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})^T &= \left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right]\left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)\right. \\
&\quad \left. -W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right] \\
&= \frac{1}{4}\operatorname{tr}^2\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) + \operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i \\
&\quad +W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i,
\end{aligned}$$

e

$$\begin{aligned}
\mathbf{E}\left[U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}^2\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) + \operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right] \\
&\quad +\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right] \\
&= \frac{1}{4}\operatorname{tr}^2\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)} - 1\right] \\
&\quad +\frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right).
\end{aligned}$$

**Informação de Fisher para $\sigma^2$**

De forma análoga, temos que

$$F_{\sigma^2\sigma^2,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T\right],$$

em que

$$U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T = \frac{1}{4}\operatorname{tr}^2\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right)+\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{r}_i+W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i,$$

com $\dot{\boldsymbol{C}}_i = \boldsymbol{V}_i^{-1/2}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1/2}$,
e

$$\begin{aligned}\mathbf{E}\left[U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}^2\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)}-1\right] \\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right).\end{aligned}$$

**Matriz de informação de Fisher para $\boldsymbol{\tau}$**

$$F_{\tau\tau,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\tau}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T\right],$$

em que

$$\begin{aligned}U_i^{\tau_j}(\boldsymbol{\theta})U_i^{\tau_l}(\boldsymbol{\theta})^T &= \left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i\right]\left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right)\right. \\ &\quad \left. -W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{l,i}\boldsymbol{z}_i\right] \\ &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right) + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{l,i}\boldsymbol{z}_i \\ &\quad + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i + W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{l,i}\boldsymbol{z}_i,\end{aligned}$$

com $\dot{\boldsymbol{G}}_{j,i} = \boldsymbol{V}_i^{-1/2}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\boldsymbol{V}_i^{-1/2}$ e $\dot{\boldsymbol{G}}_{l,i} = \boldsymbol{V}_i^{-1/2}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\boldsymbol{V}_i^{-1/2}$,
e

$$\begin{aligned}\mathbf{E}\left[U_i^{\tau_j}(\boldsymbol{\theta})U_i^{\tau_l}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right) + \operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{l,i}\boldsymbol{z}_i\right] \\ &\quad + \operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right)\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i\right] + \mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{l,i}\boldsymbol{z}_i\right] \\ &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)}-1\right] \\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_l}\right).\end{aligned}$$

**Matriz de informação de Fisher para $\gamma$ e $\boldsymbol{\beta}$**

$$F_{\gamma\beta,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\gamma}(\boldsymbol{\theta})U_i^{\beta}(\boldsymbol{\theta})\right],$$

em que

$$\begin{aligned}
U_i^{\gamma}(\boldsymbol{\theta})U_i^{\beta}(\boldsymbol{\theta}) &= \left[-\frac{1}{2}\boldsymbol{a}_i - 2W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right]\\
&\quad\times\left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&= \boldsymbol{a}_iW_g(\boldsymbol{\delta}_i)\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i + 4W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\\
&\quad\times\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i + 2W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i,
\end{aligned}$$

e

$$\begin{aligned}
\mathbf{E}\left[U_i^{\gamma}(\boldsymbol{\theta})U_i^{\beta}(\boldsymbol{\theta})^T\right] &= \boldsymbol{a}_i\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\,\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i\right]\\
&\quad+4\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&\quad+2\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&= 4\frac{d_{g_i}}{2m_i}\,\mathrm{tr}\left[\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\right]\\
&= \frac{2d_{g_i}}{m_i}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}.
\end{aligned}$$

**Informação de Fisher para $\gamma$ e $\mu$**

$$F_{\gamma\mu,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\gamma}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})\right],$$

em que

$$
\begin{aligned}
U_i^{\gamma}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta}) &= \left[-\frac{1}{2}\boldsymbol{a}_i - 2W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right]\\
&\quad\times\left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&= \boldsymbol{a}_iW_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\\
&\quad+4W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\\
&\quad+2W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T\boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i,
\end{aligned}
$$

e

$$
\mathbf{E}\left[U_i^{\gamma}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})^T\right] = \frac{2d_{g_i}}{m_i}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}^T\boldsymbol{V}_i^{-1}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}.
$$

**Matriz de informação de Fisher para $\gamma$ e $\sigma_u^2$**

$$
F_{\gamma\sigma_u^2,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})^T\right],
$$

em que

$$
\begin{aligned}
U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})^T &= \left[-\frac{1}{2}\boldsymbol{a}_i - 2W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right]\times\\
&\quad\times\left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right]\\
&= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right) + W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)\\
&\quad+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\gamma}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i + 2W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\\
&\quad+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i + W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i
\end{aligned}
$$

e

$$
\begin{aligned}
\mathbf{E}\left[U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma_u^2}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) \\
&\quad + \mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\right]\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) \\
&\quad + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right] \\
&\quad + 2\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right] \\
&\quad + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right] + \mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right] \\
&= -\frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right) \\
&\quad + \frac{f_{g_i}}{4m_i(m_i+1)}\left[2\operatorname{tr}\left(\dot{\boldsymbol{A}}_i\dot{\boldsymbol{C}}_i\right) + \operatorname{tr}\left(\dot{\boldsymbol{A}}_i\right)\operatorname{tr}\left(\dot{\boldsymbol{C}}_i\right)\right] \\
&= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)} - 1\right] \\
&\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right).
\end{aligned}
$$

