

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Modelo de regressão de valor extremo para dados agrupados

Jonatas Silva do Espirito Santo

São Carlos

2013

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Modelo de regressão de valor extremo para dados agrupados

Jonatas Silva do Espirito Santo

Orientador: Carlos Alberto Ribeiro Diniz

Co-Orientador: Marcio Luis Lanfredi Viola

São Carlos, 11 de Março de 2013

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Jonatas Silva do Espirito Santo

Modelo de regressão de valor extremo para dados agrupados

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos como parte dos requisitos para obtenção do título de Mestre em Estatística, na área de concentração Modelos de Regressão.

Orientador: Carlos Alberto Ribeiro Diniz

Co-Orientador: Marcio Luis Lanfredi Viola

São Carlos, 11 de Março de 2013.

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

E77mr

Espirito Santo, Jonatas Silva do.

Modelo de regressão de valor extremo para dados agrupados / Jonatas Silva do Espirito Santo. -- São Carlos : UFSCar, 2013.

67 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2013.

1. Estatística. 2. Modelos de regressão. 3. Distribuição valor extremo. 4. Dados agrupados. 5. Equações de estimação generalizadas. I. Título.

CDD: 519.5 (20^a)

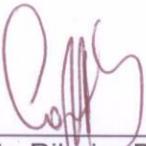


UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Estatística
Via Washington Luís, Km 235 - C.P.676 - CGC 45358058/0001-40
FONE: (016) 3351-8292 – Email: ppgest@ufscar.br
13565-905 - SÃO CARLOS-SP - BRASIL

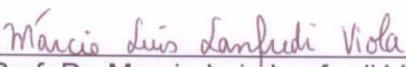
FOLHA DE APROVAÇÃO

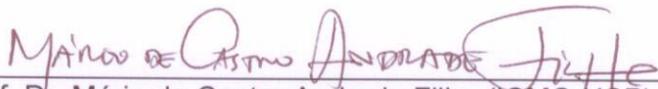
Aluno(a) : Jonatas Silva do Espirito Santo

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 11/03/2013
PELA COMISSÃO JULGADORA:

Presidente 
Prof. Dr. Carlos Alberto Ribeiro Diniz (DEs-UFSCar/Orientador)

1º Examinador 
Prof. Dr. Filidor Edilfonso Vilca Labra (UNICAMP)

2º Examinador 
Prof. Dr. Marcio Luis Lanfredi Viola (DEs-UFSCar/
Co-Orientador)

3º Examinador 
Prof. Dr. Mário de Castro Andrade Filho (ICMC-USP)

Agradecimentos

Primeiramente agradeço a Deus por ter me concedido o dom da vida, saúde, oportunidade de fazer este curso e tudo que tem proporcionado em minha vida.

Agradeço aos meus pais, Claudemiro do Espirito Santo e Analuce dos Santos Silva do Espirito Santo, por tudo que fizeram por mim, desde o ventre de minha mãe até os dias atuais. Por todo amor e carinho, paciência e educação. Sem eles, não chegaria até aqui. Aprendi com eles a ser guerreiro. Aprendi que as conquistas não viriam de forma fácil, que assim como eles, eu precisaria lutar e não ter medo da vida. Obrigado por tudo.

Agradeço aos meus irmãos Enock, Isaac, Felipe e Camile, cunhadas e cunhado, aos meus sobrinhos, e todos os familiares, pelo carinho e paciência. Por apoiar todas as minhas escolhas, e tolerar a minha ausência.

Agradeço à Monalisa, minha eterna Namorada (agora Noiva e em breve Esposa), por todo amor e companheirismo, por apoiar minha ida à São Carlos, pela paciência que teve comigo, por suportar minha falta, pelas palavras de carinho e conforto. Sem você eu não teria conseguido.

Aos colegas do Programa de pós-graduação em Estatística, por compartilhar o conhecimento, pela troca de experiências acadêmicas, pelos momentos em sala de aula e as exaustivas horas de estudos, e também pelos momentos descontraídos e divertidos. Jamais esquecerei deste período em São Carlos. Alguns de vocês foram mais do que colegas, se tornando meus grandes amigos. Valeu por tudo Breno, Broquinha (Marcelo Brocco), Coda (Thiago), Stag (Paulo), Xandy (Alexandre), Jaqueline e Verônica.

Ao meu orientador Carlos Diniz e co-orientador Marcio Viola, pelo apoio, compreensão, ensino e orientação para elaboração deste trabalho.

Aos amigos de república, Bichão (Lucas Monteiro), Miro e Jurandir (mesmo sendo somente um mês), Mateus e Lucas Cogo, este último principalmente, por ter me aturado por mais tempo. Agradeço pela paciência com minhas brincadeiras, pelo apoio e companheirismo. Pelas conversas, companhia nas refeições, pelas séries e filmes. Pelos encontros na cozinha e cafés nas madrugadas. Pelo apoio acadêmico e ajuda com definições matemáticas (Lucas Cogo você manda muito bem!!!).

A todos irmãos e amigos que conquistei em São Carlos, por me acolher e terem feito eu me sentir em casa, em especial aos casais que me ajudaram muito, Ivani e Manuel, Zé e Iraci.

Aos meus professores do programa de pós-graduação em estatística: Carlos Diniz, José Galvão, Luis Milan, Marcio Viola e Vera Tomazella.

Aos funcionários do Departamento de estatística, em especial Isabel, pela paciência e pelo excelente trabalho que realiza na pós-graduação em Estatística.

À UFSCar pela oportunidade e por toda infra-estrutura que desfrutei neste período.

À CAPES pelo auxílio financeiro que permitiu a realização deste trabalho.

Resumo

A distribuição valor extremo tipo I, também conhecida como distribuição Gumbel, é uma das distribuições utilizadas para a modelagem de eventos extremos. Os modelos existentes de regressão valor extremo supõem que as observações sejam independentes, inviabilizando o uso desses modelos quando existe dependência entre as observações. Nesta dissertação, utilizando modelos lineares generalizados (McCullagh e Nelder, 1989) e equações de estimação generalizadas (Liang e Zeger, 1986), desenvolvemos o modelo de regressão valor extremo para o caso em que há grupos independentes formados por respostas dependentes. O comportamento dos estimadores dos parâmetros do modelo proposto é avaliada através de simulações Monte Carlo.

Palavras-chave: *Modelos de Regressão, Distribuição Valor Extremo Tipo I, Dados Agrupados, Equações de Estimação Generalizadas (EEG).*

Abstract

One of the distributions used to model extremal events is the type I extreme-value distribution (Gumbel distribution). The usual extreme-value regression model requires independent observations. In this work, using generalized linear model (McCullagh e Nelder, 1989) and generalized estimating equations (Liang e Zeger, 1986), we developed the extreme-value regression model when there are independent clusters formed by dependent variables. The behavior of parameter estimators of the proposed model is studied through Monte Carlo simulations.

Palavras-chave: *Regression Model, Extreme Value, Cluster Data, Generalized Estimation Equation (GEE).*

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 2 | Modelos de regressão Gumbel para dados independentes | 5 |
| 2.1 | Distribuição Valor Extremo | 6 |
| 2.2 | Modelos de regressão Gumbel com parâmetro de escala conhecido . . | 7 |
| 2.3 | Modelos de regressão Gumbel com parâmetro de escala desconhecido | 9 |
| 2.4 | Modelos de regressão Gumbel generalizado | 12 |
| 3 | Modelo de regressão Gumbel para dados agrupados | 17 |
| 3.1 | Modelagem conjunta da locação e escala | 17 |
| 3.1.1 | Estimação de τ | 25 |
| 3.1.2 | Propriedades assintóticas de $\hat{\tau}$ | 27 |
| 3.1.3 | Matriz de correlações de trabalho | 28 |
| 3.1.4 | Medidas de diagnóstico | 31 |
| 4 | Estudo de simulação | 36 |
| 4.1 | Resultados numéricos | 36 |
| 5 | Discussões e considerações finais | 50 |

| | | |
|----------|--|-----------|
| 6 | Apêndice | 53 |
| 6.1 | Apêndice A | 53 |
| 6.1.1 | Gerando valores de v.a. Gumbel | 53 |
| 6.1.2 | Gerando valores de vetores aleatórios Gumbel | 54 |
| 6.2 | Apêndice B | 55 |
| 6.3 | Apêndice D | 60 |
| | Referências Bibliográficas | 64 |

Capítulo 1

Introdução

A teoria de valores extremos é utilizada quando se tem interesse em estudar eventos raros, ou seja, aqueles que têm pequena probabilidade de ocorrência, mas que quando ocorrem podem trazer consequências indesejáveis (desejáveis) e grandes prejuízos (acarretar benefícios e lucros).

Geralmente o interesse dos pesquisadores é construir modelos que serão utilizados para auxiliar nas tomadas de decisões ou medidas de prevenção contra eventos raros.

Em termos estatísticos, o interesse do pesquisador é a modelagem probabilística e a inferência estatística nas caudas (superior e inferior) da distribuição de uma variável de interesse (Viola, 2007).

Existem algumas distribuições que modelam eventos raros, sendo uma delas a distribuição valor extremo tipo I, também conhecida como distribuição Gumbel ou distribuição log-Weibull. Algumas situações em que podemos utilizar a distribuição valor extremo tipo I são: precipitação pluviométrica (chuva), temperaturas extremas (mínima e máxima), inundações, velocidade de vento, seguros, perdas financeiras (risco operacional), entre outras.

Diversos artigos com aplicações da distribuição Gumbel, principalmente na área de Hidrologia e em cálculos do VaR (Valor em Risco), estão disponíveis na literatura. Encontram-se, também, artigos com desenvolvimento teórico e aplicações de modelos de regressão com resposta Gumbel, como os citados nos próximos parágrafos.

O modelo linear de valor extremo, descrito em Ramos e Cordeiro (2006), assume que a variável resposta segue uma distribuição valor extremo tipo I com parâmetro de escala conhecido e o parâmetro de localização relacionado com covariáveis através de uma função de ligação identidade. Nesse artigo os autores apresentam correção de viés de segunda ordem para os estimadores de máxima verossimilhança dos parâmetros de regressão. Corrigem também as estatísticas de teste da razão de verossimilhanças e escore em modelos lineares de valor extremo. Os resultados da simulação, apresentada no artigo, apontam que, para problemas práticos com pequenas amostras, o uso dos estimadores corrigidos conduz a inferências mais precisas do que quando se utiliza estimadores usuais (sem correção).

Clarke (2002) apresenta um modelo de regressão Gumbel em que o parâmetro de localização é escrito como uma combinação linear de covariáveis e o parâmetro de escala é desconhecido. A estimação dos parâmetros é feita através de um algoritmo iterativo que combina o método dos momentos, o método da máxima verossimilhança e o de mínimos quadrados ponderados. Com uma estimativa prévia do parâmetro de escala, é feita uma transformação na variável resposta permitindo que a função densidade possa ser reescrita na forma da família exponencial, e assim o parâmetro de localização é modelado como em um modelo linear generalizado. A motivação para o artigo é baseada no fato de que, em regiões com mudanças climáticas ou mudanças no uso do solo, a suposição comum de que séries temporais hidrológicas são estacionárias pode ser posta em questão. Registros de inundações anuais (tipicamente, sequências anuais de vazões máximas diárias) podem mostrar mudanças ao longo dos anos como

consequência de escoamento mais rápido, associados com o aumento da urbanização, mudanças das práticas agrícolas ou desmatamento na bacia hidrográfica. O artigo analisa uma base de dados de vazões máximas anuais em dois locais de medição do rio Jacuí, situado no sul do Brasil, motivado pelo fato de que, nas últimas décadas, desmatamentos na bacia do rio Jacuí têm sido intensos com florestas nativas sendo substituídas por agricultura.

O modelo geral de regressão de valor extremo foi proposto por Barreto-Souza e Vasconcellos (2011), assumindo que os parâmetros de locação e escala estão relacionados com covariáveis através de duas estruturas não lineares de regressão, supondo independência entre as observações. A estimação dos parâmetros é feita através do método da máxima verossimilhança. Nesse trabalho os autores apresentam ainda fórmulas para o viés de segunda ordem das estimativas de máxima verossimilhança, visto que os estimadores corrigidos são importantes para se fazer melhores inferências quando o tamanho da amostra é pequeno. Uma aplicação do modelo é feita com dados de crescimento de trigo no inverno.

Observe que os três artigos resumidos nos parágrafos anteriores abordam modelos de regressão Gumbel, mas todos assumem que as variáveis respostas são independentes. É possível, em algumas situações, a existência de algum tipo de dependência entre as observações. Por exemplo, suponha que seja de interesse modelar a precipitação pluviométrica (quantidade de chuva) extrema de uma determinada cidade como consequência de mudanças climáticas e que dados diários sobre a quantidade de chuva estejam disponíveis. O fato de ter chovido (ou não) em um determinado dia pode influenciar a ocorrência (ou não) de chuva nos dias posteriores, ficando evidente certa dependência intrínseca entre as observações.

Outra abordagem para dados correlacionados são os dados agrupados, em que as observações em cada grupo são correlacionadas, existindo porém independência

entre os grupos. Pode-se pensar novamente em estudos de precipitação pluviométrica extrema, observando-se a quantidade de chuva em várias cidades, considerando cada cidade como um grupo e independência entre as cidades, porém com correlação entre as observações em cada grupo.

Em algumas situações, como a citada no parágrafo anterior, a suposição de independência entre as observações nem sempre é satisfeita, tornando incorreto o uso da função escore como função de estimação, e conseqüentemente inviabilizando, nesses casos, o uso dos modelos sugeridos por Clarke (2002), Ramos e Cordeiro (2006) e Barreto-Souza e Vasconcellos (2011). Uma alternativa é utilizar equações de estimação generalizadas (EEG), metodologia proposta por Liang e Zeger (1986).

Desde a publicação de Liang e Zeger (1986), estudos em diversas áreas envolvendo experimentos com medidas repetidas utilizando EEG têm sido realizados. Entre estes, Artes e Jørgensen (2000) para modelos de dispersão, Rodrigues (2009) para modelos com respostas binárias, Artes (2009) para modelos com resposta inflacionada de zeros e Venezuela (2008) para modelos beta e simplex.

Nesta dissertação, baseados em modelos lineares generalizados (Nelder e Wedderburn, 1972) e equações de estimação generalizadas (Liang e Zeger, 1986) propomos o modelo de regressão de valor extremo para dados agrupados, considerando grupos independentes de respostas correlacionadas, seguindo distribuição valor extremo tipo I e assumindo que os parâmetros de locação e escala possuem relação com covariáveis através de estruturas não lineares.

O texto está dividido em capítulos da seguinte forma. No Capítulo 2 apresentamos alguns modelos de regressão Gumbel para dados independentes. No Capítulo 3 apresentamos o desenvolvimento do modelo de regressão Gumbel para dados agrupados. No Capítulo 4 elaboramos um estudo de simulação Monte Carlo. No Capítulo 5 encontram-se os resultados finais e discussões.

Capítulo 2

Modelos de regressão Gumbel para dados independentes

Iniciamos o capítulo com um breve resumo da distribuição Gumbel, mostrando as funções que a caracterizam (densidade, distribuição e geradora de momentos) e sua média e variância. Em seguida descrevemos três modelos de regressão Gumbel para dados independentes. No primeiro, Ramos e Cordeiro (2006) assumem que a variável resposta segue uma distribuição Gumbel com parâmetro de escala conhecido e o parâmetro de locação relacionado com covariáveis através de uma estrutura linear. No segundo, Clarke (2002) também escreve o parâmetro de locação como uma combinação linear de covariáveis, mas diferente do primeiro modelo, o parâmetro de escala é desconhecido. O terceiro é o modelo de regressão geral de valor extremo, sugerido por Barreto-Souza e Vasconcellos (2011), no qual assumem que ambos, parâmetro de locação e de escala, estão relacionados com covariáveis através de estruturas não lineares. Apresentamos o desenvolvimento teórico dos modelos e também resultados numéricos e uma aplicação.

2.1 Distribuição Valor Extremo

Desenvolvida principalmente para suprir necessidades de astrônomos em estudos de observações longínquas e com diversas aplicações descritas em Gumbel (1958), Kotz e Nadarajah (2000), entre outros, a distribuição valor extremo tipo I é indexada por um parâmetro de locação $\theta \in \mathbb{R}$ e um parâmetro de escala $\phi > 0$. Se uma variável aleatória Y segue uma distribuição $Gumbel(\theta, \phi)$, então sua função densidade de probabilidade (f.d.p.) é dada por

$$f(y|\theta, \phi) = \frac{1}{\phi} \exp\left(-\frac{y-\theta}{\phi}\right) \exp\left[-\exp\left(-\frac{y-\theta}{\phi}\right)\right], \quad y \in \mathbb{R}, \quad (2.1)$$

com função de distribuição acumulada dada por

$$F(y|\theta, \phi) = \exp\left[-\exp\left(-\frac{y-\theta}{\phi}\right)\right] \quad (2.2)$$

e função geradora de momentos por

$$M_Y(t) = E(e^{tY}) = e^{t\theta} \Gamma(1 - \phi t), \quad |t| < \phi^{-1},$$

sendo $\Gamma(\cdot)$ a função gama definida como

$$\Gamma(a) = \int_0^\infty e^{-y} y^{a-1} dy, \quad a > 0. \quad (2.3)$$

A esperança e a variância de Y são dadas por

$$E(Y) = \theta + \gamma\phi \quad \text{e} \quad \text{Var}(Y) = \frac{\pi^2}{6}\phi^2, \quad (2.4)$$

em que γ é o número de Euler-Mascheroni, que vale aproximadamente 0,5772156649.

2.2 Modelos de regressão Gumbel com parâmetro de escala conhecido

Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes, com $Y_i \sim \text{Gumbel}(\theta_i, \phi)$, sendo $\theta_i \in \mathbb{R}$ o parâmetro de locação e $\phi > 0$ o parâmetro de escala, $i = 1, \dots, n$. A função densidade de probabilidade de Y_i é dada por

$$f(y_i|\theta_i, \phi) = \frac{1}{\phi} \exp\left(-\frac{y_i - \theta_i}{\phi}\right) \exp\left[-\exp\left(-\frac{y_i - \theta_i}{\phi}\right)\right], \quad y_i \in \mathbb{R}. \quad (2.5)$$

O modelo linear de valor extremo (Ramos e Cordeiro, 2006) assume que o parâmetro de escala é conhecido e igual para todas as variáveis aleatórias Y_i , $i = 1, \dots, n$. Considerando o vetor de parâmetros de locação $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)'$, a componente sistemática do modelo é dada por

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}, \quad (2.6)$$

sendo $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ a matriz do modelo, de dimensão $n \times p$ ($n > p$) e posto completo. Cada vetor $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})$ representa um conjunto de p -covariáveis associadas à i -ésima observação e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ um vetor de parâmetros de regressão desconhecidos que desejamos estimar.

A estimação de $\boldsymbol{\beta}$ pode ser feita pelo método da máxima verossimilhança. Assim, a função log-verossimilhança do vetor de parâmetros de regressão no modelo linear de valor extremo com relação à amostra $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ é dada por

$$\ell(\boldsymbol{\beta}|\mathbf{y}, \phi) = -n \log(\phi) - \frac{1}{\phi} \sum_{i=1}^n (y_i - \theta_i) - \sum_{i=1}^n \exp\left(-\frac{y_i - \theta_i}{\phi}\right). \quad (2.7)$$

A função escore de $\boldsymbol{\beta}$, definida como $\mathbf{U}_\beta = \partial\ell/\partial\boldsymbol{\beta}$, na forma matricial é

$$\mathbf{U}_\beta = -\phi^{-1}\mathbf{X}'\mathbf{g}, \quad (2.8)$$

sendo $\mathbf{g} = (g_1, \dots, g_n)'$, com componentes definidas por $g_i = \exp[-(y_i - \theta_i)/\phi] - 1$, para $i = 1, \dots, n$. A matriz de informação de Fisher, associada a $\boldsymbol{\beta}$, em relação à amostra \mathbf{y} é dada por

$$\mathbf{K}_\beta = \phi^{-2}\mathbf{X}'\mathbf{X}. \quad (2.9)$$

O estimador de máxima verossimilhança (EMV), $\hat{\boldsymbol{\beta}}$, pode ser obtido utilizando um algoritmo iterativo de mínimos quadrados ponderados, que pode ser implementado em diversos pacotes estatísticos, e é dado por

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \phi(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{g}^{(m)}, \quad (2.10)$$

em que $\boldsymbol{\beta}^{(m+1)}$ é a estimativa de $\boldsymbol{\beta}$ no passo $m + 1$ e no lado direito da equação, o índice m indica que as quantidades são calculadas com valores obtidos no passo m .