**Matriz de informação de Fisher para $\gamma$ e $\sigma^2$**

$$
F_{\gamma\sigma^2,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T\right],
$$

em que

$$
\begin{aligned}
U_i^{\gamma}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T &= \left[-\frac{1}{2}\boldsymbol{a}_i - 2W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right] \\
&\quad \times\left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i\right] \\
&= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) + W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right) \\
&\quad + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i + 2W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i \\
&\quad + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i + W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i
\end{aligned}
$$

e

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

**Matriz de informação de Fisher para $\gamma$ e $\boldsymbol{\tau}$**

$$F_{\gamma\tau,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\gamma}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T\right],$$

em que

$$\begin{aligned}U_i^{\gamma}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T &= \left[-\frac{1}{2}\boldsymbol{a}_i - 2W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix} - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\right] \\ &\quad \times\left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i\right] \\ &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right) + W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right) \\ &\quad + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i + 2W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mu\mathbf{1}_{m_i}\\ \mathbf{0}\end{pmatrix}\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i \\ &\quad + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i + W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{A}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{G}}_{j,i}\boldsymbol{z}_i\end{aligned}$$

e

$$\begin{aligned}\mathbf{E}\left[U_i^{\gamma}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)}-1\right] \\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \gamma}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right).\end{aligned}$$

**Matriz de Informação de Fisher para $\boldsymbol{\beta}$ e $\mu$**

$$F_{\beta\mu,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\beta}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})^T\right],$$

em que

$$
\begin{aligned}
U_i^{\beta}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})^T &= \left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\left[-2W_g(\boldsymbol{\delta}_i)\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]^T\\
&= 4W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i
\end{aligned}
$$

e

$$
\begin{aligned}
\mathbf{E}\left[U_i^{\beta}(\boldsymbol{\theta})U_i^{\mu}(\boldsymbol{\theta})^T\right] &= 4\,\mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\boldsymbol{V}_i^{-1/2}\begin{pmatrix}\mathbf{x}_{ij}^*\\ \mathbf{0}\end{pmatrix}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}^T \boldsymbol{V}_i^{-1/2}\boldsymbol{z}_i\right]\\
&= \frac{2d_{g_i}}{m_i}\begin{pmatrix}\boldsymbol{X}_i\\ \mathbf{0}\end{pmatrix}^T \boldsymbol{V}_i^{-1}\begin{pmatrix}\gamma\mathbf{1}_{m_i}\\ \mathbf{1}_{m_i}\end{pmatrix}
\end{aligned}
$$

**Informação de Fisher para $\sigma_u^2$ e $\sigma^2$**

$$
F_{\sigma_u^2\sigma^2,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T\right],
$$

em que

$$
\begin{aligned}
U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T &= \left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right]\left[-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right) - W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i\right]\\
&= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right) + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i\\
&\quad+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right)W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i + W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i,
\end{aligned}
$$

e

$$
\begin{aligned}
\mathbf{E}\left[U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\sigma^2}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right) + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i\right]\\
&\quad\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right)\mathbf{E}\left[W_g(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\right] + \mathbf{E}\left[W_g^2(\boldsymbol{\delta}_i)\boldsymbol{z}_i^T\dot{\boldsymbol{B}}_i\boldsymbol{z}_i\boldsymbol{z}_i^T\dot{\boldsymbol{C}}_i\boldsymbol{z}_i\right]\\
&= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)} - 1\right]\\
&\quad+\frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial\boldsymbol{V}_i}{\partial\sigma^2}\right).
\end{aligned}
$$

De forma análoga aos cálculos realizados para obtenção da informação de Fisher para $\sigma_u^2$ e $\sigma^2$, temos que

**Matriz de informação de Fisher para $\sigma_u^2$ e $\boldsymbol{\tau}$**

$$F_{\sigma_u^2\tau,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T\right],$$

em que

$$\begin{aligned}\mathbf{E}\left[U_i^{\sigma_u^2}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)}-1\right] \\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma_u^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right).\end{aligned}$$

**Matriz de informação de Fisher para $\sigma^2$ e $\boldsymbol{\tau}$**

$$F_{\sigma^2\tau,i}(\boldsymbol{\theta}) = \mathbf{E}\left[\frac{1}{n}U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T\right],$$

em que

$$\begin{aligned}\mathbf{E}\left[U_i^{\sigma^2}(\boldsymbol{\theta})U_i^{\tau}(\boldsymbol{\theta})^T\right] &= \frac{1}{4}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\right)\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right)\left[\frac{f_{g_i}}{m_i(m_i+1)}-1\right] \\ &\quad + \frac{f_{g_i}}{2m_i(m_i+1)}\operatorname{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \tau_j}\right).\end{aligned}$$

Para algumas distribuições pertencentes à classe das distribuições elípticas as quantidades $d_{g_i}$ e $f_{g_i}$ que aparecem nas expressões acima têm forma fechada, como é o caso das distribuições normal, $t$ de Student e exponencial potência. Para outras distribuições, como a normal contaminada e logísticas tipo I e II, as quantidades $d_{g_i}$ e $f_{g_i}$ devem ser calculadas mediante algum método de aproximação.