Interessados em corrigir o viés de segunda ordem do estimador de máxima verossimilhança de $\boldsymbol{\beta}$, Ramos e Cordeiro (2006) desenvolveram uma fórmula geral em notação matricial para o viés e corrigiram o estimador. O EMV corrigido é dado por

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \text{Viés}(\hat{\boldsymbol{\beta}}), \quad (2.11)$$

sendo

$$\text{Viés}(\hat{\boldsymbol{\beta}}) = -\frac{\phi}{2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}, \quad (2.12)$$

em que $\text{Viés}(\hat{\boldsymbol{\beta}})$ é o viés de segunda ordem de $\hat{\boldsymbol{\beta}}$, e h_{rs} , $r, s = 1, \dots, n$, são os elementos da diagonal principal de $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Através de estudos de simulação, Ramos e Cordeiro (2006) mostram que os EMV corrigidos têm melhores propriedades amostrais do que os não corrigidos, pois os vieses dos componentes de $\tilde{\beta}$ foram menores do que os de $\hat{\beta}$, ou seja, as estimativas obtidas pelos EMV corrigidos estão mais próximas dos verdadeiros valores dos parâmetros do que as obtidas por EMV usuais.

Os autores apresentaram também correções para as estatísticas de teste da razão de verossimilhanças e do escore em modelos lineares de valor extremo. Com estudos de simulação, mostrou-se que a distribuição das estatísticas modificadas se aproximam mais da distribuição χ^2 , resultando em inferências mais refinadas, principalmente em problemas práticos com amostras reduzidas nas quais as estatísticas de testes usuais fornecem resultados imprecisos em termos de aproximação assintótica.

2.3 Modelos de regressão Gumbel com parâmetro de escala desconhecido

Suponha Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes, cada uma seguindo distribuição Gumbel com parâmetro de locação $\theta_i \in \mathbb{R}$ e parâmetro de escala $\phi > 0$, $i = 1, \dots, n$. A f.d.p. de cada Y_i é dada pela equação (2.5) e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)'$ tem a mesma estrutura linear (2.6) que o modelo apresentado na seção anterior.

O modelo apresentado em Clarke (2002), diferente do descrito em Ramos e Cordeiro (2006), assume que o parâmetro de escala é desconhecido. Em Clarke (2002) dois métodos são utilizados para estimar os parâmetros de regressão e o parâmetro de escala, o método da máxima verossimilhança e outro combinando método dos momentos, da máxima verossimilhança e o de mínimos quadrados ponderados.

No método da máxima verossimilhança, as estimativas são as solução do sistema linear composto pelas equações $\partial\ell/\partial\phi = 0$ e $\partial\ell/\partial\boldsymbol{\beta} = \mathbf{0}$, obtidas a partir de um método numérico como o Newton-Raphson, sendo ℓ a função log-verossimilhança (2.7) e $\boldsymbol{\beta}$ o vetor de parâmetros de regressão.

No segundo método, o autor se baseia em modelos lineares generalizados (MLG), que foram desenvolvidos por Nelder e Wedderburn (1972). Em um MLG, os dados podem ser uma amostra de variáveis aleatórias cujas distribuições pertencem à família exponencial. Dizemos que uma distribuição pertence à família exponencial se sua função densidade de probabilidade (fdp) pode ser escrita da forma

$$f(y|\delta, \sigma) = \exp \{(y\delta - b(\delta))/a(\sigma) + c(y, \sigma)\} \quad (2.13)$$

em que $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções específicas de cada distribuição. Pode-se mostrar que algumas distribuições como normal, Poisson, binomial e gama, têm fdp que podem ser escritas na forma (2.13) e portanto, pertencem à família exponencial.

Considerando a teoria de MLG (McCullagh and Nelder, 1989), o modelo Gumbel é descrito substituindo o parâmetro de locação (θ_i) na equação (2.5) por $\mathbf{x}'_i\boldsymbol{\beta}$, sendo \mathbf{x}_i o vetor de covariáveis. Suponha, inicialmente, que o parâmetro ϕ seja conhecido ou estimado previamente. Seja $z_i = -\exp(-y_i/\phi)$ uma transformação nos dados originais y_1, \dots, y_n . Então, (2.5) pode ser escrita na forma da família exponencial

$$f(z|\theta^*, \phi) = \exp \{z\theta^* + \log(\theta^*) + \log(-z/\phi)\}; \quad -\infty < z < 0; \quad \theta^* > 0 \quad (2.14)$$

sendo $\theta^* = \exp(\theta/\phi)$. Comparando a equação (2.14) com (2.13), temos que $a(\phi) = 1$, $b(\theta^*) = -\log(\theta^*)$ e $c(y, \phi) = \log(-z/\phi)$ e assim $\mu_z = E(Z) = b'(\theta^*) = -1/\theta^*$ e $Var(Z) = b''(\theta^*)a(\phi) = 1/\theta^{*2}$.

O desenvolvimento do MLG com resposta Gumbel assume que o parâmetro ϕ é conhecido, de modo que a nova variável Z possa ser calculada, porém na prática ϕ não é conhecido, mas pode ser estimado. Uma alternativa é fazer a estimação pela solução da equação $\partial\ell/\partial\phi = 0$ com $\hat{\boldsymbol{\beta}}$ fixado. O seguinte algoritmo fornece uma estimação iterativa de ϕ e do vetor de parâmetros $\boldsymbol{\beta}$:

1. Estime ϕ através dos dados utilizando o método dos momentos $\tilde{\phi} = (s\sqrt{6})/\pi$, sendo s o desvio padrão amostral de \mathbf{y} .
2. Use a estimativa de ϕ obtida no passo 1 para obter a estimativa de $\boldsymbol{\beta}$ usando o algoritmo iterativo de mínimos quadrados ponderados do modelo expressado em termos de um MLG.
3. Use a estimativa de $\boldsymbol{\beta}$ obtida no passo 2 para resolver $\partial\ell/\partial\phi = 0$:

$$-n/\phi + \sum_{i=1}^n [y_i - \mathbf{x}'_i\boldsymbol{\beta}] / \phi^2 - \sum_{i=1}^n [(y_i - \mathbf{x}'_i\boldsymbol{\beta}) / \phi^2] \exp [(y_i - \mathbf{x}'_i\boldsymbol{\beta}) / \phi] = 0. \quad (2.15)$$

4. Com a estimativa de ϕ obtida no passo 3, retorne ao passo 2, e itere os passos 2 e 3 até convergir.

Clarke (2002) apresentou uma aplicação do modelo a dados de vazões máximas anuais em dois locais de medição (Espumoso e Passo Bela Vista), no Rio Jacuí, situado no sul do Brasil. Uma das motivações da análise é o fato de que, nas últimas décadas, desmatamentos na bacia do rio Jacuí têm sido intensos, com florestas nativas sendo substituídas por áreas de atividades agrícolas. Ajustou-se então um modelo de regressão Gumbel, em que as vazões máximas anuais seguem distribuição valor extremo com a média variando no tempo, $(m + \gamma/\hat{\phi}) + \beta t$, sendo t a covariável tempo, ϕ o parâmetro de escala desconhecido e γ o número de Euler.

A estimação foi feita com o algoritmo iterativo proposto por Clarke (2002), ajustando no passo 2 um MLG com função de ligação $\log(\cdot)$ para o parâmetro da distribuição de $z = -\exp(-y/\phi)$ e a equação $\partial\ell/\partial\phi = 0$ do passo 3 dada por

$$-n/\phi + \sum_{i=1}^n [y_i - m_i - \beta t_i]/\phi^2 - \sum_{i=1}^n [(y_i - m_i - \beta t_i)/\phi^2] \exp[(y_i - m_i - \beta t_i)/\phi] = 0$$

Para Espumoso, a tendência ajustada é $(\hat{m} + \gamma/\hat{\phi}) + \hat{\beta}t = 484,72 + 5,96t$ resultando uma estatística da razão de verossimilhanças de $\chi^2_{(1)} = 15.842$ e para Passo Bela Vista a tendência ajustada é dada por $804,82 + 5,26t$ com $\chi^2_{(1)} = 2.382$, para testar a hipótese $\beta = 0$. Valores tabelados da distribuição $\chi^2_{(1)}$ para os níveis de significância 5%, 1% e 0,1% são respectivamente 3,841, 6,635 e 10,827, mostrando alta significância para o parâmetro de tendência do modelo de Espumoso e não significância para o parâmetro de tendência de Passo Bela Vista.

2.4 Modelos de regressão Gumbel generalizado

Considere Y_1, Y_2, \dots, Y_n v.a. independentes, em que $Y_i \sim Gumbel(\theta_i, \phi_i)$, para $i = 1, \dots, n$, e função densidade de probabilidade dada por

$$f(y_i|\theta_i, \phi_i) = \frac{1}{\phi_i} \exp\left\{-\frac{y_i - \theta_i}{\phi_i} - \exp\left(-\frac{y_i - \theta_i}{\phi_i}\right)\right\}, \quad y, \theta_i \in \mathbb{R}; \quad \phi_i > 0 \quad (2.16)$$

Note que, diferentemente das duas últimas situações apresentadas nas Seções 2.2 e 2.3, o parâmetro de escala, além de ser desconhecido, difere de uma observação para outra. Portanto, os modelos apresentados nas seções anteriores não seriam adequados para modelar dados com essa estrutura, sendo então necessária a formulação de um modelo que leve em consideração esta heteroscedasticidade.

O modelo de regressão Gumbel generalizado (Barreto-Souza e Vasconcellos, 2011), incorpora o termo “generalizado” em seu nome pois, além de levar em consideração a heterocedasticidade, permite que a relação entre os parâmetros da distribuição e as covariáveis não fique restrita somente às formas lineares.

Este modelo assume que os parâmetros de locação e os de escala estão relacionados com covariáveis através de funções não lineares (podendo ser também lineares), sendo estruturado por duas componentes sistemáticas, uma para os parâmetros de locação e outra para os parâmetros de escala:

$$g_1(\theta_i) = \eta_{1i} = f_1(\mathbf{x}_i, \boldsymbol{\beta}) \quad \text{e} \quad g_2(\phi_i) = \eta_{2i} = f_2(\mathbf{z}_i, \boldsymbol{\xi}), \quad (2.17)$$

em que $g_1(\cdot)$ e $g_2(\cdot)$ são funções de ligação conhecidas, estritamente monótonas e que admitem derivada de segunda ordem, com $Im[g_1(\cdot)] = \mathbb{R}$ e $Im[g_2(\cdot)] = \mathbb{R}^+$. As funções $f_1(\cdot, \boldsymbol{\beta})$ e $f_2(\cdot, \boldsymbol{\xi})$ são contínuas e admitem derivadas de segunda ordem. Os vetores de parâmetros desconhecidos, os quais devemos estimar, são $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, $\boldsymbol{\beta} \in \mathbb{R}^p$, e $\boldsymbol{\xi} = (\xi_1, \dots, \xi_q)'$, $\boldsymbol{\xi} \in \mathbb{R}^q$. Os vetores \mathbf{x}_i e \mathbf{z}_i , de dimensões p e q respectivamente, são compostos pelos valores das covariáveis para a i -ésima observação. Vale ressaltar que \mathbf{x}_i e \mathbf{z}_i não são, necessariamente, diferentes, ou seja, pode-se utilizar as mesmas covariáveis para as duas componentes sistemáticas.

O método da máxima verossimilhança é utilizado para estimar os parâmetros. Sejam $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)'$ e $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_n)'$ os vetores de parâmetros de locação e de escala, com θ_i e ϕ_i definidos em (2.17), para $i = 1, \dots, n$. A função log-verossimilhança de $\boldsymbol{\theta}$ e $\boldsymbol{\phi}$ para a amostra $\mathbf{y} = (y_1, \dots, y_n)'$ é dada por

$$\ell(\boldsymbol{\theta}, \boldsymbol{\phi} | \mathbf{y}) = \sum_{i=1}^n \left\{ -\log(\phi_i) - \left(\frac{y_i - \theta_i}{\phi_i} \right) - \exp\left(-\frac{y_i - \theta_i}{\phi_i} \right) \right\}. \quad (2.18)$$

Considere $Y_i^* = \exp(-Y_i/\phi_i)$, $y_i^* = \exp(-y_i/\phi_i)$, $\mu^* = E(Y_i^*) = \exp(-\theta_i/\phi_i)$ e $\nu_i = -1 + (y_i - \theta_i)[1 - \exp(-(y_i - \theta_i)/\phi_i)]/\phi_i$, para $i = 1, \dots, n$. A função escore é definida por $\mathbf{U} = (\partial\ell/\partial\beta_1, \dots, \partial\ell/\partial\beta_p, \partial\ell/\partial\xi_1, \dots, \partial\ell/\partial\xi_q)'$ sendo

$$U_{\beta_j} = \frac{\partial\ell}{\partial\beta_j} = \sum_{i=1}^n (y_i^* - \mu_i^*) \left(-\frac{1}{\mu_i^* \phi_i} \right) \frac{\partial\theta_i}{\partial\eta_{1i}} \frac{\partial\eta_{1i}}{\partial\beta_j}, \quad j = 1, \dots, p, \quad (2.19)$$

$$e$$

$$U_{\xi_l} = \frac{\partial\ell}{\partial\xi_l} = \sum_{i=1}^n \nu_i \frac{1}{\phi_i} \frac{\partial\phi_i}{\partial\eta_{2i}} \frac{\partial\eta_{2i}}{\partial\xi_l}, \quad l = 1, \dots, q. \quad (2.20)$$

Seja $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{\xi})'$. Então, a função escore $\mathbf{U}_\tau = (\mathbf{U}_\beta, \mathbf{U}_\xi)'$ na forma matricial é

$$\mathbf{U}_\tau = \left(\frac{\partial\ell}{\partial\boldsymbol{\beta}'}, \frac{\partial\ell}{\partial\boldsymbol{\xi}'} \right)' = \left(\tilde{\mathbf{X}}' \boldsymbol{\Omega}^{-1} \mathbf{M}_1 (\mathbf{y}^* - \boldsymbol{\mu}^*), \tilde{\mathbf{Z}}' \boldsymbol{\Phi}^{-1} \mathbf{M}_2 \boldsymbol{\nu} \right)'_{(p+q) \times 1}, \quad (2.21)$$

sendo $\mathbf{y}^* = (y_1^*, \dots, y_n^*)'$, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)'$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)'$,

$$\tilde{\mathbf{X}} = [\partial\eta_{1i}/\partial\beta_j]_{i,j}, \quad \boldsymbol{\Omega} = \text{diag} [-\mu_i^* \phi_i], \quad \mathbf{M}_1 = \text{diag} [\partial\theta_i/\partial\eta_{1i}],$$

$$\tilde{\mathbf{Z}} = [\partial\eta_{2i}/\partial\xi_l]_{i,l}, \quad \boldsymbol{\Phi} = \text{diag} [\phi_i] \quad e \quad \mathbf{M}_2 = \text{diag} [\partial\phi_i/\partial\eta_{2i}].$$

A matriz de informação de Fisher, associada a $\boldsymbol{\tau}$, é dada por

$$\mathbf{K}_\tau = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta\xi} \\ \mathbf{K}_{\xi\beta} & \mathbf{K}_{\xi\xi} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{X}}' \mathbf{W}_{\beta\beta} \tilde{\mathbf{X}} & \tilde{\mathbf{X}}' \mathbf{W}_{\beta\xi} \tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}' \mathbf{W}_{\xi\beta} \tilde{\mathbf{X}} & \tilde{\mathbf{Z}}' \mathbf{W}_{\xi\xi} \tilde{\mathbf{Z}} \end{bmatrix},$$

sendo $\mathbf{W}_{\beta\beta} = \text{diag} \left[\left(\frac{1}{\phi_i} \frac{\partial\theta_i}{\partial\eta_{1i}} \right)^2 \right]$, $\mathbf{W}_{\beta\xi} = \mathbf{W}_{\xi\beta} = \text{diag} \left[\frac{(\gamma - 1)}{\phi_i^2} \frac{\partial\theta_i}{\partial\eta_{1i}} \frac{\partial\phi_i}{\partial\eta_{2i}} \right]$ e

$$\mathbf{W}_{\xi\xi} = \text{diag} \left[\frac{\{\pi^2/6 + (\gamma - 1)^2\}}{\phi_i^2} \left(\frac{\partial\phi_i}{\partial\eta_{2i}} \right)^2 \right].$$

Podemos reescrever a matriz de informação de Fisher como $\mathbf{K} = \mathbb{X}\tilde{\mathbf{W}}\mathbb{X}'$ sendo

$$\mathbb{X} = \begin{bmatrix} \tilde{\mathbf{X}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{Z}} \end{bmatrix}_{2n \times (p+q)} \quad e \quad \tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_{\beta\beta} & \mathbf{W}_{\beta\xi} \\ \mathbf{W}_{\xi\beta} & \mathbf{W}_{\xi\xi} \end{bmatrix}_{2n \times 2n}.$$

O estimador de máxima verossimilhança, $\hat{\boldsymbol{\tau}}$, do vetor de parâmetros $\boldsymbol{\tau}$ é obtido a partir da solução do sistema de equações não lineares $\mathbf{U}_{\boldsymbol{\tau}} = \mathbf{0}$, que, para resolvê-lo, é necessário utilizar métodos numéricos como Newton-Raphson ou Escore de Fisher. Sob certas condições de regularidade e assumindo que $\mathbf{J} = \lim_{n \rightarrow \infty} \{\mathbf{K}_{\boldsymbol{\tau}}/n\}$ existe e é não singular (vide Barreto-Souza, 2009), temos que a distribuição assintótica de $\hat{\boldsymbol{\tau}}$ é dada por

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \xrightarrow{D} N_{p+q}(\mathbf{0}, \mathbf{J}_{\boldsymbol{\tau}}^{-1})$$

Barreto-Souza e Vasconcellos (2011) apresentam ainda correção do viés de segunda ordem dos EMV. Os autores demonstram, através de estudos de simulação, que os estimadores sugeridos possuem menores vieses do que os EMV usuais. Mostram também que a inserção da componente sistemática associada ao parâmetro de dispersão resulta em melhores ajustes do modelo, ficando evidente que a utilização de covariáveis para estimar o parâmetro de dispersão é bastante útil na presença de heterocedasticidade.

Barreto-Souza e Vasconcellos (2011) apresentam uma aplicação do modelo a dados reais sobre o crescimento de trigo no inverno, interessados na diferença dos pesos secos dos perfilhos de trigos. A covariável, x , é medida em uma escala de graus-dias, e corresponde aos dias em que o trigo é submetido a temperaturas que estão acima da menor temperatura em que o trigo possa se desenvolver. A temperatura é medida em graus Celsius e o tempo em dias, e o tempo inicial sendo determinado pelo estado fisiológico do trigo. Foram escolhidas aleatoriamente 18 pequenas áreas de

aproximadamente $0,15m^2$ e a cada semana o peso seco dos perfilhos de cada planta coletada em cada área era avaliado, sendo que os pesos foram medidos em miligramas. Assumindo, então, que os pesos secos dos perfilhos (y) seguem uma distribuição de valor extremo tipo I, ajustou-se um modelo de regressão Gumbel generalizado com as seguintes componentes sistemáticas:

$$\theta_i = \beta_0 + e^{\beta_1 + \beta_2 x_i} \quad \text{e} \quad \log(\phi_i) = \xi x_i.$$

A Figura 2.1 mostra as curvas das estimativas de máxima verossimilhança usuais e corrigidas para a média e a dispersão juntamente com os dados reais. Observa-se que as curvas se ajustam bem aos dados, indicando que o modelo poderia ser usado para estimação do crescimento de trigo no inverno. Percebe-se também que as curvas com as estimativas usuais e corrigidas da dispersão foram diferentes.

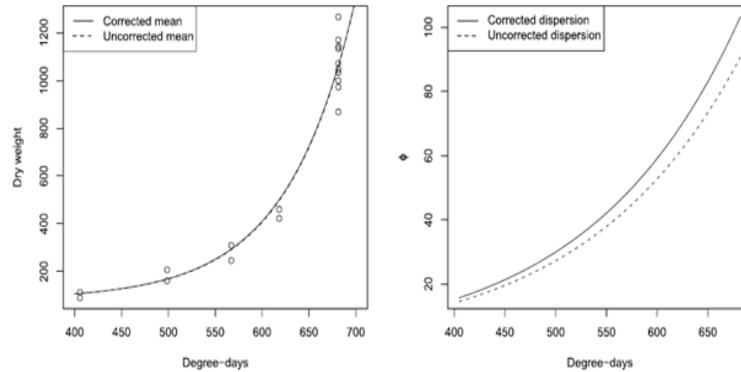


Figura 2.1: Curvas das estimativas corrigidas e não corrigidas da média e curvas do parâmetro de dispersão para os dados de trigo (Barreto-Souza e Vasconcellos, 2011).

No próximo capítulo apresentaremos um modelo alternativo aos descritos neste capítulo, para situações em que o interesse seja modelar eventos raros utilizando a distribuição valor extremo para grupos de dados correlacionados.

Capítulo 3

Modelo de regressão Gumbel para dados agrupados

Os três artigos discutidos no Capítulo 2 apresentam modelos de regressão Gumbel que são construídos assumindo independência entre as observações, condição que nem sempre é satisfeita. Em algumas situações, os dados disponíveis possuem algum tipo de dependência e, portanto, é necessário a formulação de modelos que levem em consideração a estrutura de dependência entre as observações. Neste capítulo apresentamos um modelo de regressão Gumbel para dados agrupados.

3.1 Modelagem conjunta da locação e escala

Sejam $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ vetores aleatórios independentes, sendo $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})'$, com $Y_{it} \sim Gumbel(\theta_{it}, \phi_{it})$. A função densidade de probabilidade de Y_{it} é dada por

$$f(y_{it}|\theta_{it}, \phi_{it}) = \frac{1}{\phi_{it}} \exp\left(-\frac{y_{it} - \theta_{it}}{\phi_{it}}\right) \exp\left[-\exp\left(-\frac{y_{it} - \theta_{it}}{\phi_{it}}\right)\right], \quad y_{it} \in \mathbb{R}, \quad (3.1)$$

para $t = 1, \dots, m_i$ e $i = 1, \dots, n$, em que $\theta_{it} \in \mathbb{R}$ é o parâmetro de locação e $\phi_{it} > 0$ é o parâmetro de escala. A esperança e variância de Y_{it} são, respectivamente, $E(Y_{it}) = \mu_{it} = \theta_{it} + \gamma\phi_{it}$ e $Var(Y_{it}) = \nu_{it} = \pi^2\phi_{it}^2/6$, sendo γ o número de Euler.

Considere ainda que existe dependência intragrupal mesmo havendo independência entre os grupos, ou seja, apesar de os vetores aleatórios $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ serem independentes entre si, as componentes destes vetores aleatórios são correlacionadas. Portanto, para $i = 1, \dots, n$, a matriz de variâncias dos elementos do vetor aleatório $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})'$ é dada por $\mathbf{V}_i = \text{Var}(\mathbf{Y}_i) = \text{diag}[\nu_{i1}, \dots, \nu_{im_i}]$, e a matriz de covariâncias por $\text{Cov}(\mathbf{Y}_i) = \boldsymbol{\Sigma}_i = \mathbf{V}_i^{1/2} \mathbf{R}(\mathbf{Y}_i) \mathbf{V}_i^{1/2}$, sendo $\mathbf{R}(\mathbf{Y}_i)$ a matriz de correlações do vetor \mathbf{Y}_i . Algumas possíveis estruturas de \mathbf{R} são apresentadas abaixo.

1) **Diagonal.** Caracterizada pela correlação nula entre qualquer par de observações distintas. Não existe nenhum parâmetro a ser estimado;

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (3.2)$$

2) **Permutável.** Assume que a correlação entre quaisquer duas observações diferentes de um mesmo grupo é igual. ρ é um escalar;

$$\mathbf{R}(\rho) = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}, \quad |\rho| < 1. \quad (3.3)$$

3) **Toeplitz.** Estrutura associada a dados de séries temporais igualmente espaçados.

$\boldsymbol{\rho}$ é um vetor de ordem m_i ;

$$\mathbf{R}(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{m_i-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{m_i-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{m_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{m_i-1} & \rho_{m_i-2} & \rho_{m_i-3} & \cdots & 1 \end{bmatrix}, \quad |\rho_k| < 1. \quad (3.4)$$

4) **Auto-regressiva de ordem 1.** A correlação diminui exponencialmente com o aumento da diferença de tempo entre as observações. ρ é um escalar;

$$\mathbf{R}(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{m_i-1} \\ \rho & 1 & \rho & \cdots & \rho^{m_i-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{m_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m_i-1} & \rho^{m_i-2} & \rho^{m_i-3} & \cdots & 1 \end{bmatrix}, \quad |\rho| < 1. \quad (3.5)$$

5) **Não estruturada.** Um parâmetro de correlação é associado a cada par de observações do grupo. $\boldsymbol{\rho}$ é um vetor de ordem $m_i(m_i - 1)/2$;

$$\mathbf{R}(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1m_i} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2m_i} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3m_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{m_i1} & \rho_{m_i2} & \rho_{m_i3} & \cdots & 1 \end{bmatrix}, \quad |\rho_{tt'}| < 1. \quad (3.6)$$

Considere os vetores de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n)'$ e $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)'$, em que $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{im_i})'$ e $\boldsymbol{\phi}_i = (\phi_{i1}, \dots, \phi_{im_i})'$, $i = 1, \dots, n$, e admita que cada componente θ_{it} esteja relacionado com um conjunto de p -covariáveis representadas pelo vetor $\mathbf{x}_{it} = (x_{it_1}, \dots, x_{it_p})'$ e cada ϕ_{it} relacionado com um conjunto de q -covariáveis representadas pelo vetor $\mathbf{z}_{it} = (z_{it_1}, \dots, z_{it_q})'$ através de funções não lineares.

O modelo de regressão de valor extremo para dados agrupados é especificado por duas componentes sistemáticas, uma para os parâmetros de locação e outra para os parâmetros de escala, dadas, respectivamente, pelas expressões

$$g_1(\theta_{it}) = \eta_{1it} = f_1(\mathbf{x}_{it}, \boldsymbol{\beta}) \quad \text{e} \quad g_2(\phi_{it}) = \eta_{2it} = f_2(\mathbf{z}_{it}, \boldsymbol{\xi}), \quad (3.7)$$

em que $g_1(\cdot)$ e $g_2(\cdot)$ são funções de ligação conhecidas, estritamente monótonas e que admitem derivadas de segunda ordem, com $Im[g_1(\cdot)] = \mathbb{R}$ e $Im[g_2(\cdot)] = \mathbb{R}^+$. As funções $f_1(\cdot, \boldsymbol{\beta})$ e $f_2(\cdot, \boldsymbol{\xi})$ são contínuas e possuem derivadas de segunda ordem. Os vetores de parâmetros desconhecidos são $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, $\boldsymbol{\beta} \in \mathbb{R}^p$, e $\boldsymbol{\xi} = (\xi_1, \dots, \xi_q)'$, $\boldsymbol{\xi} \in \mathbb{R}^q$. As covariáveis que compõem \mathbf{x}_{it} e \mathbf{z}_{it} não são necessariamente diferentes.

A função log-verossimilhança dos vetores de parâmetros $\boldsymbol{\theta}$ e $\boldsymbol{\phi}$ para a amostra $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)'$, em que $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})'$, $t = 1, \dots, m_i$, $i = 1, \dots, n$, supondo independência entre as observações, é dada por

$$\ell(\boldsymbol{\theta}, \boldsymbol{\phi} | \mathbf{y}) = \sum_{i=1}^n \sum_{t=1}^{m_i} \left\{ -\log(\phi_{it}) - \frac{y_{it} - \theta_{it}}{\phi_{it}} - \exp\left(-\frac{y_{it} - \theta_{it}}{\phi_{it}}\right) \right\} \quad (3.8)$$

Considere a v.a. $Y_{it}^* = \exp(-Y_{it}/\phi_{it})$ com $E(Y_{it}^*) = \mu_{it}^* = \exp(-\theta_{it}/\phi_{it})$ e $Var(Y_{it}^*) = \nu_{it}^* = \mu_{it}^{*2}$, e seja $\lambda_{it} = Y_{it}^* - \mu_{it}^*$, em que $E(\lambda_{it}) = 0$ e $Var(\lambda_{it}) = \mu_{it}^{*2} = \nu_{it}^*$ (resultados obtidos das resoluções de algumas integrais detalhados no apêndice C).

A função escore de β_j , escrita diferentemente de (2.19), é dada por

$$\begin{aligned}
U_{\beta_j} = \frac{\partial \ell}{\partial \beta_j} &= \sum_{i=1}^n \sum_{t=1}^{m_i} \frac{\partial \ell_{it}}{\partial \theta_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \eta_{1it}}{\partial \beta_j} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} \left[\frac{1}{\phi_{it}} - \exp\left(-\frac{Y_{it} - \theta_{it}}{\phi_{it}}\right) \frac{1}{\phi_{it}} \right] \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \eta_{1it}}{\partial \beta_j} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} (Y_{it}^* - \mu_{it}^*) \left(-\frac{\mu_{it}^*}{\phi_{it}}\right) \frac{1}{\mu_{it}^{*2}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \eta_{1it}}{\partial \beta_j} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} \lambda_{it} \left(-\frac{\mu_{it}^*}{\phi_{it}}\right) \left(\frac{1}{\text{Var}(\lambda_{it})}\right) \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \eta_{1it}}{\partial \beta_j}.
\end{aligned}$$

Considere também a v.a. $\delta_{it} = \mu_{it}^* + [(Y_{it} - \theta_{it})/\phi_{it}](Y_{it}^* - \mu_{it}^*)$, em que $E(\delta_{it}) = 0$ e $\text{Var}(\delta_{it}) = c\mu_{it}^{*2} = c\nu_{it}^*$, sendo $c = \pi^2/6 + (\gamma - 1)^2$. Então, a função escore de ξ_l , escrita diferentemente de (2.20), é dada por

$$\begin{aligned}
U_{\xi_l} = \frac{\partial \ell}{\partial \xi_l} &= \sum_{i=1}^n \sum_{t=1}^{m_i} \frac{\partial \ell_{it}}{\partial \phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \frac{\partial \eta_{2it}}{\partial \xi_l} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} \left[-\frac{1}{\phi_{it}} + \frac{Y_{it} - \theta_{it}}{\phi_{it}^2} - \exp\left(-\frac{Y_{it} - \theta_{it}}{\phi_{it}}\right) \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}^2}\right) \right] \frac{\partial \phi_{it}}{\partial \eta_{2it}} \frac{\partial \eta_{2it}}{\partial \xi_l} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} -\frac{1}{\mu_{it}^* \phi_{it}} \left[\mu_{it}^* + \frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right] \frac{\partial \phi_{it}}{\partial \eta_{2it}} \frac{\partial \eta_{2it}}{\partial \xi_l} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} -\frac{\delta_{it}}{\mu_{it}^* \phi_{it}} \frac{c\mu_{it}^*}{c\mu_{it}^*} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \frac{\partial \eta_{2it}}{\partial \xi_l} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} \delta_{it} \left(-c \frac{\mu_{it}^*}{\phi_{it}}\right) \frac{1}{c\mu_{it}^{*2}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \frac{\partial \eta_{2it}}{\partial \xi_l} \\
&= \sum_{i=1}^n \sum_{t=1}^{m_i} \delta_{it} \left(-c \frac{\mu_{it}^*}{\phi_{it}}\right) \frac{1}{\text{Var}(\delta_{it})} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \frac{\partial \eta_{2it}}{\partial \xi_l}.
\end{aligned}$$

Na forma matricial, a função escore, $\mathbf{U}_\tau = (\mathbf{U}_\beta, \mathbf{U}_\xi)'$, de $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{\xi})'$ é dada por

$$\mathbf{U}_\tau = \left(\frac{\partial \ell}{\partial \boldsymbol{\beta}}, \frac{\partial \ell}{\partial \boldsymbol{\xi}} \right)' = \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\lambda}_i, \sum_{i=1}^n \tilde{\mathbf{Z}}_i' \mathbf{B}_i \boldsymbol{\Delta}_i^{-1} \boldsymbol{\delta}_i \right)'_{(p+q) \times 1}, \quad (3.9)$$

sendo $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{im_i})' = (Y_{i1}^* - \mu_{i1}^*, \dots, Y_{im_i}^* - \mu_{im_i}^*)'$, $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{im_i})'$,

$$Y_{it}^* = \exp(-Y_{it}/\phi_{it}), \quad \mu_{it}^* = E(Y_{it}^*) = \exp(-\theta_{it}/\phi_{it}), \quad \nu_{it}^* = \text{Var}(Y_{it}^*) = \mu_{it}^{*2},$$

$$\delta_{it} = \mu_{it}^* + [(Y_{it} - \theta_{it})/\phi_{it}] (Y_{it}^* - \mu_{it}^*), \quad \mathbf{V}_i^* = \text{Var}(\mathbf{Y}_i^*) = \text{diag} [\nu_{i1}^*, \dots, \nu_{im_i}^*]_{m_i},$$

$$\tilde{\mathbf{X}}_i = \left[\frac{\partial \eta_{1it}}{\partial \beta_j} \right]_{t,j}, \quad \mathbf{A}_i = \text{diag} \left[-\frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \right]_{m_i}, \quad \boldsymbol{\Lambda}_i = \text{Var}(\boldsymbol{\lambda}_i) = \mathbf{V}_i^*$$

$$\tilde{\mathbf{Z}}_i = \left[\frac{\partial \eta_{2it}}{\partial \xi_l} \right]_{t,l}, \quad \mathbf{B}_i = \text{diag} \left[-c \frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right]_{m_i} \quad \text{e} \quad \boldsymbol{\Delta}_i = \text{Var}(\boldsymbol{\delta}_i) = c \mathbf{V}_i^*$$

Dizemos que \mathbf{G} é uma função de estimação de $\boldsymbol{\theta}$, se esta estiver em função do parâmetro desconhecido $\boldsymbol{\theta}$ o qual desejamos estimar e dos dados amostrais \mathbf{x} , e suas raízes forem estimativas das componentes de $\boldsymbol{\theta}$, ou seja, $\mathbf{G}(\mathbf{x}, \hat{\boldsymbol{\theta}}) = \mathbf{0}$. Estamos interessados em construir funções de estimação que permitam obter estimadores com boas propriedades assintóticas. Alguns critérios de otimalidade de funções de estimação foram estabelecidos por Godambe (1960).

Assumindo independência intragrupo, podemos usar a função escore como função de estimação (3.9) e o estimador do vetor de parâmetros de regressão $\boldsymbol{\tau}$ pode ser obtido através da solução da equação de estimação $\mathbf{U}_\tau = \mathbf{0}$. Porém, as componentes dos vetores aleatórios $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ são correlacionadas, mesmo que os vetores sejam independentes entre si, sendo $\boldsymbol{\Sigma}_i$ a matriz de covariâncias do i -ésimo vetor. Isso implica que as componentes de cada vetor aleatório \mathbf{Y}_i^* , $i = 1, \dots, n$, também são correlacionadas. Portanto, não podemos usar \mathbf{U}_τ como função de estimação de $\boldsymbol{\tau}$, visto que esta foi construída assumindo independência entre as componentes de cada

vetor aleatório, e desta forma, $\mathbf{V}_i^* = \mathbf{V}_i^{*1/2} \mathbf{I}_{m_i} \mathbf{V}_i^{*1/2} = \text{Cov}(\mathbf{Y}_i^*)$ é a matriz de covariâncias de \mathbf{Y}_i^* , e a matriz de correlações de \mathbf{Y}_i^* possui estrutura diagonal. Então, faz-se necessário a formulação de uma nova função de estimação que leve em conta a estrutura de dependência.

Seguindo a ideia de Liang e Zeger (1986), que estendem a metodologia de modelos lineares generalizados para análise de dados longitudinais, propomos uma nova função de estimação inserindo uma estrutura de dependência na função escore (3.9), substituindo a matriz identidade pelas verdadeiras matrizes de correlações, ou seja, substituindo $\mathbf{\Lambda}_i = \text{Var}(\boldsymbol{\lambda}_i) = \mathbf{V}_i^{*1/2} \mathbf{I}_{m_i} \mathbf{V}_i^{*1/2}$ por $\mathbf{\Lambda}_i^0 = \text{Cov}(\boldsymbol{\lambda}_i) = \mathbf{V}_i^{*1/2} \mathbf{R}(\mathbf{Y}_i^*) \mathbf{V}_i^{*1/2}$ e $\mathbf{\Delta}_i = \text{Var}(\boldsymbol{\delta}_i) = c \mathbf{V}_i^{*1/2} \mathbf{I}_{m_i} \mathbf{V}_i^{*1/2}$ por $\mathbf{\Delta}_i^0 = \text{Cov}(\boldsymbol{\delta}_i) = c \mathbf{V}_i^{*1/2} \mathbf{R}(\boldsymbol{\delta}_i) \mathbf{V}_i^{*1/2}$, sendo $\mathbf{R}(\mathbf{Y}_i^*)$ e $\mathbf{R}(\boldsymbol{\delta}_i)$ as matrizes de correlações de \mathbf{Y}_i^* e $\boldsymbol{\delta}_i$, respectivamente. Assim, a função de estimação ótima de $\boldsymbol{\tau}$ é dada por

$$\boldsymbol{\Psi}_\tau^0 = (\boldsymbol{\Psi}_\beta^0, \boldsymbol{\Psi}_\xi^0) = \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i' \mathbf{A}_i (\mathbf{\Lambda}_i^0)^{-1} \boldsymbol{\lambda}_i, \sum_{i=1}^n \tilde{\mathbf{Z}}_i' \mathbf{B}_i (\mathbf{\Delta}_i^0)^{-1} \boldsymbol{\delta}_i \right)'. \quad (3.10)$$

Na prática, geralmente $\mathbf{R}(\mathbf{Y}_i^*)$ e $\mathbf{R}(\boldsymbol{\delta}_i)$ são desconhecidas, e portanto devemos fazer algumas suposições sobre estas verdadeiras matrizes de correlações, como substituí-las por matrizes de correlações de trabalho, $\tilde{\mathbf{R}}(\boldsymbol{\rho}_1)$ e $\tilde{\mathbf{R}}(\boldsymbol{\rho}_2)$, em que $\boldsymbol{\rho}_1$ e $\boldsymbol{\rho}_2$ são parâmetros a serem estimados, de forma que $\tilde{\mathbf{R}}(\boldsymbol{\rho}_1)$ e $\tilde{\mathbf{R}}(\boldsymbol{\rho}_2)$ satisfaçam as condições de matrizes de correlações. Com o intuito de reduzir a quantidade de parâmetros a serem estimados, Artes e Jørgesen (2000) sugerem assumir que $\text{Cov}(\boldsymbol{\lambda}, \boldsymbol{\delta}) = \mathbf{0}$ e que $\text{Cov}(\boldsymbol{\delta}) = \mathbf{I}_{m_i}$. Venezuela (2008) sugere que em modelos lineares generalizados com medidas repetidas, assumir $\text{Cov}(\boldsymbol{\delta}) = \mathbf{\Pi} \text{Cov}(\mathbf{Y}_i^*) \mathbf{\Pi}$, sendo $\mathbf{\Pi}$ uma matriz diagonal com valores do parâmetro canônico da distribuição da variável resposta envolvida no modelo, resulta em estimativas com vieses menores do que se assumir que $\text{Cov}(\boldsymbol{\delta}) = \mathbf{I}_{m_i}$, como Artes e Jørgesen (2000) propõem.

Nesta dissertação, assumimos que $\text{Cov}(\boldsymbol{\lambda}, \boldsymbol{\delta}) = \mathbf{0}$ e utilizamos uma matriz de correlação de trabalho, $\tilde{\mathbf{R}}(\boldsymbol{\rho})$, substituindo $\boldsymbol{\Lambda}_i^0 = \mathbf{V}_i^{*1/2} \mathbf{R}(\mathbf{Y}_i^*) \mathbf{V}_i^{*1/2}$ em (3.10) por $\boldsymbol{\Omega}_i = \mathbf{V}_i^{*1/2} \tilde{\mathbf{R}}(\boldsymbol{\rho}) \mathbf{V}_i^{*1/2}$ e $\boldsymbol{\Delta}_i^0 = c \mathbf{V}_i^{*1/2} \mathbf{R}(\boldsymbol{\delta}_i) \mathbf{V}_i^{*1/2}$ por $c \boldsymbol{\Omega}_i = c \mathbf{V}_i^{*1/2} \tilde{\mathbf{R}}(\boldsymbol{\rho}) \mathbf{V}_i^{*1/2}$. Assim, a função de estimação de $\boldsymbol{\tau}$ do modelo de regressão de valor extremo para dados agrupados é dada por

$$\boldsymbol{\Psi}_{\boldsymbol{\tau}} = \begin{pmatrix} \boldsymbol{\Psi}_{\boldsymbol{\beta}} \\ \boldsymbol{\Psi}_{\boldsymbol{\xi}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \boldsymbol{\lambda}_i \\ \sum_{i=1}^n \tilde{\mathbf{Z}}_i' \mathbf{D}_i \boldsymbol{\Omega}_i^{-1} \boldsymbol{\delta}_i \end{pmatrix} = \sum_{i=1}^n \mathbb{X}_i' \mathbb{A}_i \mathbf{b}_i \quad (3.11)$$

sendo $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{im_i})'$, $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{im_i})'$, $\mathbf{b}_i = (\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)'$,

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \frac{\partial \eta_{1it}}{\partial \beta_j} \end{bmatrix}_{t,j}, \quad \mathbf{A}_i = \text{diag} \left[-\frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \right]_{m_i}, \quad \mathbb{X}_i = \begin{pmatrix} \tilde{\mathbf{X}}_i & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{Z}}_i \end{pmatrix}$$

$$\tilde{\mathbf{Z}}_i = \begin{bmatrix} \frac{\partial \eta_{2it}}{\partial \xi_l} \end{bmatrix}_{t,l}, \quad \mathbf{D}_i = \text{diag} \left[-\frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right]_{m_i} \quad \text{e} \quad \mathbb{A}_i = \begin{pmatrix} \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_i \boldsymbol{\Omega}_i^{-1} \end{pmatrix}.$$

As suposições feitas sobre as estruturas de dependência diminuem a quantidade de parâmetros a serem estimados, mas fazem com que a função de estimação se afaste da função de estimação ótima (3.10). Então, para que a raiz de (3.11), $\hat{\boldsymbol{\tau}}$, seja um estimador consistente de $\boldsymbol{\tau}$ e que encontremos a distribuição limite do estimador, $\hat{\boldsymbol{\rho}}$ deve satisfazer algumas condições (vide Artes e Botter, 2005, Apêndice A).