# Apêndice D

# Dados reduzidos dos setores censitários de Boston

A seguir são apresentados os dados do exemplo de aplicação, que também foram utilizados na aplicação, apresentada na Seção 5.2.

Tabela D.1: Apresentação dos dados de 132 setores censitários de 15 distritos da cidade de Boston

| #Setor | Setor | CRIM | ZN | INDUS | CHAS | NOX | RM | AGE |
|---|---|---|---|---|---|---|---|---|
| 357 | 75 | 8,98296 | 0 | 18,1 | 1 | 0,77 | 6,212 | 97,4 |
| 358 | 75 | 3,8497 | 0 | 18,1 | 1 | 0,77 | 6,395 | 91 |
| 359 | 75 | 5,20177 | 0 | 18,1 | 1 | 0,77 | 6,127 | 83,4 |
| 360 | 75 | 4,26131 | 0 | 18,1 | 0 | 0,77 | 6,112 | 81,3 |
| 361 | 75 | 4,54192 | 0 | 18,1 | 0 | 0,77 | 6,398 | 88 |
| 362 | 75 | 3,83684 | 0 | 18,1 | 0 | 0,77 | 6,251 | 91,1 |
| 363 | 75 | 3,67822 | 0 | 18,1 | 0 | 0,77 | 5,362 | 96,2 |
| 364 | 75 | 4,22239 | 0 | 18,1 | 1 | 0,77 | 5,803 | 89 |
| 365 | 76 | 3,47428 | 0 | 18,1 | 1 | 0,718 | 8,78 | 82,9 |
| 366 | 76 | 4,55587 | 0 | 18,1 | 0 | 0,718 | 3,561 | 87,9 |
| 367 | 76 | 3,69695 | 0 | 18,1 | 0 | 0,718 | 4,963 | 91,4 |
| 368 | 76 | 13,5222 | 0 | 18,1 | 0 | 0,631 | 3,863 | 100 |
| 369 | 76 | 4,89822 | 0 | 18,1 | 0 | 0,631 | 4,97 | 100 |
| 370 | 76 | 5,66998 | 0 | 18,1 | 1 | 0,631 | 6,683 | 96,8 |
| 371 | 77 | 6,53876 | 0 | 18,1 | 1 | 0,631 | 7,016 | 97,5 |
| 372 | 77 | 9,2323 | 0 | 18,1 | 0 | 0,631 | 6,216 | 100 |
| 373 | 77 | 8,26725 | 0 | 18,1 | 1 | 0,668 | 5,875 | 89,6 |
| 374 | 78 | 11,1081 | 0 | 18,1 | 0 | 0,668 | 4,906 | 100 |
| 375 | 78 | 18,4982 | 0 | 18,1 | 0 | 0,668 | 4,138 | 100 |
| 376 | 79 | 19,6091 | 0 | 18,1 | 0 | 0,671 | 7,313 | 97,9 |
| 377 | 79 | 15,288 | 0 | 18,1 | 0 | 0,671 | 6,649 | 93,3 |
| 378 | 79 | 9,82349 | 0 | 18,1 | 0 | 0,671 | 6,794 | 98,8 |
| 379 | 79 | 23,6482 | 0 | 18,1 | 0 | 0,671 | 6,38 | 96,2 |
| 380 | 79 | 17,8667 | 0 | 18,1 | 0 | 0,671 | 6,223 | 100 |
| 381 | 79 | 88,9762 | 0 | 18,1 | 0 | 0,671 | 6,968 | 91,9 |
| 382 | 79 | 15,8744 | 0 | 18,1 | 0 | 0,671 | 6,545 | 99,1 |
| 383 | 80 | 9,18702 | 0 | 18,1 | 0 | 0,7 | 5,536 | 100 |
| 384 | 80 | 7,99248 | 0 | 18,1 | 0 | 0,7 | 5,52 | 100 |
| 385 | 80 | 20,0849 | 0 | 18,1 | 0 | 0,7 | 4,368 | 91,2 |
| 386 | 80 | 16,8118 | 0 | 18,1 | 0 | 0,7 | 5,277 | 98,1 |
| 387 | 80 | 24,3938 | 0 | 18,1 | 0 | 0,7 | 4,652 | 100 |
| 388 | 80 | 22,5971 | 0 | 18,1 | 0 | 0,7 | 5 | 89,5 |
| 389 | 80 | 14,3337 | 0 | 18,1 | 0 | 0,7 | 4,88 | 100 |
| 390 | 80 | 8,15174 | 0 | 18,1 | 0 | 0,7 | 5,39 | 98,9 |
| 391 | 80 | 6,96215 | 0 | 18,1 | 0 | 0,7 | 5,713 | 97 |
| 392 | 80 | 5,29305 | 0 | 18,1 | 0 | 0,7 | 6,051 | 82,5 |
| 393 | 80 | 11,5779 | 0 | 18,1 | 0 | 0,7 | 5,036 | 97 |
| 394 | 81 | 8,64476 | 0 | 18,1 | 0 | 0,693 | 6,193 | 92,6 |