Assim, quando for de interesse modelar dados que são oriundos de vetores aleatórios cujas componentes seguem uma distribuição valor extremo tipo I e são correlacionadas entre si, sugerimos utilizar a função de estimação expressa em (3.11) e não a função escore, visto que $\boldsymbol{\Psi}_{\boldsymbol{\tau}}$ leva em consideração a estrutura de correlação entre as componentes do vetor aleatório e a função escore não considera a estrutura.

3.1.1 Estimação de τ

Um estimador, $\hat{\tau}$, do vetor de parâmetros de regressão, τ , do modelo de regressão Gumbel para dados agrupados, é obtido através da solução da equação de estimação $\Psi_{\tau} = \mathbf{0}$.

Como o sistema de equações gerado de $\Psi_{\tau} = \mathbf{0}$ não tem solução analítica, utilizamos o algoritmo Newton-Raphson para resolvê-lo. As iterações da versão multivariada deste algoritmo, para solucionar a equação $\Psi_{\tau} = \mathbf{0}$, são obtidas por

$$\hat{\tau}^{(m+1)} = \hat{\tau}^{(m)} - \left(\frac{\partial \Psi_{\tau}^{(m)}}{\partial \tau} \right)^{-1} \Psi_{\tau}^{(m)}, \quad (3.12)$$

em que os índices m e $m + 1$, indicam, respectivamente, que os elementos da equação (3.12) estão sendo avaliados nos passos m e $m + 1$ do algoritmo.

Para facilitar os cálculos, substituímos $\partial \Psi_{\tau} / \partial \tau$ em (3.12) pela matriz de sensibilidade, dada por $\mathbf{S}_{\tau} = E(\partial \Psi_{\tau} / \partial \tau)$. Logo, obtemos as iterações do algoritmo modificado, conhecido como algoritmo Escore de Fisher, para solução de $\Psi_{\tau} = \mathbf{0}$, por

$$\hat{\tau}^{(m+1)} = \hat{\tau}^{(m)} - \left(\mathbf{S}_{\tau}^{(m)} \right)^{-1} \Psi_{\tau}^{(m)}, \quad (3.13)$$

Os blocos da matriz de sensibilidade são dadas por $\mathbf{S}_{\beta\beta} = E(\partial \Psi_{\beta} / \partial \beta)$, $\mathbf{S}_{\beta\xi} = E(\partial \Psi_{\beta} / \partial \xi)$, $\mathbf{S}_{\xi\beta} = E(\partial \Psi_{\xi} / \partial \beta)$ e $\mathbf{S}_{\xi\xi} = E(\partial \Psi_{\xi} / \partial \xi)$ (vide Apêndice B), e

$$\mathbf{S}_{\tau} = \begin{bmatrix} \mathbf{S}_{\beta\beta} & \mathbf{S}_{\beta\xi} \\ \mathbf{S}_{\xi\beta} & \mathbf{S}_{\xi\xi} \end{bmatrix} = - \sum_{i=1}^n \mathbf{X}_i' \mathbb{W}_i \mathbf{X}_i,$$

em que, para todo $i = 1, \dots, n$, a matriz \mathbb{W}_i é dada por

$$\mathbb{W}_i = \begin{bmatrix} \mathbf{W}_{(\beta\beta)_i} & \mathbf{W}_{(\beta\xi)_i} \\ \mathbf{W}_{(\xi\beta)_i} & \mathbf{W}_{(\xi\xi)_i} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \mathbf{A}_i & (\gamma - 1) \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \\ (\gamma - 1) \mathbf{D}_i \boldsymbol{\Omega}_i^{-1} \mathbf{A}_i & c \mathbf{D}_i \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \end{bmatrix}.$$

Podemos escrever as iterações do algoritmo Escore de Fisher (3.13) como um processo iterativo de mínimos quadrados ponderados, dado por

$$\begin{aligned}
\hat{\boldsymbol{\tau}}^{(m+1)} &= \hat{\boldsymbol{\tau}}^{(m)} - [\mathbf{S}_{\boldsymbol{\tau}}^{(m)}]^{-1} \boldsymbol{\Psi}_{\boldsymbol{\tau}}^{(m)} \\
&= [\mathbf{S}_{\boldsymbol{\tau}}^{(m)}]^{-1} \{ \mathbf{S}_{\boldsymbol{\tau}}^{(m)} \hat{\boldsymbol{\tau}}^{(m)} - \boldsymbol{\Psi}_{\boldsymbol{\tau}}^{(m)} \} \\
&= \left\{ \left[\sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{W}}_i \mathbb{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{W}}_i \mathbb{X}_i \hat{\boldsymbol{\tau}} + \sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{A}}_i \hat{\boldsymbol{b}}_i \right] \right\}^{(m)} \\
&= \left\{ \left[\sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{W}}_i \mathbb{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{W}}_i \mathbb{X}_i \hat{\boldsymbol{\tau}} + \sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{W}}_i \hat{\mathbb{W}}_i^{-1} \hat{\mathbb{A}}_i \hat{\boldsymbol{b}}_i \right] \right\}^{(m)} \\
&= \left\{ \left[\sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{W}}_i \mathbb{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbb{X}'_i \hat{\mathbb{W}}_i \boldsymbol{\zeta}_i \right] \right\}^{(m)} \tag{3.14}
\end{aligned}$$

em que $\boldsymbol{\zeta}_i = \mathbb{X}_i \hat{\boldsymbol{\tau}} + \hat{\mathbb{W}}_i^{-1} \hat{\mathbb{A}}_i \hat{\boldsymbol{b}}_i$ faz o papel de uma variável dependente transformada, \mathbb{X} da matriz do modelo (de planejamento) e $\hat{\mathbb{W}}$ de uma matriz de pesos. O índice m no lado direito da equação indica que as quantidades estão sendo avaliadas nas estimativas dos parâmetros no passo m , e a inversa da matriz \mathbb{W}_i é dada por

$$\mathbb{W}_i^{-1} = \begin{bmatrix} \mathbf{W}_i^{\beta\beta} & \mathbf{W}_i^{\beta\xi} \\ \mathbf{W}_i^{\xi\beta} & \mathbf{W}_i^{\xi\xi} \end{bmatrix},$$

em que as matrizes $\mathbf{W}_i^{\beta\beta}$, $\mathbf{W}_i^{\beta\xi}$, $\mathbf{W}_i^{\xi\beta}$ e $\mathbf{W}_i^{\xi\xi}$ são dados por

$$\begin{aligned}
\mathbf{W}_i^{\beta\beta} &= \left[\mathbf{W}_{(\beta\beta)_i} - \mathbf{W}_{(\beta\xi)_i} \mathbf{W}_{(\xi\xi)_i}^{-1} \mathbf{W}_{(\xi\beta)_i} \right]^{-1}, \\
\mathbf{W}_i^{\beta\xi} &= -\mathbf{W}_i^{\beta\beta} \mathbf{W}_{(\beta\xi)_i} \mathbf{W}_{(\xi\xi)_i}^{-1}, \\
\mathbf{W}_i^{\xi\beta} &= -\mathbf{W}_{(\xi\xi)_i}^{-1} \mathbf{W}_{(\xi\beta)_i} \mathbf{W}_i^{\beta\beta} \text{ e} \\
\mathbf{W}_i^{\xi\xi} &= \mathbf{W}_{(\xi\xi)_i}^{-1} \left[\mathbf{I} - \mathbf{W}_{(\xi\beta)_i} \mathbf{W}_i^{\beta\xi} \right].
\end{aligned}$$

O valor inicial, $\widehat{\boldsymbol{\tau}}^{(0)}$, do processo iterativo do método modificado de Newton escrito como mínimos quadrados ponderados (3.14) pode ser obtido como segue:

1. Como $\text{Var}(Y_{it}) = \pi^2 \phi_{it}^2 / 6$, assumindo que a variância intragrupal é constante, ou seja, $\text{Var}(Y_{it}) = \text{Var}(\mathbf{Y}_i)$, para todo $t = 1, \dots, m_i$, e que haja independência entre as observações de cada grupo, podemos estimar ϕ_{it} pelo método dos momentos, utilizando o estimador $\tilde{\phi}_i = (s_i \sqrt{6}) / \pi$, sendo s_i o desvio padrão amostral calculado com os elementos de \mathbf{y}_i ;
2. Como $E(Y_{it}) = \theta_{it} + \gamma \phi_{it}$, assumindo que $E(Y_{it}) = Y_{it}$ e $\phi_{it} = \tilde{\phi}_i$, um estimador de θ_{it} é dado por $\tilde{\theta}_{it} = Y_{it} - \gamma \tilde{\phi}_i$;
3. Encontrar os valores iniciais $\boldsymbol{\beta}^{(0)}$ ajustando um modelo de regressão não linear $g_1(\tilde{\theta}_{it}) = f_1(\mathbf{x}_{it}, \boldsymbol{\beta})$;
4. Calcular $\theta_{it}^{(0)} = g_1^{-1} [f_1(\mathbf{x}_{it}, \boldsymbol{\beta}^{(0)})]$ e $\phi_{it}^{(0)} = (Y_{it} - \theta_{it}^{(0)}) / \gamma$;
5. Encontrar os valores iniciais $\boldsymbol{\xi}^{(0)}$ ajustando um modelo de regressão não linear $g_2(\tilde{\phi}_{it}^{(0)}) = f_2(\mathbf{z}_{it}, \boldsymbol{\xi})$.

3.1.2 Propriedades assintóticas de $\widehat{\boldsymbol{\tau}}$

Para grandes amostras ($n \rightarrow \infty$) e sob certas condições de regularidade, $\widehat{\boldsymbol{\tau}}$ (3.14) é um estimador consistente de $\boldsymbol{\tau}$ (vide Artes e Botter, 2005, Teorema 10) e

$$\sqrt{n}(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \xrightarrow{D} N_{p+q}(\mathbf{0}, \mathbf{K}_{\boldsymbol{\tau}}^{-1}), \quad (3.15)$$

em que $\mathbf{K}_{\boldsymbol{\tau}} = \mathbf{S}'_{\boldsymbol{\tau}} \mathbf{T}_{\boldsymbol{\tau}}^{-1} \mathbf{S}_{\boldsymbol{\tau}}$ é a matriz de informação de Godambe, sendo $\mathbf{S}_{\boldsymbol{\tau}}$ a matriz de sensibilidade e $\mathbf{T}_{\boldsymbol{\tau}}$ a matriz de variabilidade definida por $\mathbf{T}_{\boldsymbol{\tau}} = \sum_{i=1}^n E \left(\boldsymbol{\psi}_{(\boldsymbol{\tau})_i} \boldsymbol{\psi}'_{(\boldsymbol{\tau})_i} \right)$, sendo $\boldsymbol{\psi}_{(\boldsymbol{\tau})_i} = (\boldsymbol{\psi}_{(\boldsymbol{\beta})_i}, \boldsymbol{\psi}_{(\boldsymbol{\xi})_i})' = \left(\tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \boldsymbol{\lambda}_i, \tilde{\mathbf{Z}}_i' \mathbf{D}_i \boldsymbol{\Omega}_i^{-1} \boldsymbol{\delta}_i \right)'$, e dada por

$$\mathbf{T}_\tau = \begin{bmatrix} \mathbf{T}_{\beta\beta} & \mathbf{T}_{\beta\xi} \\ \mathbf{T}_{\xi\beta} & \mathbf{T}_{\xi\xi} \end{bmatrix} = \sum_{i=1}^n \mathbf{X}'_i \mathbf{A}_i \mathbf{Q}_i \mathbf{A}'_i \mathbf{X}_i, \quad (3.16)$$

e como $E(\boldsymbol{\lambda}_i) = \mathbf{0}$ e $E(\boldsymbol{\delta}_i) = \mathbf{0}$, e estamos assumindo que $\text{Cov}(\boldsymbol{\delta}, \boldsymbol{\lambda}) = \mathbf{0}$, então

$$\mathbf{Q}_i = \begin{bmatrix} \text{Cov}(\boldsymbol{\delta}) & \mathbf{0} \\ \mathbf{0} & \text{Cov}(\boldsymbol{\lambda}) \end{bmatrix}, \quad \forall i = 1, \dots, n. \quad (3.17)$$

Um estimador consistente para a matriz de covariâncias de $\hat{\boldsymbol{\tau}}$ é dado por

$$\widehat{\mathbf{K}}_\tau^{-1} = \widehat{\mathbf{S}}_\tau^{-1} \left\{ \sum_{i=1}^n \begin{bmatrix} \widehat{\mathbf{X}}_i' \widehat{\mathbf{A}}_i \widehat{\boldsymbol{\Omega}}_i^{-1} \widehat{\boldsymbol{\lambda}}_i \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{\Omega}}_i'^{-1} \widehat{\mathbf{A}}_i \widehat{\mathbf{X}}_i & \widehat{\mathbf{X}}_i' \widehat{\mathbf{A}}_i \widehat{\boldsymbol{\Omega}}_i^{-1} \widehat{\boldsymbol{\lambda}}_i \widehat{\boldsymbol{\delta}}_i' \widehat{\boldsymbol{\Omega}}_i'^{-1} \widehat{\mathbf{D}}_i \widehat{\mathbf{Z}}_i \\ \widehat{\mathbf{Z}}_i' \widehat{\mathbf{D}}_i \widehat{\boldsymbol{\Omega}}_i^{-1} \widehat{\boldsymbol{\delta}}_i \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{\Omega}}_i'^{-1} \widehat{\mathbf{A}}_i \widehat{\mathbf{X}}_i & \widehat{\mathbf{Z}}_i' \widehat{\mathbf{D}}_i \widehat{\boldsymbol{\Omega}}_i^{-1} \widehat{\boldsymbol{\delta}}_i \widehat{\boldsymbol{\delta}}_i' \widehat{\boldsymbol{\Omega}}_i'^{-1} \widehat{\mathbf{D}}_i \widehat{\mathbf{Z}}_i \end{bmatrix} \right\} \widehat{\mathbf{S}}_\tau^{-1},$$

em que as estimativas das matrizes são obtidas através da substituição de $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{\xi})'$ e $\boldsymbol{\rho}$ pelas suas estimativas consistentes.

De posse da distribuição assintótica do estimador $\hat{\boldsymbol{\tau}}$, podemos construir intervalos de confiança assintóticos e elaborar testes de hipóteses para o vetor de parâmetros de regressão $\boldsymbol{\tau}$ do modelo de regressão de valor extremo para dados agrupados, ou para as componentes deste deste vetor.

3.1.3 Matriz de correlações de trabalho

Existem diversas formas de definir a estrutura de correlação e estimar o vetor de parâmetros de correlação $\boldsymbol{\rho}$. Uma delas consiste em desprezar a dependência entre as observações, e utilizar a matriz de correlações com estrutura diagonal, ou seja, a matriz de correlação de trabalho será um matriz identidade de ordem m_i , fazendo com que a função de estimação se reduza à função escore ($\boldsymbol{\Psi}_\tau = \mathbf{U}_\tau$).

Outra estrutura que pode ser utilizada para a matriz de correlações de trabalho $\tilde{\mathbf{R}}(\boldsymbol{\rho})$, é a não estruturada, definida em (3.6), em que cada par de observações do grupo possui seu próprio parâmetro de correlação. Um estimador \sqrt{n} -consistente da matriz de correlação de trabalho não estruturada é dado por

$$\hat{\mathbf{R}} = \frac{\sum_{i=1}^n \mathbf{r}_i^* \mathbf{r}_i^{*'}}{n - (p + q)}, \quad (3.18)$$

em que $\mathbf{r}_i^* = (r_{i1}^*, \dots, r_{im_i}^*)$ é o vetor de resíduos de Pearson das componentes do vetor \mathbf{y}_i^* . Por definição, o resíduo de Pearson para a observação y_{it}^* é dado por

$$\hat{r}_{it}^* = \frac{y_{it}^* - \hat{\mu}_{it}^*}{\sqrt{\hat{\nu}_{it}^*}} = \frac{y_{it}^* - \hat{\mu}_{it}^*}{\sqrt{(\hat{\mu}_{it}^*)^2}} = \frac{y_{it}^* - \hat{\mu}_{it}^*}{\hat{\mu}_{it}^*}. \quad (3.19)$$

Entre os dois métodos extremos, ou seja, utilizar uma estrutura de correlação que despreza qualquer dependência existente nos dados e utilizar uma matriz de correlações de trabalho não estruturada, existem outras alternativas para definir e estimar a estrutura de correlação. Uma delas é utilizar uma matriz de correlações de trabalho com estrutura permutável, definida em (3.3), em que, em cada grupo de observações, \mathbf{y}_i^* , a correlação é a mesma para quaisquer pares de observações. Um estimador consistente para ρ é

$$\hat{\rho} = \frac{\sum_{i=1}^n \sum_{t>t'} \hat{r}_{it}^* \hat{r}_{i't'}^*}{\sum_{i=1}^n \frac{1}{2} m_i (m_i - 1) - (p + q)}. \quad (3.20)$$

Pode-se também utilizar uma estrutura autoregressiva de ordem 1, definida em (3.5). Nessa estrutura, que é bastante utilizada, assume-se que a correlação diminui exponencialmente de acordo com o aumento da diferença de tempo das observações, e o estimador consistente de ρ é dado por

$$\hat{\rho} = \frac{\sum_{i=1}^n \sum_{t=1}^{m_i-1} r_{it}^* r_{i(t+1)}^*}{(\sum_{i=1}^n m_i) - n}. \quad (3.21)$$

A escolha correta da matriz de correlações de trabalho é de suma importância, visto que ela tem grande influência sobre a eficiência dos estimadores. Porém, na maioria dos estudos não conhecemos a verdadeira estrutura da matriz de correlação. Logo, devemos utilizar algum critério para selecionar a matriz de correlação de trabalho. Baseado no AIC (Akaike's information criteria), que é um critério bem difundido e bastante utilizado em seleções de modelos, Pan (2001) propôs o QIC, um critério para seleção de estruturas da matriz de correlações de trabalho, que é dado por

$$QIC = -2\ell(\hat{\boldsymbol{\tau}}(\tilde{\mathbf{R}})|\mathbb{D}) + 2\text{traço}(\widehat{\mathbf{J}}\widehat{\mathbf{K}}_{\boldsymbol{\tau}}^{-1}), \quad (3.22)$$

em que ℓ é a função log-verossimilhança (3.8), $\hat{\boldsymbol{\tau}}(\tilde{\mathbf{R}})$ é o vetor de estimadores sob o modelo com matriz de correlação de trabalho $\tilde{\mathbf{R}}$, $\mathbb{D} = (\mathbf{y}, \mathbf{x}, \mathbf{z})$ são os dados observados, $\widehat{\mathbf{J}} = -\hat{\mathbf{S}}[\hat{\boldsymbol{\tau}}(\tilde{\mathbf{R}}); \tilde{\mathbf{R}} = \mathbf{I}]$ é a matriz de sensibilidade avaliada sob o estimador $\hat{\boldsymbol{\tau}}(\tilde{\mathbf{R}})$ e a matriz de correlações de trabalho igual à matriz identidade e $\widehat{\mathbf{K}}_{\boldsymbol{\tau}}^{-1}$ é o estimador robusto da matriz de covariâncias de $\hat{\boldsymbol{\tau}}$. A estrutura da matriz de correlações de trabalho considerada mais adequada, é aquela que resulta em um valor de QIC mais próximo de zero.