| #Setor | Setor | CRIM | ZN | INDUS | CHAS | NOX | RM | AGE |
|---|---|---|---|---|---|---|---|---|
| 395 | 81 | 13,3598 | 0 | 18,1 | 0 | 0,693 | 5,887 | 94,7 |
| 396 | 81 | 8,71675 | 0 | 18,1 | 0 | 0,693 | 6,471 | 98,8 |
| 397 | 81 | 5,87205 | 0 | 18,1 | 0 | 0,693 | 6,405 | 96 |
| 398 | 81 | 7,67202 | 0 | 18,1 | 0 | 0,693 | 5,747 | 98,9 |
| 399 | 81 | 38,3518 | 0 | 18,1 | 0 | 0,693 | 5,453 | 100 |
| 400 | 81 | 9,91655 | 0 | 18,1 | 0 | 0,693 | 5,852 | 77,8 |
| 401 | 81 | 25,0461 | 0 | 18,1 | 0 | 0,693 | 5,987 | 100 |
| 402 | 81 | 14,2362 | 0 | 18,1 | 0 | 0,693 | 6,343 | 100 |
| 403 | 81 | 9,59571 | 0 | 18,1 | 0 | 0,693 | 6,404 | 100 |
| 404 | 81 | 24,8017 | 0 | 18,1 | 0 | 0,693 | 5,349 | 96 |
| 405 | 81 | 41,5292 | 0 | 18,1 | 0 | 0,693 | 5,531 | 85,4 |
| 406 | 81 | 67,9208 | 0 | 18,1 | 0 | 0,693 | 5,683 | 100 |
| 407 | 82 | 20,7162 | 0 | 18,1 | 0 | 0,659 | 4,138 | 100 |
| 408 | 82 | 11,9511 | 0 | 18,1 | 0 | 0,659 | 5,608 | 100 |
| 409 | 82 | 7,40389 | 0 | 18,1 | 0 | 0,597 | 5,617 | 97,9 |
| 410 | 82 | 14,4383 | 0 | 18,1 | 0 | 0,597 | 6,852 | 100 |
| 411 | 82 | 51,1358 | 0 | 18,1 | 0 | 0,597 | 5,757 | 100 |
| 412 | 82 | 14,0507 | 0 | 18,1 | 0 | 0,597 | 6,657 | 100 |
| 413 | 82 | 18,811 | 0 | 18,1 | 0 | 0,597 | 4,628 | 100 |
| 414 | 82 | 28,6558 | 0 | 18,1 | 0 | 0,597 | 5,155 | 100 |
| 415 | 83 | 45,7461 | 0 | 18,1 | 0 | 0,693 | 4,519 | 100 |
| 416 | 83 | 18,0846 | 0 | 18,1 | 0 | 0,679 | 6,434 | 100 |
| 417 | 83 | 10,8342 | 0 | 18,1 | 0 | 0,679 | 6,782 | 90,8 |
| 418 | 83 | 25,9406 | 0 | 18,1 | 0 | 0,679 | 5,304 | 89,1 |
| 419 | 83 | 73,5341 | 0 | 18,1 | 0 | 0,679 | 5,957 | 100 |
| 420 | 83 | 11,8123 | 0 | 18,1 | 0 | 0,718 | 6,824 | 76,5 |
| 421 | 83 | 11,0874 | 0 | 18,1 | 0 | 0,718 | 6,411 | 100 |
| 422 | 83 | 7,02259 | 0 | 18,1 | 0 | 0,718 | 6,006 | 95,3 |
| 423 | 83 | 12,0482 | 0 | 18,1 | 0 | 0,614 | 5,648 | 87,6 |
| 424 | 83 | 7,05042 | 0 | 18,1 | 0 | 0,614 | 6,103 | 85,1 |
| 425 | 83 | 8,79212 | 0 | 18,1 | 0 | 0,584 | 5,565 | 70,6 |
| 426 | 83 | 15,8603 | 0 | 18,1 | 0 | 0,679 | 5,896 | 95,4 |
| 427 | 83 | 12,2472 | 0 | 18,1 | 0 | 0,584 | 5,837 | 59,7 |
| 428 | 83 | 37,6619 | 0 | 18,1 | 0 | 0,679 | 6,202 | 78,7 |
| 429 | 83 | 7,36711 | 0 | 18,1 | 0 | 0,679 | 6,193 | 78,1 |
| 430 | 83 | 9,33889 | 0 | 18,1 | 0 | 0,679 | 6,38 | 95,6 |
| 431 | 83 | 8,49213 | 0 | 18,1 | 0 | 0,584 | 6,348 | 86,1 |
| 432 | 83 | 10,0623 | 0 | 18,1 | 0 | 0,584 | 6,833 | 94,3 |
| 433 | 83 | 6,44405 | 0 | 18,1 | 0 | 0,584 | 6,425 | 74,8 |

| #Setor | Setor | CRIM | ZN | INDUS | CHAS | NOX | RM | AGE |
|---|---|---|---|---|---|---|---|---|
| 442 | 84 | 9,72418 | 0 | 18,1 | 0 | 0,74 | 6,406 | 97,2 |
| 443 | 84 | 5,66637 | 0 | 18,1 | 0 | 0,74 | 6,219 | 100 |
| 444 | 84 | 9,96654 | 0 | 18,1 | 0 | 0,74 | 6,485 | 100 |
| 445 | 84 | 12,8023 | 0 | 18,1 | 0 | 0,74 | 5,854 | 96,6 |
| 446 | 84 | 10,6718 | 0 | 18,1 | 0 | 0,74 | 6,459 | 94,8 |
| 447 | 84 | 6,28807 | 0 | 18,1 | 0 | 0,74 | 6,341 | 96,4 |
| 448 | 84 | 9,92485 | 0 | 18,1 | 0 | 0,74 | 6,251 | 96,6 |
| 449 | 84 | 9,32909 | 0 | 18,1 | 0 | 0,713 | 6,185 | 98,7 |
| 450 | 84 | 7,52601 | 0 | 18,1 | 0 | 0,713 | 6,417 | 98,3 |
| 451 | 84 | 6,71772 | 0 | 18,1 | 0 | 0,713 | 6,749 | 92,6 |
| 452 | 84 | 5,44114 | 0 | 18,1 | 0 | 0,713 | 6,655 | 98,2 |
| 453 | 84 | 5,09017 | 0 | 18,1 | 0 | 0,713 | 6,297 | 91,8 |
| 454 | 84 | 8,24809 | 0 | 18,1 | 0 | 0,713 | 7,393 | 99,3 |
| 455 | 84 | 9,51363 | 0 | 18,1 | 0 | 0,713 | 6,728 | 94,1 |
| 456 | 84 | 4,75237 | 0 | 18,1 | 0 | 0,713 | 6,525 | 86,5 |
| 457 | 85 | 4,66883 | 0 | 18,1 | 0 | 0,713 | 5,976 | 87,9 |
| 458 | 85 | 8,20058 | 0 | 18,1 | 0 | 0,713 | 5,936 | 80,3 |
| 459 | 85 | 7,75223 | 0 | 18,1 | 0 | 0,713 | 6,301 | 83,7 |
| 460 | 85 | 6,80117 | 0 | 18,1 | 0 | 0,713 | 6,081 | 84,4 |
| 461 | 85 | 4,81213 | 0 | 18,1 | 0 | 0,713 | 6,701 | 90 |
| 462 | 85 | 3,69311 | 0 | 18,1 | 0 | 0,713 | 6,376 | 88,4 |
| 463 | 85 | 6,65492 | 0 | 18,1 | 0 | 0,713 | 6,317 | 83 |
| 464 | 85 | 5,82115 | 0 | 18,1 | 0 | 0,713 | 6,513 | 89,9 |
| 465 | 85 | 7,83932 | 0 | 18,1 | 0 | 0,655 | 6,209 | 65,4 |
| 466 | 85 | 3,1636 | 0 | 18,1 | 0 | 0,655 | 5,759 | 48,2 |
| 467 | 85 | 3,77498 | 0 | 18,1 | 0 | 0,655 | 5,952 | 84,7 |
| 468 | 86 | 4,42228 | 0 | 18,1 | 0 | 0,584 | 6,003 | 94,5 |
| 469 | 86 | 15,5757 | 0 | 18,1 | 0 | 0,58 | 5,926 | 71 |
| 470 | 86 | 13,0751 | 0 | 18,1 | 0 | 0,58 | 5,713 | 56,7 |
| 471 | 86 | 4,34879 | 0 | 18,1 | 0 | 0,58 | 6,167 | 84 |
| 472 | 86 | 4,03841 | 0 | 18,1 | 0 | 0,532 | 6,229 | 90,7 |
| 473 | 86 | 3,56868 | 0 | 18,1 | 0 | 0,58 | 6,437 | 75 |
| 474 | 87 | 4,64689 | 0 | 18,1 | 0 | 0,614 | 6,98 | 67,6 |
| 475 | 87 | 8,05579 | 0 | 18,1 | 0 | 0,584 | 5,427 | 95,4 |
| 476 | 87 | 6,39312 | 0 | 18,1 | 0 | 0,584 | 6,162 | 97,4 |
| 477 | 87 | 4,87141 | 0 | 18,1 | 0 | 0,614 | 6,484 | 93,6 |
| 478 | 87 | 15,0234 | 0 | 18,1 | 0 | 0,614 | 5,304 | 97,3 |
| 479 | 87 | 10,233 | 0 | 18,1 | 0 | 0,614 | 6,185 | 96,7 |
| 480 | 87 | 14,3337 | 0 | 18,1 | 0 | 0,614 | 6,229 | 88 |