Como estamos assumindo que o parâmetro de dispersão é desconhecido e difere de uma observação para outra, então devemos calcular QIC utilizando o máximo das estimativas do parâmetro de dispersão, obtidas a partir de todos os modelos candidatos. Isso deve ser feito para que as possíveis alterações nos valores de QIC sejam causadas somente por mudanças da matriz de correlação de trabalho.

3.1.4 Medidas de diagnóstico

Na convergência do processo iterativo escrito na forma de mínimos quadrados ponderados, podemos simplificar a expressão (3.14), retirando os somatórios na forma

$$\hat{\boldsymbol{\tau}} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} \right)^{-1} \mathbf{X}'\hat{\mathbf{W}}\boldsymbol{\zeta} \quad (3.23)$$

em que $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$, $\hat{\mathbf{W}} = \text{diag} \{ \hat{\mathbf{W}}_1, \dots, \hat{\mathbf{W}}_n \}$ e $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_n)'$. Podemos interpretar $\hat{\boldsymbol{\tau}}$, na expressão (3.23), como a solução de mínimos quadrados de uma regressão linear, com $\hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta}$ sendo a variável resposta e $\hat{\mathbf{W}}^{1/2}\mathbf{X}$ a matriz do modelo (de planejamento). Assim, o resíduo ordinário (\mathbf{r}_O), que é definido como a diferença entre o valor observado ($\hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta}$) e o valor ajustado ($\hat{\mathbf{W}}^{1/2}\mathbf{X}\hat{\boldsymbol{\tau}}$) é dado por

$$\begin{aligned} \mathbf{r}_O &= \hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta} - \hat{\mathbf{W}}^{1/2}\mathbf{X}\hat{\boldsymbol{\tau}} \\ &= \hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta} - \hat{\mathbf{W}}^{1/2}\mathbf{X} \left[\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} \right)^{-1} \mathbf{X}'\hat{\mathbf{W}}\boldsymbol{\zeta} \right] \\ &= \left[\mathbf{I} - \hat{\mathbf{W}}^{1/2}\mathbf{X} \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} \right)^{-1} \mathbf{X}'\hat{\mathbf{W}}^{1/2} \right] \hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta} \\ &= (\mathbf{I} - \mathbf{H}) \hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta} \end{aligned} \quad (3.24)$$

em que $\mathbf{H} = \text{diag} \{ \mathbf{H}_1, \dots, \mathbf{H}_n \}$ é uma matriz simétrica ($\mathbf{H}' = \mathbf{H}$) e idempotente ($\mathbf{H}\mathbf{H} = \mathbf{H}$) e, como na regressão linear, pode ser denominada de matriz de projeção ou matriz chapéu, e seus elementos são dados por $\mathbf{H}_i = \hat{\mathbf{W}}_i^{1/2}\mathbf{X}_i \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} \right)^{-1} \mathbf{X}'_i\hat{\mathbf{W}}_i^{1/2}$.

Como $\text{Cov}(\boldsymbol{\lambda}) \approx \boldsymbol{\Omega}$, $\text{Cov}(\boldsymbol{\delta}) \approx c\boldsymbol{\Omega}$, e estamos assumindo que $\text{Cov}(\boldsymbol{\lambda}, \boldsymbol{\delta}) = \mathbf{0}$, então

$$\text{Cov}(\mathbf{b}_i) \approx \begin{bmatrix} \boldsymbol{\Omega}_i^{-1} & \mathbf{0} \\ \mathbf{0} & c\boldsymbol{\Omega}_i^{-1} \end{bmatrix}, \quad \mathbf{b}_i = (\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)'$$

e as respectivas matrizes de covariâncias de $\boldsymbol{\zeta}_i$ e \mathbf{r}_O são dadas aproximadamente por

$$\begin{aligned}
\text{Cov}(\boldsymbol{\zeta}_i) &= \text{Cov}\left(\mathbb{X}_i \hat{\boldsymbol{\tau}} + \hat{\mathbb{W}}_i^{-1} \hat{\mathbf{A}}_i \mathbf{b}_i\right) \\
&= \hat{\mathbb{W}}_i^{-1} \hat{\mathbf{A}}_i \text{Cov}(\mathbf{b}_i) \hat{\mathbf{A}}_i' \hat{\mathbb{W}}_i^{-1} \\
&\approx \hat{\mathbb{W}}_i^{-1} \hat{\mathbb{M}}_i \hat{\mathbb{W}}_i^{-1},
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\mathbf{r}_O) &= \text{Cov}\left[(\mathbf{I} - \mathbf{H}) \hat{\mathbb{W}}^{1/2} \boldsymbol{\zeta}\right] \\
&= (\mathbf{I} - \mathbf{H}) \hat{\mathbb{W}}^{1/2} \text{Cov}(\boldsymbol{\zeta}) \hat{\mathbb{W}}^{1/2} (\mathbf{I} - \mathbf{H}) \\
&\approx (\mathbf{I} - \mathbf{H}) \hat{\mathbb{W}}^{1/2} \hat{\mathbb{W}}^{-1} \hat{\mathbb{M}} \hat{\mathbb{W}}^{-1} \hat{\mathbb{W}}^{1/2} (\mathbf{I} - \mathbf{H}) \\
&\approx (\mathbf{I} - \mathbf{H}) \hat{\mathbb{W}}^{-1/2} \hat{\mathbb{M}} \hat{\mathbb{W}}^{-1/2} (\mathbf{I} - \mathbf{H})
\end{aligned}$$

em que $\hat{\mathbb{M}} = \text{diag}\{\hat{\mathbb{M}}_1, \dots, \hat{\mathbb{M}}_n\}$ e para todo $i = 1, \dots, n$

$$\hat{\mathbb{M}}_i = \begin{bmatrix} \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \mathbf{A}_i & \mathbf{0} \\ \mathbf{0} & c \mathbf{D}_i \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \end{bmatrix}$$

Pontos de alavanca

Um ponto de alavanca é uma observação que se destaca das demais no espaço das variáveis explicativas (independentes) e que causa grandes mudanças nas estimativas de alguns (ou todos) parâmetros do modelo quando ela é omitida do conjunto de dados.

Para identificar pontos de alavancas em modelos de regressão linear, utiliza-se valores da matriz chapéu (\mathbf{H}). O uso da matriz chapéu para identificar pontos de alavanca foi estendido para diversos casos particulares de modelos como Pregibon (1981) para MLG e Venezuela et al. (2007) para MLG para medidas repetidas.

Dizemos que a observação y_{it} pode ser um ponto de alavanca no ajuste do vetor de parâmetros β se o t -ésimo elemento da diagonal principal de \mathbf{H}_i , h_{it} (leverage), apresentar um valor mais alto que os das demais observações e um ponto de alavanca no ajuste do vetor de parâmetros ξ se o $2t$ -ésimo elemento da diagonal principal de \mathbf{H}_i , apresentar um valor mais alto que os das demais observações, para $t = 1, \dots, m_i$.

Em relação aos grupos, podemos considerar o i -ésimo grupo como alavanca no ajuste do vetor de parâmetros β se $h_{\beta i} = \sum_{t=1}^{m_i} h_{it}/m_i$ for grande em relação aos $h_{\beta i'} = \sum_{t=1}^{m_{i'}} h_{i't}/m_{i'}$ para todo $i' \neq i$, e para verificar grupos alavanca no ajuste de ξ analisamos $h_{\xi i} = \sum_{t=m_i+1}^{2m_i} h_{it}/m_i$, para todo $i = 1, \dots, n$, sendo h_{it} o t -ésimo elemento da diagonal principal de \mathbf{H}_i , $t = 1, \dots, m_i, m_i + 1, \dots, 2m_i$.

Ponto aberrante

Dizemos que uma observação é aberrante em relação à variável resposta quando esta observação possui perfil diferente das outras observações e apresenta valor baixo na diagonal principal da matriz chapéu (\mathbf{H}).

Uma maneira de detectar pontos aberrantes é através da identificação de pontos que se destacam dos demais em um diagrama de ponto entre $r_{(pad)it}$ versus i , para todo $i = 1, \dots, n$ e $t = 1, \dots, m_i, m_i + 1, \dots, 2m_i$, sendo $r_{(pad)it}$ os resíduos padronizados que são definidos como

$$r_{(pad)it} = \frac{e^{(it)'} \left(\hat{\mathbb{W}}^{1/2} \zeta - \hat{\mathbb{W}}^{1/2} \mathbb{X} \hat{\tau} \right) e^{(it)}}{\sqrt{e^{(it)'} \text{Cov}(\mathbf{r}_O) e^{(it)}}},$$

em que $e^{(it)} = (e'_1, \dots, e'_n)'$, sendo e_k vetores nulos de dimensão $2m_k$ para todo $k \neq i$ e o vetor e_i , de dimensão $2m_i$, é um vetor com valor 1 na posição t e 0 nas outras posições, para $t = 1, \dots, m_i, m_i + 1, \dots, 2m_i$.

Ponto influente

Dizemos que uma observação é influente em relação à variável resposta quando esta observação possui perfil diferente das outras observações e apresenta valor grande na diagonal principal da matriz chapéu (\mathbf{H}).

A distância de Cook é uma medida que serve para detectar pontos influentes, mensurando a discrepância entre a estimativa do vetor de parâmetros da regressão considerando todas as observações ($\hat{\boldsymbol{\tau}}$) e a estimativa de $\boldsymbol{\tau}$ retirando a observação y_{it} do conjunto de dados ($\hat{\boldsymbol{\tau}}_{(it)}$).

Para obter $\hat{\boldsymbol{\tau}}_{(it)}$ em modelos lineares generalizados, Pregibon (1981) propõe uma aproximação baseada na primeira iteração do método Escore de Fisher quando o mesmo é iniciado em $\hat{\boldsymbol{\tau}}$. Uma extensão dessa aproximação para modelos de regressão com medidas repetidas foi apresentada por Venezuela et al. (2007). Para o modelo de regressão de valor extremo para dados agrupados, $\hat{\boldsymbol{\tau}}_{(it)}$ é dado por

$$\hat{\boldsymbol{\tau}}_{(it)} = \hat{\boldsymbol{\tau}} - \frac{\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1} \left(\mathbf{X}'\hat{\mathbf{W}}^{1/2} e^{(it)}\right) \left[e^{(it)'} \left(\hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta} - \hat{\mathbf{W}}^{1/2}\mathbf{X}\hat{\boldsymbol{\tau}}\right)\right]}{1 - e^{(it)'} \mathbf{H} e^{(it)}}$$

A distância de Cook com a deleção da observação y_{it} é definida por

$$\begin{aligned} \text{DCook}_{(it)} &= \frac{1}{p+q} (\hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\tau}}_{(it)})' \mathbf{X}'\hat{\mathbf{W}}\mathbf{X} (\hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\tau}}_{(it)}) \\ &= \frac{\left[e^{(it)'} \left(\hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta} - \hat{\mathbf{W}}^{1/2}\mathbf{X}\hat{\boldsymbol{\tau}}\right)\right]^2 e^{(it)'} \hat{\mathbf{W}}^{1/2} \mathbf{X} \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1} \mathbf{X}'\hat{\mathbf{W}}^{1/2} e^{(it)}}{(p+q) \left[1 - e^{(it)'} \mathbf{H} e^{(it)}\right]^2} \\ &= \frac{e^{(it)'} \text{Cov}(\mathbf{r}_O) e^{(it)} \left[e^{(it)'} \left(\hat{\mathbf{W}}^{1/2}\boldsymbol{\zeta} - \hat{\mathbf{W}}^{1/2}\mathbf{X}\hat{\boldsymbol{\tau}}\right)\right]^2 e^{(it)'} \mathbf{H} e^{(it)}}{e^{(it)'} \text{Cov}(\mathbf{r}_O) e^{(it)} (p+q) \left[1 - e^{(it)'} \mathbf{H} e^{(it)}\right]^2} \\ &= \left[r_{(pad)it}\right]^2 \frac{e^{(it)'} \text{Cov}(\mathbf{r}_O) e^{(it)} \left(e^{(it)'} \mathbf{H} e^{(it)}\right)}{(p+q) \left[1 - e^{(it)'} \mathbf{H} e^{(it)}\right]^2}. \end{aligned}$$

Uma observação y_{it} pode ser considerada ponto influente no ajuste do vetor de parâmetros β se $\text{DCook}_{(it)}$ apresentar um valor mais alto que os valores da distância de Cook para as demais observações, para $t = 1, \dots, m_i$ e um ponto influente no ajuste do vetor de parâmetros ξ se $\text{DCook}_{(it')}$ apresentar um valor mais alto que os valores da distância de Cook para as demais observações, para $t' = m_i + 1, \dots, 2m_i$.

Em relação aos grupos, podemos considerar o i -ésimo grupo como influente no ajuste do vetor de parâmetros β se $\text{DCook}_{\beta i} = \sum_{t=1}^{m_i} \text{DCook}_{(it)}/m_i$ for grande em relação aos $\text{DCook}_{\beta i'} = \sum_{t=1}^{m_{i'}} \text{DCook}_{(i't)}/m_{i'}$ para todos $i' \neq i$, e para verificar grupos alavancas no ajuste de ξ analisamos $\text{DCook}_{\xi i} = \sum_{t'=m_i+1}^{2m_i} \text{DCook}_{(it')}/m_i$, para todo $i = 1, \dots, n$.

De acordo com Paula (2012), a deleção de pontos talvez seja a técnica mais conhecida para avaliar o impacto da retirada de uma observação particular nas estimativas da regressão. Ao se detectar candidatos a pontos de alavanca, aberrantes ou influentes, deve-se verificar se essas observações exercem um peso desproporcional ou inferencial nas estimativas dos parâmetros do modelo, através de comparação das estimativas dos parâmetros da regressão com e sem a presença desses pontos. Ao encontrarmos explicações para os valores atípicos, podemos entender melhor a relação entre as covariáveis e o fenômenos que estamos estudando, e até ajuda a traçarmos estratégias de utilização do modelo ajustado. Todas as possíveis situações devem ser analisadas com bastante cuidado antes de qualquer decisão ser tomada em relação aos candidatos a pontos de alavanca, aberrantes ou influentes. A eliminação desses pontos deve ser deixada como a última opção de decisão.

No próximo capítulo realizaremos simulações Monte Carlo para avaliar a performance dos estimadores dos parâmetros do modelo de regressão de valor extremo para dados agrupados, proposto nesta dissertação.

Capítulo 4

Estudo de simulação

Neste capítulo apresentamos resultados de estudos de simulação Monte Carlo, conduzidos para avaliar o comportamento dos estimadores dos parâmetros do modelo de regressão Gumbel para dados correlacionados, proposto nesta dissertação.

4.1 Resultados numéricos

Para avaliar o comportamento dos estimadores dos parâmetros do modelo proposto, conduzimos um estudo de simulação Monte Carlo, utilizando a linguagem R (R Core Team, 2012). Simulamos valores de vetores aleatórios, com componentes correlacionadas, e cada componente seguindo distribuição Gumbel com parâmetro de locação (θ_{it}) e parâmetro de escala (ϕ_{it}) escritos, respectivamente, da forma

$$\log(\theta_{it}) = \beta_1 x_{it} + \exp(\beta_2 z_{it}) \quad \text{e} \quad \log(\phi_{it}) = \xi_1 + \xi_2 x_{it}, \quad (4.1)$$

para $t = 1, \dots, m_i$. Assumimos também que foi amostrada a mesma quantidade de observações em cada grupo, ou seja, $m_i = m$, $i = 1, \dots, n$.

Os valores foram simulados considerando $\beta_1 = 1$, $\beta_2 = -1$, $\xi_1 = -0.7$ e $\xi_2 = 0.5$, e as covariáveis x e z geradas aleatoriamente de $X \sim N(-1, 1)$ e $Z \sim U(0, 1)$. Utilizamos uma matriz de correlações com estrutura auto-regressiva de ordem 1.

Consideramos amostras com 5, 10, 20 e 50 grupos, cada grupo com 5, 10 e 20 observações, para três diferentes valores do parâmetro de correlação, (ρ), 0,1, 0,3 e 0,5. Em cada um dos 36 ($4 \times 3 \times 3$) cenários, geramos 10.000 replicações Monte Carlo, e calculamos estimativas do viés e da variância dos estimadores dos parâmetros de regressão do modelo considerado. Os resultados são mostrados nas Tabelas 4.1, 4.2 e 4.3, com os valores entre parênteses sendo as estimativas da variância. Para maiores detalhes sobre a geração dos valores, ver Apêndice A.

Através dos resultados apresentados nas tabelas, observamos que os estimadores apresentados nesta dissertação produzem boas estimativas para os parâmetros do modelo de regressão de valor extremo para dados agrupados, tanto para os parâmetros de regressão associados com a locação, quanto os associados à escala.

A Tabela 4.1 apresenta as estimativas do módulo do viés e variância para amostras com 5 observações em cada grupo. Nesse cenário, as estimativas dos vieses dos parâmetros β_1 , β_2 e ξ_2 foram bem pequenas, indicando que em média, as estimativas são muito próximas dos verdadeiros valores dos parâmetros, principalmente para as estimativas de β_1 . Os pequenos vieses desses estimadores diminuem à medida que aumenta a quantidade de grupos. Já o estimador de ξ_1 apresentou um viés considerável em amostras pequenas, porém, assim como os outros estimadores, teve o vício diminuído à medida que o número de unidades amostrais crescia.

Percebemos também que, quando os grupos possuem poucas unidades amostrais, o aumento da correlação, mesmo que de forma sutil, contribui para o aumento dos vieses de $\hat{\xi}_1$, mas, com o aumento do número de grupos, a dependência vai perdendo a influência sobre o vício deste estimador.

Tabela 4.1: Estimativas do módulo do Viés e da (Variância) dos estimadores dos parâmetros de regressão, para amostras com 5 observações em cada grupo.

| | | β_1 | β_2 | ξ_1 | ξ_2 |
|----------|---------------|----------------------|----------------------|----------------------|----------------------|
| $n = 5$ | $\rho = 0, 1$ | 0,00232 (0,01099) | 0,02368 (0,18541) | 0,11942 (0,10443) | 0,02554 (0,05116) |
| | $\rho = 0, 3$ | 0,00082 (0,01062) | 0,02258 (0,18051) | 0,14378 (0,11259) | 0,02828 (0,04816) |
| | $\rho = 0, 5$ | 0,00418 (0,00994) | 0,00029 (0,21428) | 0,15743 (0,12146) | 0,02532 (0,04498) |
| $n = 10$ | $\rho = 0, 1$ | 0,00406 (0,00412) | 0,00253 (0,05203) | 0,06259 (0,03770) | 0,01544 (0,01763) |
| | $\rho = 0, 3$ | 0,00322 (0,00405) | 0,00150 (0,06005) | 0,06620 (0,04000) | 0,01322 (0,01753) |
| | $\rho = 0, 5$ | 0,00099 (0,00383) | 0,00577 (0,06753) | 0,07514 (0,04549) | 0,01162 (0,01645) |
| $n = 20$ | $\rho = 0, 1$ | 0,00254 (0,00178) | 0,00113 (0,02361) | 0,03270 (0,01561) | 0,00844 (0,00740) |
| | $\rho = 0, 3$ | 0,00237 (0,00177) | 0,00051 (0,02752) | 0,03382 (0,01696) | 0,00684 (0,00745) |
| | $\rho = 0, 5$ | 0,00175 (0,00168) | 0,00443 (0,03184) | 0,03778 (0,01936) | 0,00627 (0,00689) |
| $n = 50$ | $\rho = 0, 1$ | 0,00124 (0,00062) | 0,00070 (0,00865) | 0,01402 (0,00533) | 0,00349 (0,00260) |
| | $\rho = 0, 3$ | 0,00114 (0,00063) | 0,00136 (0,01039) | 0,01332 (0,00589) | 0,00277 (0,00268) |
| | $\rho = 0, 5$ | 0,00045 (0,00060) | 0,00253 (0,01218) | 0,01270 (0,00696) | 0,00196 (0,00246) |

As Tabelas 4.2 e 4.3 apresentam as estimativas do módulo do viés e variância para amostras com 10 e 20 observações em cada grupo, respectivamente. Nesses cenários os estimadores eram mais precisos do que em amostras com 5 observações em cada grupo. Porém, assim como em amostras com 5 observações por grupos, o estimador de ξ_1 possui um vício maior do que os dos outros estimadores, mas, com o aumento

de 5 para 10 observações em cada grupos, o vício teve uma redução considerável, diminuindo ainda mais, quando se tinha 20 observações por grupo.