| #Setor | Setor | DIS | RAD | TAX | PTRATIO | BLACK | LSTAT | LMV |
|---|---|---|---|---|---|---|---|---|
| 357 | 75 | 2,1222 | 24 | 666 | 20,2 | 377,73 | 17,6 | 17,8 |
| 358 | 75 | 2,5052 | 24 | 666 | 20,2 | 391,34 | 13,27 | 21,7 |
| 359 | 75 | 2,7227 | 24 | 666 | 20,2 | 395,43 | 11,48 | 22,7 |
| 360 | 75 | 2,5091 | 24 | 666 | 20,2 | 390,74 | 12,67 | 22,6 |
| 361 | 75 | 2,5182 | 24 | 666 | 20,2 | 374,56 | 7,79 | 25 |
| 362 | 75 | 2,2955 | 24 | 666 | 20,2 | 350,65 | 14,19 | 19,9 |
| 363 | 75 | 2,1036 | 24 | 666 | 20,2 | 380,79 | 10,19 | 20,8 |
| 364 | 75 | 1,9047 | 24 | 666 | 20,2 | 353,04 | 14,64 | 16,8 |
| 365 | 76 | 1,9047 | 24 | 666 | 20,2 | 354,55 | 5,29 | 21,9 |
| 366 | 76 | 1,6132 | 24 | 666 | 20,2 | 354,7 | 7,12 | 27,5 |
| 367 | 76 | 1,7523 | 24 | 666 | 20,2 | 316,03 | 14 | 21,9 |
| 368 | 76 | 1,5106 | 24 | 666 | 20,2 | 131,42 | 13,33 | 23,1 |
| 369 | 76 | 1,3325 | 24 | 666 | 20,2 | 375,52 | 3,26 | 50 |
| 370 | 76 | 1,3567 | 24 | 666 | 20,2 | 375,33 | 3,73 | 50 |
| 371 | 77 | 1,2024 | 24 | 666 | 20,2 | 392,05 | 2,96 | 50 |
| 372 | 77 | 1,1691 | 24 | 666 | 20,2 | 366,15 | 9,53 | 50 |
| 373 | 77 | 1,1296 | 24 | 666 | 20,2 | 347,88 | 8,88 | 50 |
| 374 | 78 | 1,1742 | 24 | 666 | 20,2 | 396,9 | 34,77 | 13,8 |
| 375 | 78 | 1,137 | 24 | 666 | 20,2 | 396,9 | 37,97 | 13,8 |
| 376 | 79 | 1,3163 | 24 | 666 | 20,2 | 396,9 | 13,44 | 15 |
| 377 | 79 | 1,3449 | 24 | 666 | 20,2 | 363,02 | 23,24 | 13,9 |
| 378 | 79 | 1,358 | 24 | 666 | 20,2 | 396,9 | 21,24 | 13,3 |
| 379 | 79 | 1,3861 | 24 | 666 | 20,2 | 396,9 | 23,69 | 13,1 |
| 380 | 79 | 1,3861 | 24 | 666 | 20,2 | 393,74 | 21,78 | 10,2 |
| 381 | 79 | 1,4165 | 24 | 666 | 20,2 | 396,9 | 17,21 | 10,4 |
| 382 | 79 | 1,5192 | 24 | 666 | 20,2 | 396,9 | 21,08 | 10,9 |
| 383 | 80 | 1,5804 | 24 | 666 | 20,2 | 396,9 | 23,6 | 11,3 |
| 384 | 80 | 1,5331 | 24 | 666 | 20,2 | 396,9 | 24,56 | 12,3 |
| 385 | 80 | 1,4395 | 24 | 666 | 20,2 | 285,83 | 30,63 | 8,8 |
| 386 | 80 | 1,4261 | 24 | 666 | 20,2 | 396,9 | 30,81 | 7,2 |
| 387 | 80 | 1,4672 | 24 | 666 | 20,2 | 396,9 | 28,28 | 10,5 |
| 388 | 80 | 1,5184 | 24 | 666 | 20,2 | 396,9 | 31,99 | 7,4 |
| 389 | 80 | 1,5895 | 24 | 666 | 20,2 | 372,92 | 30,62 | 10,2 |
| 390 | 80 | 1,7281 | 24 | 666 | 20,2 | 396,9 | 20,85 | 11,5 |
| 391 | 80 | 1,9265 | 24 | 666 | 20,2 | 394,43 | 17,11 | 15,1 |
| 392 | 80 | 2,1678 | 24 | 666 | 20,2 | 378,38 | 18,76 | 23,2 |
| 393 | 80 | 1,77 | 24 | 666 | 20,2 | 396,9 | 25,68 | 9,7 |
| 394 | 81 | 1,7912 | 24 | 666 | 20,2 | 396,9 | 15,17 | 13,8 |

| #Setor | Setor | DIS | RAD | TAX | PTRATIO | BLACK | LSTAT | LMV |
|---|---|---|---|---|---|---|---|---|
| 395 | 81 | 1,7821 | 24 | 666 | 20,2 | 396,9 | 16,35 | 12,7 |
| 396 | 81 | 1,7257 | 24 | 666 | 20,2 | 391,98 | 17,12 | 13,1 |
| 397 | 81 | 1,6768 | 24 | 666 | 20,2 | 396,9 | 19,37 | 12,5 |
| 398 | 81 | 1,6334 | 24 | 666 | 20,2 | 393,1 | 19,92 | 8,5 |
| 399 | 81 | 1,4896 | 24 | 666 | 20,2 | 396,9 | 30,59 | 5 |
| 400 | 81 | 1,5004 | 24 | 666 | 20,2 | 338,16 | 29,97 | 6,3 |
| 401 | 81 | 1,5888 | 24 | 666 | 20,2 | 396,9 | 26,77 | 5,6 |
| 402 | 81 | 1,5741 | 24 | 666 | 20,2 | 396,9 | 20,32 | 7,2 |
| 403 | 81 | 1,639 | 24 | 666 | 20,2 | 376,11 | 20,31 | 12,1 |
| 404 | 81 | 1,7028 | 24 | 666 | 20,2 | 396,9 | 19,77 | 8,3 |
| 405 | 81 | 1,6074 | 24 | 666 | 20,2 | 329,46 | 27,38 | 8,5 |
| 406 | 81 | 1,4254 | 24 | 666 | 20,2 | 384,97 | 22,98 | 5 |
| 407 | 82 | 1,1781 | 24 | 666 | 20,2 | 370,22 | 23,34 | 11,9 |
| 408 | 82 | 1,2852 | 24 | 666 | 20,2 | 332,09 | 12,13 | 27,9 |
| 409 | 82 | 1,4547 | 24 | 666 | 20,2 | 314,64 | 26,4 | 17,2 |
| 410 | 82 | 1,4655 | 24 | 666 | 20,2 | 179,36 | 19,78 | 27,5 |
| 411 | 82 | 1,413 | 24 | 666 | 20,2 | 2,6 | 10,11 | 15 |
| 412 | 82 | 1,5275 | 24 | 666 | 20,2 | 35,05 | 21,22 | 17,2 |
| 413 | 82 | 1,5539 | 24 | 666 | 20,2 | 28,79 | 34,37 | 17,9 |
| 414 | 82 | 1,5894 | 24 | 666 | 20,2 | 210,97 | 20,08 | 16,3 |
| 415 | 83 | 1,6582 | 24 | 666 | 20,2 | 88,27 | 36,98 | 7 |
| 416 | 83 | 1,8347 | 24 | 666 | 20,2 | 27,25 | 29,05 | 7,2 |
| 417 | 83 | 1,8195 | 24 | 666 | 20,2 | 21,57 | 25,79 | 7,5 |
| 418 | 83 | 1,6475 | 24 | 666 | 20,2 | 127,36 | 26,64 | 10,4 |
| 419 | 83 | 1,8026 | 24 | 666 | 20,2 | 16,45 | 20,62 | 8,8 |
| 420 | 83 | 1,794 | 24 | 666 | 20,2 | 48,45 | 22,74 | 8,4 |
| 421 | 83 | 1,8589 | 24 | 666 | 20,2 | 318,75 | 15,02 | 16,7 |
| 422 | 83 | 1,8746 | 24 | 666 | 20,2 | 319,98 | 15,7 | 14,2 |
| 423 | 83 | 1,9512 | 24 | 666 | 20,2 | 291,55 | 14,1 | 20,8 |
| 424 | 83 | 2,0218 | 24 | 666 | 20,2 | 2,52 | 23,29 | 13,4 |
| 425 | 83 | 2,0635 | 24 | 666 | 20,2 | 3,65 | 17,16 | 11,7 |
| 426 | 83 | 1,9096 | 24 | 666 | 20,2 | 7,68 | 24,39 | 8,3 |
| 427 | 83 | 1,9976 | 24 | 666 | 20,2 | 24,65 | 15,69 | 10,2 |
| 428 | 83 | 1,8629 | 24 | 666 | 20,2 | 18,82 | 14,52 | 10,9 |
| 429 | 83 | 1,9356 | 24 | 666 | 20,2 | 96,73 | 21,52 | 11 |
| 430 | 83 | 1,9682 | 24 | 666 | 20,2 | 60,72 | 24,08 | 9,5 |
| 431 | 83 | 2,0527 | 24 | 666 | 20,2 | 83,45 | 17,64 | 14,5 |
| 432 | 83 | 2,0882 | 24 | 666 | 20,2 | 81,33 | 19,69 | 14,1 |
| 433 | 83 | 2,2004 | 24 | 666 | 20,2 | 97,95 | 12,03 | 16,1 |