Tabela 4.2: Estimativas do módulo do Viés e da (Variância) dos estimadores dos parâmetros de regressão, para amostras com 10 observações em cada grupo.

| | | β_1 | β_2 | ξ_1 | ξ_2 |
|----------|--------------|----------------------|----------------------|----------------------|----------------------|
| $n = 5$ | $\rho = 0,1$ | 0,00432 (0,00426) | 0,00131 (0,05291) | 0,06137 (0,03776) | 0,01472 (0,01777) |
| | $\rho = 0,3$ | 0,00269 (0,00416) | 0,00163 (0,06451) | 0,07096 (0,04101) | 0,01389 (0,01773) |
| | $\rho = 0,5$ | 0,00258 (0,00401) | 0,01374 (0,07425) | 0,07884 (0,04787) | 0,01057 (0,01603) |
| $n = 10$ | $\rho = 0,1$ | 0,00206 (0,00180) | 0,00367 (0,02326) | 0,03372 (0,01567) | 0,00985 (0,00732) |
| | $\rho = 0,3$ | 0,00249 (0,00176) | 0,00119 (0,02878) | 0,03391 (0,01691) | 0,00814 (0,00725) |
| | $\rho = 0,5$ | 0,00013 (0,00169) | 0,00553 (0,03480) | 0,03796 (0,01999) | 0,00564 (0,00674) |
| $n = 20$ | $\rho = 0,1$ | 0,00174 (0,00079) | 0,00163 (0,01110) | 0,01683 (0,00707) | 0,00420 (0,00335) |
| | $\rho = 0,3$ | 0,00148 (0,00083) | 0,00033 (0,01418) | 0,01627 (0,00773) | 0,00287 (0,00334) |
| | $\rho = 0,5$ | 0,00059 (0,00077) | 0,00412 (0,01670) | 0,01923 (0,00916) | 0,00246 (0,00310) |
| $n = 50$ | $\rho = 0,1$ | 0,00081 (0,00030) | 0,00217 (0,00418) | 0,00560 (0,00252) | 0,00107 (0,00122) |
| | $\rho = 0,3$ | 0,00035 (0,00030) | 0,00021 (0,00538) | 0,00605 (0,00283) | 0,00090 (0,00126) |
| | $\rho = 0,5$ | 0,00031 (0,00029) | 0,00225 (0,00653) | 0,00819 (0,00357) | 0,00175 (0,00115) |

Tabela 4.3: Estimativas do módulo do Viés e da (Variância) dos estimadores dos parâmetros de regressão, para amostras com 20 Observações em cada grupo.

| | | β_1 | β_2 | ξ_1 | ξ_2 |
|----------|--------------|----------------------|----------------------|----------------------|----------------------|
| $n = 5$ | $\rho = 0,1$ | 0,00212 (0,00179) | 0,00222 (0,02377) | 0,03033 (0,01545) | 0,00846 (0,00715) |
| | $\rho = 0,3$ | 0,00156 (0,00175) | 0,00017 (0,02981) | 0,03375 (0,01768) | 0,00667 (0,00735) |
| | $\rho = 0,5$ | 0,00013 (0,00179) | 0,00914 (0,03687) | 0,03906 (0,02029) | 0,00573 (0,00694) |
| $n = 10$ | $\rho = 0,1$ | 0,00136 (0,00080) | 0,00104 (0,01128) | 0,01472 (0,00692) | 0,00322 (0,00330) |
| | $\rho = 0,3$ | 0,00177 (0,00084) | 0,00024 (0,01429) | 0,01713 (0,00763) | 0,00354 (0,00331) |
| | $\rho = 0,5$ | 0,00027 (0,00082) | 0,00428 (0,01760) | 0,01998 (0,00904) | 0,00343 (0,00303) |
| $n = 20$ | $\rho = 0,1$ | 0,00095 (0,00039) | 0,00047 (0,00548) | 0,00818 (0,00339) | 0,00219 (0,00158) |
| | $\rho = 0,3$ | 0,00080 (0,00040) | 0,00004 (0,00696) | 0,00878 (0,00374) | 0,00238 (0,00162) |
| | $\rho = 0,5$ | 0,00025 (0,00038) | 0,00347 (0,00873) | 0,00958 (0,00428) | 0,00157 (0,00140) |
| $n = 50$ | $\rho = 0,1$ | 0,00035 (0,00015) | 0,00089 (0,00226) | 0,00304 (0,00129) | 0,00053 (0,00060) |
| | $\rho = 0,3$ | 0,00022 (0,00015) | 0,00036 (0,00276) | 0,00338 (0,00142) | 0,00091 (0,00062) |
| | $\rho = 0,5$ | 0,00013 (0,00015) | 0,00077 (0,00341) | 0,00390 (0,00168) | 0,00087 (0,00057) |

A quantidade de grupos influencia também na variabilidade dos estimadores, pois, de acordo com a Figura 4.1, a variância dos estimadores diminui à medida que aumentamos o número de grupos. Esses resultados sugerem que os estimadores dos parâmetros do modelo de regressão de valor extremo para dados agrupados, proposto nesta dissertação, são consistentes.

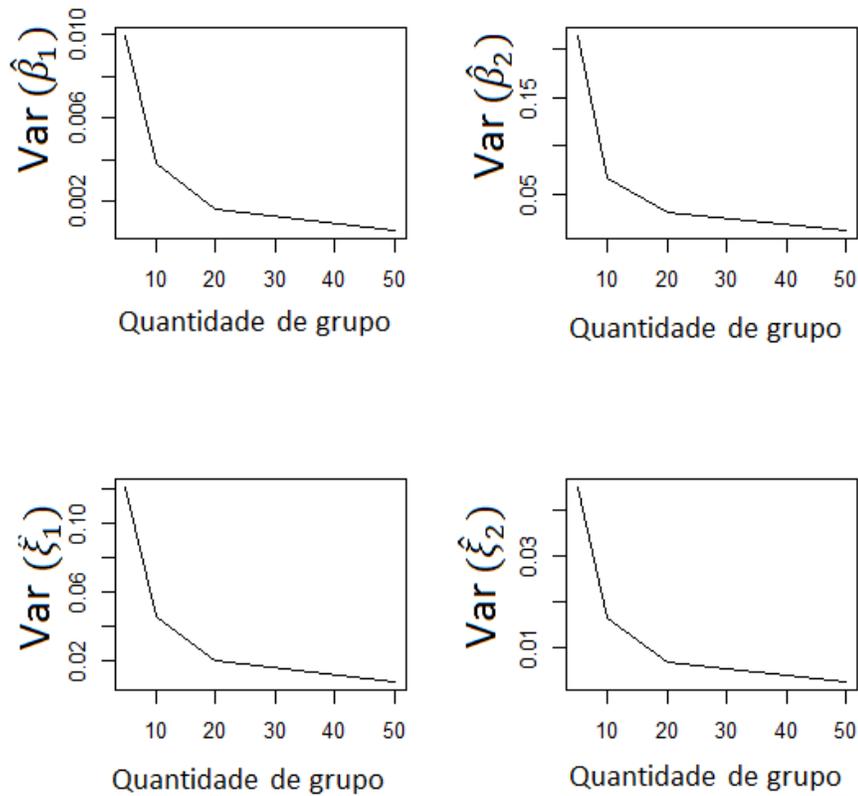


Figura 4.1: Estimativas das variâncias de $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\xi}_1$ e $\hat{\xi}_2$ para amostras 5, 10, 20 e 50 grupos com 5 observações em cada grupo.

Para avaliar a influência que a matriz de correlações de trabalho tem sobre os estimadores, simulamos dados com matrizes de correlações com estrutura da forma AR(1) e Diagonal, e estimamos os parâmetros utilizando matrizes de correlações de trabalho com estruturas AR(1) e Diagonal para amostras com 10, 20, 30, 40, 50 e 100 grupos. Os resultados destas simulações encontram-se nas Tabelas 4.4 e 4.5.

Tabela 4.4: Estimativas do módulo do Viés e da (Variância) dos estimadores dos parâmetros de regressão, gerando dados com matriz de correlação AR(1) com $\rho = 0.5$, para amostras com 10 observações em cada grupo.

| | | β_1 | β_2 | ξ_1 | ξ_2 |
|-----------|------------|----------------------|----------------------|----------------------|----------------------|
| $n = 10$ | R = AR(1) | 0,00052 (0,00183) | 0,00473 (0,03469) | 0,03880 (0,01862) | 0,00693 (0,00727) |
| | R=Diagonal | 0,00024 (0,00228) | 0,01389 (0,03951) | 0,04398 (0,01885) | 0,01037 (0,00723) |
| $n = 20$ | R = AR(1) | 0,00057 (0,00079) | 0,00998 (0,01635) | 0,01641 (0,00873) | 0,00037 (0,00329) |
| | R=Diagonal | 0,00094 (0,00096) | 0,01306 (0,01957) | 0,01819 (0,00889) | 0,00106 (0,00337) |
| $n = 30$ | R = AR(1) | 0,00077 (0,00048) | 0,00433 (0,01124) | 0,00936 (0,00591) | 0,00022 (0,00203) |
| | R=Diagonal | 0,00156 (0,00061) | 0,00498 (0,01362) | 0,00993 (0,00619) | 0,00040 (0,00216) |
| $n = 40$ | R = AR(1) | 0,00053 (0,00036) | 0,00531 (0,00802) | 0,00837 (0,00422) | 0,00130 (0,00144) |
| | R=Diagonal | 0,00078 (0,00047) | 0,00343 (0,00957) | 0,01046 (0,00423) | 0,00202 (0,00155) |
| $n = 50$ | R = AR(1) | 0,00028 (0,00028) | 0,00210 (0,00595) | 0,00565 (0,00309) | 0,00083 (0,00115) |
| | R=Diagonal | 0,00095 (0,00038) | 0,00036 (0,00718) | 0,00734 (0,00319) | 0,00168 (0,00120) |
| $n = 100$ | R = AR(1) | 0,00047 (0,00014) | 0,00260 (0,00326) | 0,00469 (0,00163) | 0,00112 (0,00056) |
| | R=Diagonal | 0,00027 (0,00020) | 0,00041 (0,00404) | 0,00478 (0,00170) | 0,00108 (0,00059) |

Tabela 4.5: Estimativas do módulo do Viés e da (Variância) dos estimadores dos parâmetros de regressão, gerando dados com matriz de correlação Diagonal, para amostras com 10 observações em cada grupo.

| | | β_1 | β_2 | ξ_1 | ξ_2 |
|-----------|------------|----------------------|----------------------|----------------------|----------------------|
| $n = 10$ | R = AR(1) | 0,00287 (0,00184) | 0,00549 (0,02120) | 0,03080 (0,01351) | 0,00770 (0,00738) |
| | R=Diagonal | 0,00307 (0,00180) | 0,00587 (0,02073) | 0,03135 (0,01318) | 0,00798 (0,00709) |
| $n = 20$ | R = AR(1) | 0,00153 (0,00080) | 0,00170 (0,00988) | 0,01742 (0,00688) | 0,00365 (0,00343) |
| | R=Diagonal | 0,00154 (0,00079) | 0,00183 (0,00992) | 0,01765 (0,00685) | 0,00383 (0,00342) |
| $n = 30$ | R = AR(1) | 0,00123 (0,00050) | 0,00142 (0,00665) | 0,01093 (0,00430) | 0,00308 (0,00203) |
| | R=Diagonal | 0,00133 (0,00051) | 0,00149 (0,00668) | 0,01126 (0,00431) | 0,00333 (0,00203) |
| $n = 40$ | R = AR(1) | 0,00046 (0,00036) | 0,00365 (0,00486) | 0,00514 (0,00319) | 0,00081 (0,00160) |
| | R=Diagonal | 0,00045 (0,00036) | 0,00372 (0,00487) | 0,00507 (0,00320) | 0,00072 (0,00159) |
| $n = 50$ | R = AR(1) | 0,00091 (0,00032) | 0,00465 (0,00384) | 0,00713 (0,00243) | 0,00100 (0,00125) |
| | R=Diagonal | 0,00089 (0,00032) | 0,00447 (0,00383) | 0,00705 (0,00241) | 0,00092 (0,00123) |
| $n = 100$ | R = AR(1) | 0,00016 (0,00015) | 0,00387 (0,00200) | 0,00140 (0,00136) | 0,00010 (0,00063) |
| | R=Diagonal | 0,00012 (0,00015) | 0,00389 (0,00199) | 0,00143 (0,00135) | 0,00012 (0,00063) |

De acordo com a Figura 4.2, quando os dados são medições de variáveis com matrizes de correlações com estrutura AR(1), a variância dos estimadores que utilizam matriz de correlações de trabalho com estrutura AR(1) é menor do que a variância dos estimadores que utilizam a matriz de correlações com estrutura diagonal.

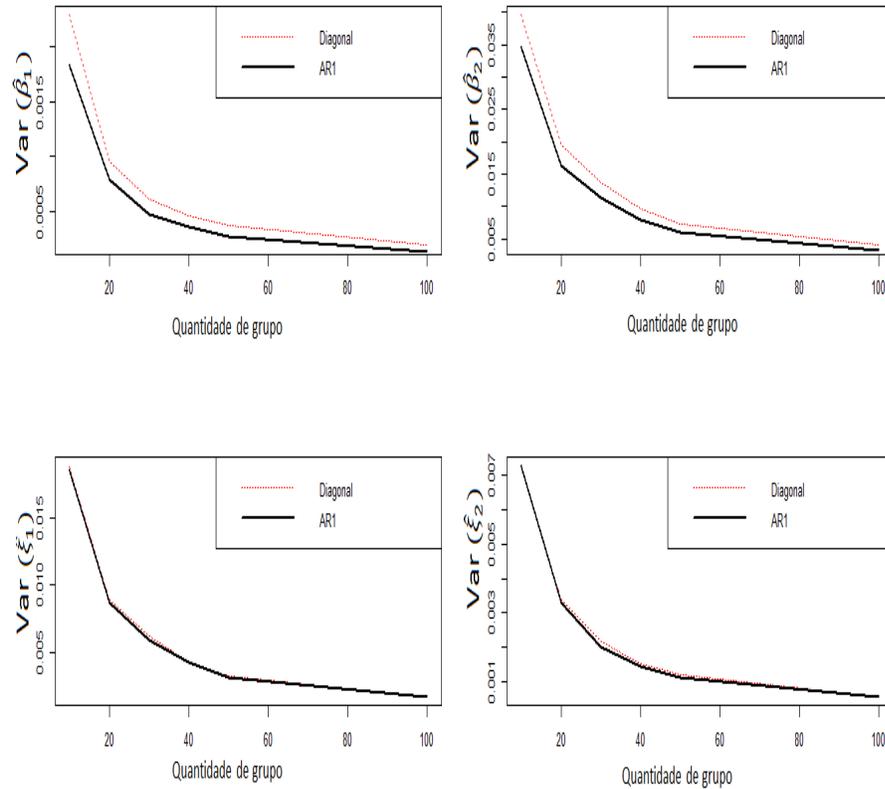


Figura 4.2: Estimativas das variâncias de $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\xi}_1$ e $\hat{\xi}_2$ para amostras com 10, 20, 30, 40, 50 e 100 grupos com 10 observações em cada grupo, geradas com matriz de correlações com estrutura AR(1).

Quando os dados são oriundos de variáveis com matrizes de correlações com estrutura diagonal (vide Figura 4.3), a variância dos estimadores que utilizam matriz de correlações de trabalho com estrutura AR(1) é praticamente igual à variância dos estimadores que utilizam a matriz de correlações com estrutura diagonal.

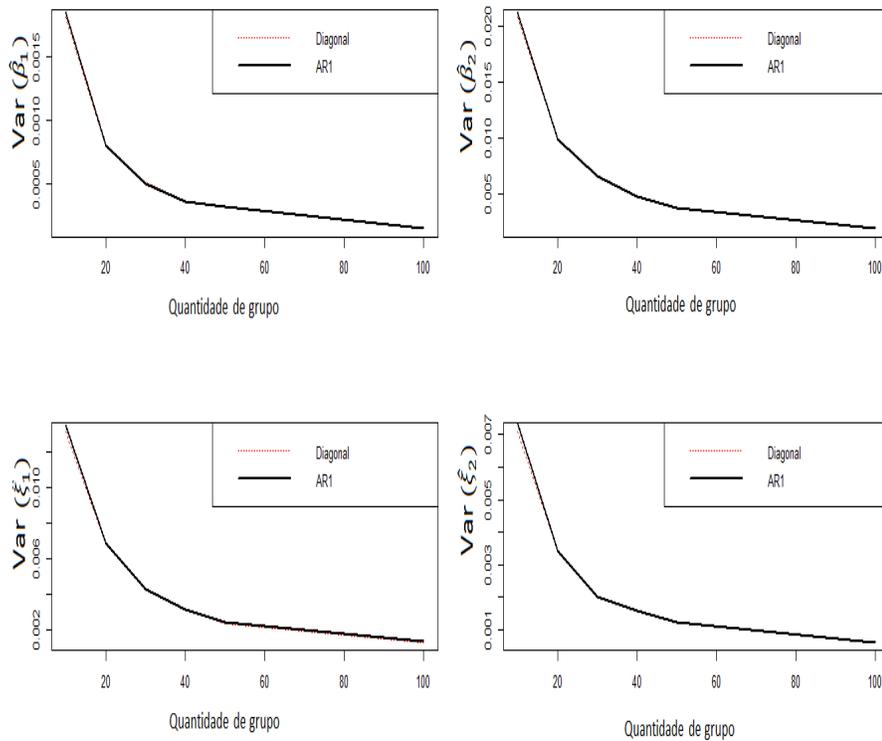


Figura 4.3: Estimativas das variâncias de $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\xi}_1$ e $\hat{\xi}_2$ para amostras com 10, 20, 30, 40, 50 e 100 grupos com 10 observações em cada grupo, geradas com matriz de correlações com estrutura Diagonal.

Através de simulações Monte Carlo, avaliamos também o quão preciso é o critério de seleção da estrutura da matriz de correlações de trabalho. Para tanto, simulamos dados com matriz de correlações com estrutura AR(1) e Diagonal para amostras com 10, 20, 30, 40, 50 e 100 grupos, com 10 observações em cada grupo, e em cada uma delas estimamos os parâmetros utilizando matrizes de correlações de trabalho com estruturas AR(1) e Diagonal, e calculamos os valores de QIC. A matriz de correlações de trabalho mais indicada é aquela que apresentar menor valor de QIC.

Tabela 4.6: Escolha da matriz de correlações de trabalho baseada no QIC.

| | Estrutura real: AR(1) | | Estrutura real: Diagonal | |
|-----------|-------------------------------------|--|-------------------------------------|--|
| | $\tilde{\mathbf{R}} = \text{AR}(1)$ | $\tilde{\mathbf{R}} = \text{Diagonal}$ | $\tilde{\mathbf{R}} = \text{AR}(1)$ | $\tilde{\mathbf{R}} = \text{Diagonal}$ |
| $n = 10$ | 62,4% | 37,6% | 51,4% | 48,6% |
| $n = 20$ | 69,9% | 30,1% | 49,4% | 50,6% |
| $n = 30$ | 78,1% | 21,9% | 48,3% | 51,7% |
| $n = 40$ | 81,6% | 18,4% | 51,4% | 48,6% |
| $n = 50$ | 82,2% | 17,8% | 49,7% | 50,3% |
| $n = 100$ | 88,6% | 11,4% | 48,4% | 51,6% |

A Tabela 4.6 apresenta as porcentagens de vezes que cada estrutura foi indicada de acordo com a verdadeira matriz de correlações e o número de grupos em cada amostra. Percebemos que para amostras geradas com estrutura AR(1), o QIC indicava a verdadeira estrutura (AR(1)) mais vezes que a estrutura diagonal, sendo que quanto maior o número de grupos, maior o acerto do QIC. E para amostras geradas com estrutura Diagonal, para os diferentes cenários analisados em termos de número de grupos, em metade das amostras o QIC indicava estrutura diagonal (verdadeira) e em metade das amostras indicava estrutura AR1. Por certo, isso acontece pelo fato de que, quando os dados são gerados com matrizes de correlação com estrutura diagonal $\hat{\rho} \approx 0$ fazendo com que $\tilde{\mathbf{R}} \approx \mathbf{I}$.

Para avaliar as medidas de diagnóstico, geramos uma amostra com 50 grupos, com 10 observações em cada grupo. As covariáveis, estrutura dos parâmetros da distribuição da resposta (equação. 4.1) e valores dos parâmetros de regressão foram os mesmos das outras simulações. Ajustamos o modelo proposto nesta dissertação aos dados e as estimativas dos parâmetros, obtidas com menos de 10 iterações do algoritmo, foram muito próximas do verdadeiros valores dos parâmetros: $\hat{\beta}_1 = 0,9963038$, $\hat{\beta}_2 = -0,9704523$, $\hat{\xi}_1 = -0,7074544$ e $\hat{\xi}_2 = 0,4940250$.