| #Setor | Setor | DIS | RAD | TAX | PTRATIO | BLACK | LSTAT | LMV |
|---|---|---|---|---|---|---|---|---|
| 442 | 84 | 2,0651 | 24 | 666 | 20,2 | 385,96 | 19,52 | 17,1 |
| 443 | 84 | 2,0048 | 24 | 666 | 20,2 | 395,69 | 16,59 | 18,4 |
| 444 | 84 | 1,9784 | 24 | 666 | 20,2 | 386,73 | 18,85 | 15,4 |
| 445 | 84 | 1,8956 | 24 | 666 | 20,2 | 240,52 | 23,79 | 10,8 |
| 446 | 84 | 1,9879 | 24 | 666 | 20,2 | 43,06 | 23,98 | 11,8 |
| 447 | 84 | 2,072 | 24 | 666 | 20,2 | 318,01 | 17,79 | 14,9 |
| 448 | 84 | 2,198 | 24 | 666 | 20,2 | 388,52 | 16,44 | 12,6 |
| 449 | 84 | 2,2616 | 24 | 666 | 20,2 | 396,9 | 18,13 | 14,1 |
| 450 | 84 | 2,185 | 24 | 666 | 20,2 | 304,21 | 19,31 | 13 |
| 451 | 84 | 2,3236 | 24 | 666 | 20,2 | 0,32 | 17,44 | 13,4 |
| 452 | 84 | 2,3552 | 24 | 666 | 20,2 | 355,29 | 17,73 | 15,2 |
| 453 | 84 | 2,3682 | 24 | 666 | 20,2 | 385,09 | 17,27 | 16,1 |
| 454 | 84 | 2,4527 | 24 | 666 | 20,2 | 375,87 | 16,74 | 17,8 |
| 455 | 84 | 2,4961 | 24 | 666 | 20,2 | 6,68 | 18,71 | 14,9 |
| 456 | 84 | 2,4358 | 24 | 666 | 20,2 | 50,92 | 18,13 | 14,1 |
| 457 | 85 | 2,5806 | 24 | 666 | 20,2 | 10,48 | 19,01 | 12,7 |
| 458 | 85 | 2,7792 | 24 | 666 | 20,2 | 3,5 | 16,94 | 13,5 |
| 459 | 85 | 2,7831 | 24 | 666 | 20,2 | 272,21 | 16,23 | 14,9 |
| 460 | 85 | 2,7175 | 24 | 666 | 20,2 | 396,9 | 14,7 | 20 |
| 461 | 85 | 2,5975 | 24 | 666 | 20,2 | 255,23 | 16,42 | 16,4 |
| 462 | 85 | 2,5671 | 24 | 666 | 20,2 | 391,43 | 14,65 | 17,7 |
| 463 | 85 | 2,7344 | 24 | 666 | 20,2 | 396,9 | 13,99 | 19,5 |
| 464 | 85 | 2,8016 | 24 | 666 | 20,2 | 393,82 | 10,29 | 20,2 |
| 465 | 85 | 2,9634 | 24 | 666 | 20,2 | 396,9 | 13,22 | 21,4 |
| 466 | 85 | 3,0665 | 24 | 666 | 20,2 | 334,4 | 14,13 | 19,9 |
| 467 | 85 | 2,8715 | 24 | 666 | 20,2 | 22,01 | 17,15 | 19 |
| 468 | 86 | 2,5403 | 24 | 666 | 20,2 | 331,29 | 21,32 | 19,1 |
| 469 | 86 | 2,9084 | 24 | 666 | 20,2 | 368,74 | 18,13 | 19,1 |
| 470 | 86 | 2,8237 | 24 | 666 | 20,2 | 396,9 | 14,76 | 20,1 |
| 471 | 86 | 3,0334 | 24 | 666 | 20,2 | 396,9 | 16,29 | 19,9 |
| 472 | 86 | 3,0993 | 24 | 666 | 20,2 | 395,33 | 12,87 | 19,6 |
| 473 | 86 | 2,8965 | 24 | 666 | 20,2 | 393,37 | 14,36 | 23,2 |
| 474 | 87 | 2,5329 | 24 | 666 | 20,2 | 374,68 | 11,66 | 29,8 |
| 475 | 87 | 2,4298 | 24 | 666 | 20,2 | 352,58 | 18,14 | 13,8 |
| 476 | 87 | 2,206 | 24 | 666 | 20,2 | 302,76 | 24,1 | 13,3 |
| 477 | 87 | 2,3053 | 24 | 666 | 20,2 | 396,21 | 18,68 | 16,7 |
| 478 | 87 | 2,1007 | 24 | 666 | 20,2 | 349,48 | 24,91 | 12 |
| 479 | 87 | 2,1705 | 24 | 666 | 20,2 | 379,7 | 18,03 | 14,6 |
| 480 | 87 | 1,9512 | 24 | 666 | 20,2 | 383,32 | 13,11 | 21,4 |