Calculamos o resíduo de Pearson, $\hat{r} = (y - \hat{\mu})/\sqrt{\widehat{\text{Var}}(Y)}$, e construímos o gráfico de resíduos de Pearson *versus* o grupo (Figura 4.4), o qual mostra que os resíduos estão distribuídos em torno do zero.

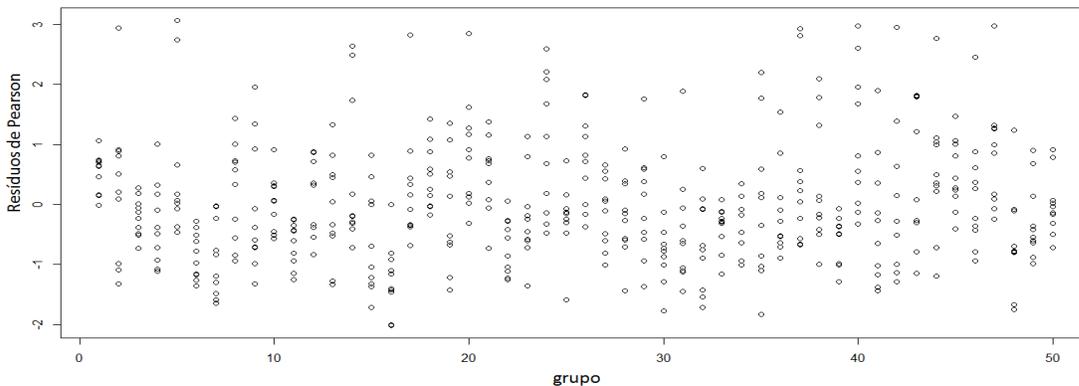


Figura 4.4: Resíduo de Pearson de y versus grupo.

Nessa amostra, a 10^a observação do 50^o grupo, apresentava os seguintes valores de variáveis respostas e covariáveis: $y_{(50,10)} = 1,826970371$, $x_{(50,10)} = -0,039403214$ e $z_{(50,10)} = 0,536839913$. As medidas de diagnóstico de modelo para essa observação foram: leverage ($h_{\beta(50,10)} = 0,005530629$ e $h_{\xi(50,10)} = 0,004016036$), resíduo padronizado ($r_{(pad)\beta(50,10)} = 1,69613957$ e $r_{(pad)\xi(50,10)} = -0,54341233$) e Distância de Cook ($\text{DCook}_{(\beta(50,10))} = 0,004640865$ e $\text{DCook}_{(\xi(50,10))} = 0,0003434844$).

Para avaliar a sensibilidade da leverage, modificamos o perfil da 10^a observação do 50^o grupo em relação às covariáveis, alterando seus valores de covariáveis para $x_{(50,10)} = -8,0$ e $z_{(50,10)} = -0,1$, e mantendo o valor da resposta. Com essa alteração, o valor de $h_{\beta(50,10)}$ não apresentou grande mudança, $h_{\beta(50,10)} = 0,005909212$, mas houve uma mudança significativa no valor da leverage dessa observação em relação a ξ , $h_{\xi(50,10)} = 0,1055051$, destacando-se das demais, conforme pode ser verificado

na Figura 4.5, indicando que com os valores alterados, essa observação seria um ponto de alavanca. Com a alteração nos valores de suas covariáveis, essa observação influenciava as estimativas dos parâmetros, que nesse cenário foram $\hat{\beta}_1 = 0,9536432$, $\hat{\beta}_2 = -1,0428038$, $\hat{\xi}_1 = -0,8873629$ e $\hat{\xi}_2 = 0,2204526$.

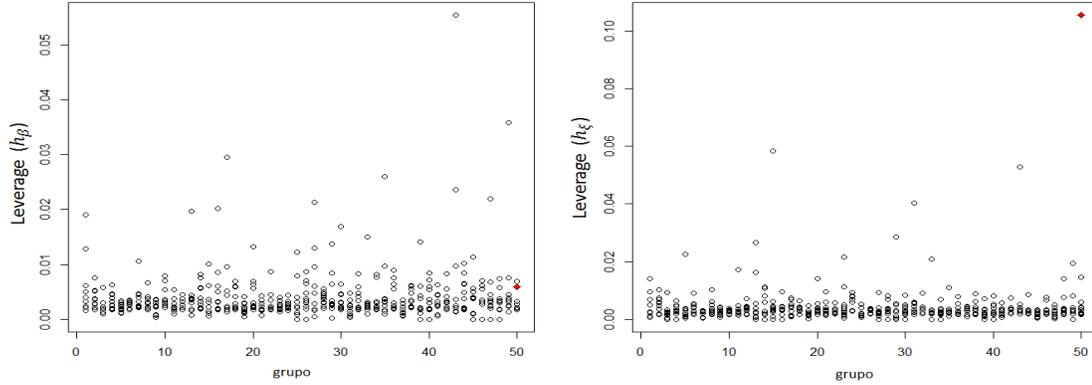


Figura 4.5: Leverages versus grupo.

Para avaliar a sensibilidade dos resíduos padronizados, modificamos o perfil da 10^a observação do 50^o grupo em relação à variável resposta, alterando o valor da resposta para $y_{(50,10)} = 13,0$, mantendo seus valores de covariáveis. Com essa alteração, os valores dos resíduos padronizados dessa observação mudaram para $r_{(pad)}\beta_{(50,10)} = 1,40159019$ e $r_{(pad)}\xi_{(50,10)} = 18,21103413$, destacando-se dos demais, conforme pode ser verificado na Figura 4.6, indicando que com os valores alterados, essa observação seria um ponto aberrante. Com a alteração no valor da resposta, essa observação influenciava as estimativas dos parâmetros, que nesse cenário foram $\hat{\beta}_1 = 0,9945247$, $\hat{\beta}_2 = -0,9419775$, $\hat{\xi}_1 = -0,6446260$ e $\hat{\xi}_2 = 0,5016673$.

Para avaliar a sensibilidade da distância de Cook, modificamos o perfil da 10^a observação do 50^o grupo, em relação à variável resposta e as covariáveis, alterando seus valores para $y_{(50,10)} = 13,0$, $x_{(50,10)} = -3,0$ e $z_{(50,10)} = -0,1$. Com

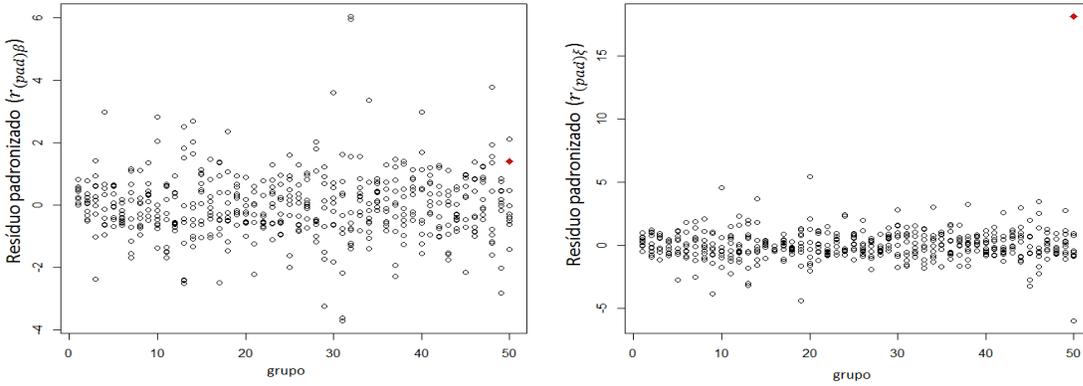


Figura 4.6: Resíduos padronizados versus grupo.

essa alteração, os valores das distâncias de Cook dessa observação mudaram para $DCook_{(\beta(50,10))} = 0,01734812$ e $DCook_{(\beta(50,10))} = 7,102191$, destacando-se dos demais, conforme pode ser verificado na Figura 4.7, indicando que com os valores alterados, essa observação seria um ponto influente. Com as alterações, essa observação influenciou as estimativas dos parâmetros, que nesse cenário foram $\hat{\beta}_1 = 0,9536223$, $\hat{\beta}_2 = -1,0008745$, $\hat{\xi}_1 = -0,7887745$ e $\hat{\xi}_2 = 0,2550430$.

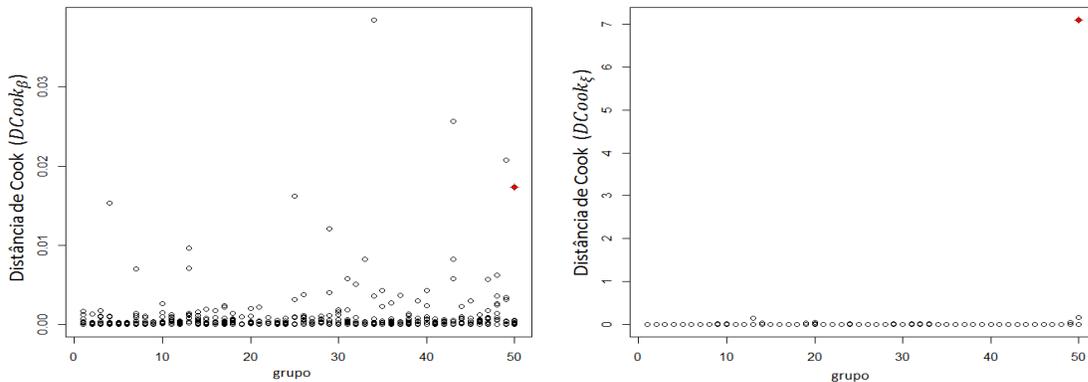


Figura 4.7: Distâncias de Cook versus grupo.

Capítulo 5

Discussões e considerações finais

Nesta dissertação apresentamos um modelo de regressão de valor extremo para dados agrupados. A motivação para este trabalho é o fato de que, em algumas situações, os dados disponíveis para modelagem de eventos raros podem apresentar alguma estrutura de dependência, e assim, usar os modelos de regressão de valor extremo disponíveis na literatura não seria adequado, visto que estes assumem independência entre as observações.

A construção do modelo foi baseada em modelos lineares generalizados e equações de estimação generalizadas. Consideramos a existência de grupos independentes, com componentes correlacionadas, cada componente sendo uma variável aleatória seguindo distribuição Gumbel com parâmetros de locação e escala escritos como funções não lineares de covariáveis.

Descrevemos o desenvolvimento teórico do modelo de regressão de valor extremo para dados agrupados, além da obtenção das funções de estimação e da matriz de informação. Também relatamos métodos de estimação dos parâmetros de regressão e para os parâmetros de correlação. O algoritmo de estimação foi escrito como um

processo iterativo de mínimos quadrados ponderados, e para encontrar os valores iniciais deste processo, propomos um algoritmo baseado em modelos de regressão não lineares.

Algumas técnicas de diagnóstico para o modelo de regressão de valor extremo para dados agrupados também foram propostas, como medidas baseadas na matriz de projeção para identificação de pontos de alavanca, os resíduos padronizados para identificar pontos aberrantes e a distância de Cook para detectar pontos influentes. Para selecionar a estrutura da matriz de correlações de trabalho utilizamos o QIC.

Para avaliar o comportamento dos estimadores dos parâmetros da regressão e algumas medidas proposta nesta dissertação, conduzimos um estudo de simulação Monte Carlo.

Os resultados do estudo de simulação mostraram que, para grandes amostras, as estimativas dos parâmetros de regressão estão muito próximas dos verdadeiros valores dos parâmetros, indicando que o modelo e o método de estimação apresentados nesta dissertação, são úteis quando se deseja modelar eventos extremos (máximos) por meio de covariáveis, e os dados disponíveis não são independentes, inviabilizando a utilização dos modelos já existentes na literatura.

Como continuação deste trabalho, faremos a aplicação do modelo aqui proposto a dados reais.

Como propostas para trabalhos futuros, pretendemos continuar o desenvolvimento de métodos de diagnóstico para o modelo de regressão de valor extremo para dados agrupados, com o estudo de influência local, que de acordo com Paula (2012) é um dos métodos mais modernos de diagnóstico de modelos. Neste método, em vez de avaliar o ajuste do modelo com a retirada individual ou conjunta de pontos, Cook (1986) propõe avaliar a influência conjunta das observações sob mudanças (perturbações ou influência local) modestas nos dados ou no modelo, ou seja, verificar

a existência de pontos que sob pequenas modificações no modelo causam variações desproporcionais nos resultados do ajuste do modelo. Essa metodologia tem sido largamente estudada em classes particulares de modelos e estendida para situações mais gerais, como o artigo de Venezuela et al. (2011), que apresenta métodos de influência local para modelos lineares generalizados para análise de medidas repetidas.

Pretendemos ainda estudar um critério para seleção da matriz de correlações de trabalho que leve em consideração a modelagem do parâmetro de dispersão, visto que o QIC assume que o parâmetro de dispersão seja homogêneo e precisamos fazer algumas considerações para utilizá-lo na modelagem conjunta dos parâmetros de locação e dispersão.

Capítulo 6

Apêndice

6.1 Apêndice A

Neste Apêndice apresentamos uma metodologia para gerar valores de variáveis aleatórias (v.a) Gumbel independentes e vetores aleatórios com marginais Gumbel.

6.1.1 Gerando valores de v.a. Gumbel

Devido à simplicidade e à facilidade de implementação do método da função inversa, a geração de números pseudo-aleatórios provenientes de distribuições de probabilidade, distintas da Uniforme(0,1), é usualmente efetuada através deste método.

Suponha que desejamos gerar valores da v.a. $Y \sim Gumbel(\theta, \phi)$ com função densidade de probabilidade, f_Y , e função de distribuição acumulada, F_Y , dadas, respectivamente, por (2.1) e (2.2). Gerando-se um valor u da v.a. $U \sim U(0, 1)$, o valor simulado y da v.a. Y é obtido isolando-se y na equação $F_Y(y) = u$, ou seja,

$$u = F_Y(y) = \exp \{ -\exp [-(y - \theta)/\phi] \} \implies y = \theta - \phi \log(-\log(u))$$

O algoritmo para gerar n valores da v.a $Y \sim Gumbel(\theta, \phi)$ é dado por:

1. Gere u da v.a. $U \sim U(0, 1)$;
2. Calcule $y = \theta - \phi \log(-\log(u))$;
3. Repita o processo até gerar os n valores desejados.

6.1.2 Gerando valores de vetores aleatórios Gumbel

Suponha que seja de interesse gerar valores do vetor aleatório $\mathbf{Y} = (Y_1, \dots, Y_m)'$ em que $Y_i \sim Gumbel(\theta_i, \phi_i)$, $i = 1, \dots, m$ e $Cov(\mathbf{Y}) = \Sigma$.

Dias e Godoi (1997) apresentaram um método para gerar valores de variáveis aleatórias correlacionadas de distribuições uniforme, triangular, exponencial, gama e binomial. Adaptamos este método para gerar valores de Gumbel multivariada.

O primeiro passo do algoritmo, consiste em gerar $\mathbf{z} = (z_1, \dots, z_m)'$ do vetor aleatório $\mathbf{Z} \sim N_m(\mathbf{0}, \Sigma)$. Valores simulados de \mathbf{Z} podem ser facilmente obtidos através de alguns pacotes do R como, por exemplo, o pacote `mnormt`.

No segundo passo, geramos valores u_1, \dots, u_m de variáveis aleatórias uniformes correlacionadas, através da equação $u_i = F_Z(z_i)$, $i = 1, \dots, m$. Computacionalmente podemos calcular a $F_Z(z)$ como

$$F_Z(z) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{z/\sqrt{2}} e^{-x^2} dx$$

Utilizando o método da inversa, os valores simulados de cada componente do vetor aleatório \mathbf{Y} são dados por $y_i = \theta_i - \phi_i \log[-\log(u_i)]$, para $i = 1, \dots, m$.

6.2 Apêndice B

Aqui calculamos algumas esperanças importantes no desenvolvimento da matriz de informação. Considere a v.a $Y_{it} \sim \text{Gumbel}(\theta_{it}, \phi_{it})$ e f.d.p. dada por (2.1). Logo,

$$\begin{aligned}
 E \left[\left(\frac{Y_{it}^*}{\mu_{it}^*} \right)^k \right] &= E \left[\exp \left(-k \frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \right] \\
 &= \int_{-\infty}^{\infty} \left\{ \exp \left(-k \frac{y_{it} - \theta_{it}}{\phi_{it}} \right) f_{Y_{it}}(y_{it} | \theta_{it}, \phi_{it}) \right\} dy_{it} \\
 &= \int_{-\infty}^{\infty} \left\{ \frac{1}{\phi_{it}} \exp \left(-(k+1) \frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \exp \left[-\exp \left(-\frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \right] \right\} dy_{it} \\
 &= \int_0^{\infty} u_{it}^k e^{-u_{it}} du_{it}, \quad u_{it} = \exp \left[-(y_{it} - \theta_{it}) / \phi_{it} \right] \\
 &= \Gamma(k+1),
 \end{aligned}$$

$$\begin{aligned}
 E \left[\left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \frac{Y_{it}^*}{\mu_{it}^*} \right] &= E \left[\left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \exp \left(-\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \right] \\
 &= \int_{-\infty}^{\infty} \left\{ \left(\frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \exp \left(-\frac{y_{it} - \theta_{it}}{\phi_{it}} \right) f_{Y_{it}}(y_{it} | \theta_{it}, \phi_{it}) \right\} dy_{it} \\
 &= \int_{-\infty}^{\infty} \left\{ \left(\frac{y_{it} - \theta_{it}}{\phi_{it}^2} \right) \exp \left(-2 \frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \exp \left[-\exp \left(-\frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \right] \right\} dy_{it} \\
 &= \gamma - 1,
 \end{aligned}$$

$$\begin{aligned}
 E \left[\frac{Y_{it}^*}{\mu_{it}^*} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \right] &= E \left[\exp \left(-\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \right] \\
 &= \int_{-\infty}^{\infty} \left\{ \exp \left(-\frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \left(\frac{y_{it} - \theta_{it}}{\phi_{it}} \right)^2 f_{y_{it}}(y_{it} | \theta_{it}, \phi_{it}) \right\} dy_{it} \\
 &= \int_{-\infty}^{\infty} \left\{ \frac{(y_{it} - \theta_{it})^2}{\phi_{it}^3} \exp \left(-2 \frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \exp \left[-\exp \left(-\frac{y_{it} - \theta_{it}}{\phi_{it}} \right) \right] \right\} dy_{it} \\
 &= \Gamma^{(2)}(2) \\
 &= \frac{\pi^2}{6} - \gamma(2 - \gamma),
 \end{aligned}$$

em que $\Gamma^{(2)}(\cdot)$ é a segunda derivada da função $\Gamma(\cdot)$ definida em (2.3) e γ é o número de Euler-Mascheroni, que vale aproximadamente 0,5772156649.

Como $E \left[\frac{Y_{it}^*}{\mu_{it}^*} \right] = 1$, $E \left[\left(\frac{Y_{it}^*}{\mu_{it}^*} \right)^2 \right] = 2$ e $E \left[\frac{Y_{it}^*}{\mu_{it}^*} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \right] = \gamma - 1$, então

$$E[\lambda_{it}] = E[Y_{it}^* - \mu_{it}^*] = \mu_{it}^* E \left[\frac{Y_{it}^*}{\mu_{it}^*} - 1 \right] = 0,$$

$$\begin{aligned} E[\delta_{it}] &= E \left\{ \mu_{it}^* + \frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right\} \\ &= \mu_{it}^* + E \left\{ \frac{Y_{it} - \theta_{it}}{\phi_{it}} Y_{it}^* \right\} - E \left\{ \frac{Y_{it} - \theta_{it}}{\phi_{it}} \mu_{it}^* \right\} \\ &= \mu_{it}^* + E \left\{ \frac{Y_{it} - \theta_{it}}{\phi_{it}} \frac{Y_{it}^* \mu_{it}^*}{\mu_{it}^*} \right\} - E(Y_{it} - \theta_{it}) \frac{\mu_{it}^*}{\phi_{it}} \\ &= \mu_{it}^* \left[1 + E \left\{ \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \frac{Y_{it}^*}{\mu_{it}^*} \right\} - \frac{(\theta_{it} + \gamma \phi_{it}) - \theta_{it}}{\phi_{it}} \right] \\ &= \mu_{it}^* [1 + (\gamma - 1) - \gamma] \\ &= 0, \end{aligned}$$

$$\begin{aligned} Var[\lambda_{it}] &= Var[Y_{it}^* - \mu_{it}^*] \\ &= Var[Y_{it}^*] \\ &= E[(Y_{it}^*)^2] - E[(Y_{it}^*)]^2 \\ &= E \left[\left(\frac{Y_{it}^*}{\mu_{it}^*} \right)^2 \right] (\mu_{it}^*)^2 - (\mu_{it}^*)^2 \\ &= 2(\mu_{it}^*)^2 - (\mu_{it}^*)^2 \\ &= (\mu_{it}^*)^2. \end{aligned}$$

Como $E \left\{ \frac{Y_{it}^*}{\mu_{it}^*} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \right\} = \frac{\pi^2}{6} - \gamma(2 - \gamma)$ e $E \left\{ \left[\frac{Y_{it}^*}{\mu_{it}^*} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \right]^2 \right\} = \frac{\pi^2}{3} + 2\gamma^2 - 6\gamma + 2$,

$$\begin{aligned} E \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right] &= E \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} \left(\frac{Y_{it}^*}{\mu_{it}^*} \right) \mu_{it}^* - \frac{Y_{it} - \theta_{it}}{\phi_{it}} \mu_{it}^* \right] \\ &= \mu_{it}^* \left\{ E \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} \left(\frac{Y_{it}^*}{\mu_{it}^*} \right) \right] - E \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right] \right\} \\ &= \mu_{it}^* \left[(\gamma - 1) - \frac{\theta_{it} + \gamma\phi_{it} - \theta_{it}}{\phi_{it}} \right] \\ &= -\mu_{it}^*, \end{aligned}$$

$$\begin{aligned} E \left\{ \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right]^2 \right\} &= E \left\{ \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \left[(Y_{it}^*)^2 - 2Y_{it}^* \mu_{it}^* + (\mu_{it}^*)^2 \right] \right\} \\ &= (\mu_{it}^*)^2 E \left\{ \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \left[\left(\frac{Y_{it}^*}{\mu_{it}^*} \right)^2 - 2 \frac{Y_{it}^*}{\mu_{it}^*} + 1 \right] \right\} \\ &= (\mu_{it}^*)^2 \left\{ \left(\frac{\pi^2}{3} + 2\gamma^2 - 6\gamma + 2 \right) \right\} \\ &+ (\mu_{it}^*)^2 \left\{ -2 \left(\frac{\pi^2}{6} - \gamma(2 - \gamma) \right) + \left(\frac{\pi^2}{6} + \gamma^2 \right) \right\} \\ &= (\mu_{it}^*)^2 \left\{ \frac{\pi^2}{6} + \gamma^2 - 2\gamma + 2 \right\}, \end{aligned}$$

$$\begin{aligned} Var[\delta_{it}] &= E[(\delta_{it})^2] - [E(\delta_{it})]^2 \\ &= E \left\{ \left[\mu_{it}^* + \frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right]^2 \right\} - 0 \\ &= E \left\{ (\mu_{it}^*)^2 + 2\mu_{it}^* \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right] \right\} \\ &+ E \left\{ \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right]^2 \right\} \\ &= (\mu_{it}^*)^2 + 2\mu_{it}^* [-\mu_{it}^*] + (\mu_{it}^*)^2 \left[\frac{\pi^2}{6} + \gamma^2 - 2\gamma + 2 \right] \\ &= (\mu_{it}^*)^2 \left[\frac{\pi^2}{6} + (\gamma - 1)^2 \right] \\ &= c(\mu_{it}^*)^2. \end{aligned}$$

Como $E \left[\frac{Y_{it}^*}{\mu_{it}^*} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right) \right] = \gamma - 1$, então

$$\begin{aligned}
E \left[\frac{\partial \lambda_{it}}{\partial \theta_{it}} \right] &= E \left\{ \frac{\partial}{\partial \theta_{it}} (Y_{it}^* - \mu_{it}^*) \right\} \\
&= E \left\{ \frac{\partial}{\partial \theta_{it}} \left[\exp \left(-\frac{Y_{it}}{\phi_{it}} \right) - \exp \left(-\frac{\theta_{it}}{\phi_{it}} \right) \right] \right\} \\
&= E \left\{ 0 - \frac{\mu_{it}^*}{\phi_{it}} \right\} \\
&= \frac{\mu_{it}^*}{\phi_{it}},
\end{aligned}$$

$$\begin{aligned}
E \left[\frac{\partial \delta_{it}}{\partial \theta_{it}} \right] &= E \left\{ \frac{\partial}{\partial \theta_{it}} \left[\mu_{it}^* + \frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right] \right\} \\
&= E \left\{ -\frac{\mu_{it}^*}{\phi_{it}} - \frac{(Y_{it}^* - \mu_{it}^*)}{\phi_{it}} + \frac{Y_{it} - \theta_{it}}{\phi_{it}} \frac{\mu_{it}^*}{\phi_{it}} \right\} \\
&= -\frac{\mu_{it}^*}{\phi_{it}} - 0 + \frac{[(\theta_{it} + \gamma \phi_{it}) - \theta_{it}] \mu_{it}^*}{\phi_{it} \phi_{it}} \\
&= (\gamma - 1) \frac{\mu_{it}^*}{\phi_{it}},
\end{aligned}$$

$$\begin{aligned}
E \left[\frac{\partial \lambda_{it}}{\partial \phi_{it}} \right] &= E \left[\frac{\partial (Y_{it}^* - \mu_{it}^*)}{\partial \phi_{it}} \right] \\
&= E \left(Y_{it}^* \frac{Y_{it}}{\phi_{it}^2} - \mu_{it}^* \frac{\theta_{it}}{\phi_{it}^2} \right) \\
&= E \left(Y_{it}^* \frac{Y_{it}}{\phi_{it}^2} - \mu_{it}^* \frac{\theta_{it}}{\phi_{it}^2} + Y_{it}^* \frac{\theta_{it}}{\phi_{it}^2} - Y_{it}^* \frac{\theta_{it}}{\phi_{it}^2} \right) \\
&= E \left([Y_{it}^* - \mu_{it}^*] \frac{\theta_{it}}{\phi_{it}^2} + [Y_{it} - \theta_{it}] \frac{Y_{it}^*}{\phi_{it}^2} \right) \\
&= E \left([Y_{it} - \theta_{it}] \frac{Y_{it}^* \mu_{it}^*}{\phi_{it}^2 \mu_{it}^*} \right) \\
&= \frac{\mu_{it}^*}{\phi_{it}} E \left(\frac{Y_{it}^*}{\mu_{it}^*} \left[\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right] \right) \\
&= (\gamma - 1) \frac{\mu_{it}^*}{\phi_{it}}.
\end{aligned}$$

Como $E \left[\exp \left(-\frac{Y_{it}-\theta_{it}}{\phi_{it}} \right) \left(\frac{Y_{it}-\theta_{it}}{\phi_{it}} \right)^2 \right] = \Gamma^{(2)}(2) = \frac{\pi^2}{6} - \gamma(2 - \gamma)$, então

$$\begin{aligned}
E \left[\frac{\partial \delta_{it}}{\partial \phi_{it}} \right] &= E \left\{ \frac{\partial}{\partial \phi_{it}} \left[\mu_{it}^* + \frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right] \right\} \\
&= E \left\{ \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} - \frac{Y_{it} - \theta_{it}}{\phi_{it}^2} (Y_{it}^* - \mu_{it}^*) + \frac{Y_{it} - \theta_{it}}{\phi_{it}} \left(\frac{Y_{it}^* Y_{it}}{\phi_{it}^2} - \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} \right) \right\} \\
&= E \left\{ \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} - \frac{Y_{it} - \theta_{it}}{\phi_{it}^2} (Y_{it}^* - \mu_{it}^*) \right\} \\
&\quad + E \left\{ \frac{Y_{it} - \theta_{it}}{\phi_{it}} \left(\frac{Y_{it}^* Y_{it}}{\phi_{it}^2} + \frac{Y_{it}^* \theta_{it}}{\phi_{it}^2} - \frac{Y_{it}^* \theta_{it}}{\phi_{it}^2} - \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} \right) \right\} \\
&= E \left\{ \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} - \frac{Y_{it} - \theta_{it}}{\phi_{it}^2} (Y_{it}^* - \mu_{it}^*) + \frac{Y_{it}^*}{\phi_{it}} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \right\} \\
&\quad + E \left\{ \frac{Y_{it} - \theta_{it}}{\phi_{it}} \left(\frac{Y_{it}^* \theta_{it}}{\phi_{it}^2} - \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} \right) \right\} \\
&= E \left\{ \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} + \frac{Y_{it} - \theta_{it}}{\phi_{it}^2} (Y_{it}^* - \mu_{it}^*) \left(\frac{\theta_{it}}{\phi_{it}} - 1 \right) + \frac{\mu_{it}^* Y_{it}^*}{\mu_{it}^* \phi_{it}} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \right\} \\
&= \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} + \left(\frac{\theta_{it}}{\phi_{it}^2} - \frac{1}{\phi_{it}} \right) E \left\{ \frac{Y_{it} - \theta_{it}}{\phi_{it}} (Y_{it}^* - \mu_{it}^*) \right\} + E \left\{ \frac{\mu_{it}^* Y_{it}^*}{\phi_{it} \mu_{it}^*} \left(\frac{Y_{it} - \theta_{it}}{\phi_{it}} \right)^2 \right\} \\
&= \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} + \left(\frac{\theta_{it}}{\phi_{it}^2} - \frac{1}{\phi_{it}} \right) \left\{ (\gamma - 1) \mu_{it}^* - \frac{[(\theta_{it} + \gamma \phi_{it}) - \theta_{it}] \mu_{it}^*}{\phi_{it}} \right\} + \frac{\mu_{it}^*}{\phi_{it}} \Gamma^{(2)}(2) \\
&= \frac{\mu_{it}^* \theta_{it}}{\phi_{it}^2} - \left(\frac{\theta_{it}}{\phi_{it}^2} - \frac{1}{\phi_{it}} \right) \mu_{it}^* + \frac{\mu_{it}^*}{\phi_{it}} \Gamma^{(2)}(2) \\
&= \left(1 + \Gamma^{(2)}(2) \right) \frac{\mu_{it}^*}{\phi_{it}} \\
&= \left(1 + \frac{\pi^2}{6} - \gamma(2 - \gamma) \right) \frac{\mu_{it}^*}{\phi_{it}} \\
&= \left(\frac{\pi^2}{6} + (\gamma - 1)^2 \right) \frac{\mu_{it}^*}{\phi_{it}} \\
&= c \frac{\mu_{it}^*}{\phi_{it}}.
\end{aligned}$$

6.3 Apêndice D

Nesse apêndice calculamos as matrizes de sensibilidade e variabilidade, considerando os vetores de parâmetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ e $\boldsymbol{\xi} = (\xi_1, \dots, \xi_q)'$, as funções de estimação $\boldsymbol{\Psi}_\tau = (\boldsymbol{\Psi}_\beta, \boldsymbol{\Psi}_\xi)$ definida em (3.11) e $E(\lambda_{it}) = E(Y_{it}^* - \mu_{it}^*) = 0$. Então, para $t = 1, \dots, m_i$; $J = 1, \dots, p$

$$\begin{aligned}
S_{\beta\beta} &= E\left(\frac{\partial \boldsymbol{\Psi}_\beta}{\partial \boldsymbol{\beta}}\right) \\
&= E\left(\frac{\partial}{\partial \boldsymbol{\beta}} \left[\sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \boldsymbol{\lambda}_i \right\} \right]\right) \\
&= \sum_{i=1}^n E\left\{ \frac{\partial}{\partial \boldsymbol{\beta}} \left[\tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \boldsymbol{\lambda}_i \right] \right\} \\
&= \sum_{i=1}^n E\left\{ \frac{\partial}{\partial \boldsymbol{\beta}} \left[\tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \right] \boldsymbol{\lambda}_i + \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \frac{\partial \boldsymbol{\lambda}_i}{\partial \boldsymbol{\beta}} \right\} \\
&= \sum_{i=1}^n \left\{ \frac{\partial}{\partial \boldsymbol{\beta}} \left[\tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \right] E(\boldsymbol{\lambda}_i) - \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} E\left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \boldsymbol{\beta}}\right) \right\} \\
&= \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \left[\frac{\partial \lambda_{it}}{\partial \theta_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \eta_{1it}}{\partial \beta_J} \right]_{t,J} \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \text{diag} \left[-\frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \right] \tilde{\mathbf{X}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \mathbf{A}_i \tilde{\mathbf{X}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{W}_{(\beta\beta)_i} \tilde{\mathbf{X}}_i \right\},
\end{aligned}$$

em que $\mathbf{W}_{(\beta\beta)_i} = \mathbf{A}_i \boldsymbol{\Omega}_i^{-1} \mathbf{A}_i = \text{diag} \left[\left(\frac{1}{\phi_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \right)^2 \right] \mathbf{R}^{-1}$.

Como $E \left[\frac{\partial \lambda_{it}}{\partial \phi_{it}} \right] = (\gamma - 1) \frac{\mu_{it}^*}{\phi_{it}}$, então para $t = 1, \dots, m_i$, $L = 1, \dots, q$,

$$\begin{aligned}
S_{\beta\xi} &= E \left(\frac{\partial \Psi_\beta}{\partial \xi} \right) \\
&= E \left(\frac{\partial}{\partial \xi} \left[\sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \Omega_i^{-1} \lambda_i \right\} \right] \right) \\
&= \sum_{i=1}^n E \left\{ \frac{\partial}{\partial \xi} \left[\tilde{\mathbf{X}}_i' \mathbf{A}_i \Omega_i^{-1} \lambda_i \right] \right\} \\
&= \sum_{i=1}^n \left\{ \frac{\partial}{\partial \xi} \left[\tilde{\mathbf{X}}_i' \mathbf{A}_i \Omega_i^{-1} \right] E(\lambda_i) + \tilde{\mathbf{X}}_i' \mathbf{A}_i \Omega_i^{-1} E \left[\frac{\partial \lambda_i}{\partial \xi} \right] \right\} \\
&= \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \Omega_i^{-1} E \left(\left[\begin{array}{ccc} \frac{\partial \lambda_{it}}{\partial \phi_{it}} & \frac{\partial \phi_{it}}{\partial \eta_{2it}} & \frac{\partial \eta_{2it}}{\partial \xi_L} \end{array} \right]_{i,t,L} \right) \right\} \\
&= \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \Omega_i^{-1} \text{diag} \left[(\gamma - 1) \frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right]_{m_i} \tilde{\mathbf{Z}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{A}_i \Omega_i^{-1} (\gamma - 1) \text{diag} \left[-\frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right]_{m_i} \tilde{\mathbf{Z}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' (\gamma - 1) \mathbf{A}_i \Omega_i^{-1} \mathbf{D}_i \tilde{\mathbf{Z}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{X}}_i' \mathbf{W}_{(\beta\xi)_i} \tilde{\mathbf{Z}}_i \right\},
\end{aligned}$$

em que $\mathbf{W}_{(\beta\xi)_i} = (\gamma - 1) \mathbf{A}_i \Omega_i^{-1} \mathbf{D}_i = \text{diag} \left[\frac{(\gamma - 1)}{\phi_{it}^2} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right]_{m_i} \mathbf{R}^{-1}$.

Como $E \left[\frac{\partial \delta_{it}}{\partial \theta_{it}} \right] = (\gamma - 1) \frac{\mu_{it}^*}{\phi_{it}}$, então, para $t = 1, \dots, m_i$; $J = 1, \dots, p$,

$$\begin{aligned}
S_{\xi\beta} &= E \left(\frac{\partial \Psi_{\xi}}{\partial \beta} \right) \\
&= E \left(\frac{\partial}{\partial \beta} \left[\sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \delta_i \right\} \right] \right) \\
&= \sum_{i=1}^n E \left\{ \frac{\partial}{\partial \beta} \left[\tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \delta_i \right] \right\} \\
&= \sum_{i=1}^n \left\{ \frac{\partial}{\partial \beta} \left[\tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \right] E(\delta_i) + \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} E \left[\frac{\partial \delta_i}{\partial \beta} \right] \right\} \\
&= \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} E \left(\left[\frac{\partial \delta_{it}}{\partial \theta_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \eta_{1it}}{\partial \beta_J} \right]_{t,J} \right) \right\} \\
&= \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \text{diag} \left[(\gamma - 1) \frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \right]_{m_i} \tilde{\mathbf{X}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} (\gamma - 1) \text{diag} \left[-\frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \right]_{m_i} \tilde{\mathbf{X}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' (\gamma - 1) \mathbf{D}_i \Omega_i^{-1} \mathbf{A}_i \tilde{\mathbf{X}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{W}_{(\beta\xi)_i} \tilde{\mathbf{X}}_i \right\},
\end{aligned}$$

em que $\mathbf{W}_{(\beta\xi)_i} = (\gamma - 1) \mathbf{D}_i \Omega_i^{-1} \mathbf{A}_i = \text{diag} \left[\frac{(\gamma - 1)}{\phi_{it}^2} \frac{\partial \theta_{it}}{\partial \eta_{1it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right] \mathbf{R}^{-1}$.

Como $E \left[\frac{\partial \delta_{it}}{\partial \phi_{it}} \right] = c \frac{\mu_{it}^*}{\phi_{it}}$, então, para $t = 1, \dots, m_i$; $L = 1, \dots, q$,

$$\begin{aligned}
S_{\xi\xi} &= E \left(\frac{\partial \Psi_{\xi}}{\partial \xi} \right) \\
&= E \left(\frac{\partial}{\partial \xi} \left[\sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \boldsymbol{\delta}_i \right\} \right] \right) \\
&= \sum_{i=1}^n E \left\{ \frac{\partial}{\partial \xi} \left[\tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \boldsymbol{\delta}_i \right] \right\} \\
&= \sum_{i=1}^n \left\{ \frac{\partial}{\partial \xi} \left[\tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \right] E(\boldsymbol{\delta}_i) + \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} E \left[\frac{\partial \boldsymbol{\delta}_i}{\partial \xi} \right] \right\} \\
&= \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} E \left(\left[\frac{\partial \delta_{it}}{\partial \phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \frac{\partial \eta_{2it}}{\partial \xi_L} \right]_{t,L} \right) \right\} \\
&= \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{D}_i \Omega_i^{-1} \text{diag} \left[c \frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right]_{m_i} \tilde{\mathbf{Z}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' c \mathbf{D}_i \Omega_i^{-1} \text{diag} \left[-\frac{\mu_{it}^*}{\phi_{it}} \frac{\partial \phi_{it}}{\partial \eta_{2it}} \right]_{m_i} \tilde{\mathbf{Z}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' c \mathbf{D}_i \Omega_i^{-1} \mathbf{D}_i \tilde{\mathbf{Z}}_i \right\} \\
&= - \sum_{i=1}^n \left\{ \tilde{\mathbf{Z}}_i' \mathbf{W}_{(\xi\xi)_i} \tilde{\mathbf{Z}}_i \right\},
\end{aligned}$$

em que $\mathbf{W}_{(\xi\xi)_i} = c \mathbf{D}_i \Omega_i^{-1} \mathbf{D}_i = \text{diag} \left[\frac{c}{\phi_{it}^2} \left(\frac{\partial \phi_{it}}{\partial \eta_{2it}} \right)^2 \right] \mathbf{R}^{-1}$.

Referências Bibliográficas

- [1] ARTES, R.; JØRGENSEN, B., *Longitudinal data estimating equations for dispersion models*, SCANDINAVIAN JOURNAL OF STATISTICS, 27, 321-334, 2000.
- [2] ARTES, R.; BOTTER, D. A., *Função de estimação em modelos de regressão*, 9^A ESCOLA DE MODELOS DE REGRESSÃO - ABE, SÃO PEDRO, 2005.
- [3] ARTES, R., *Modelo de regressão para distribuições inflacionadas de zeros: correção dos erros-padrão para dados em painel*, 11^A ESCOLA DE MODELOS DE REGRESSÃO - ABE, RECIFE, 2009.
- [4] BARRETO-SOUZA, W.; VASCONCELLOS, K. L. P., *Bias and skewness in a general extreme-value regression model*, COMPUTATIONAL STATISTICS AND DATA ANALYSIS 55, 1379-1393, 2011.
- [5] CLARKE, R. T., *Estimating time trends in Gumbel-distributed data by means of generalized linear models*, WATER RESOURCES RESEARCH, 38, 1111-1121, 2002.
- [6] COOK, R. D., *Assessment of local influence (with discussion)*, JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B 48, 133-169, 1986.

- [7] CORDEIRO, G. M., *Introdução à teoria assintóticas*, 22º COLÓQUIO BRASILEIRO DE MATEMÁTICA - IMPA, RIO DE JANEIRO, 1999.
- [8] CORDEIRO, G. M.; DEMÉTRIO, C. G. B., *Modelos lineares generalizados e Extensões*, DEPARTAMENTO DE CIÊNCIAS EXATAS, ESALQ, USP, 2007.
- [9] DIAS, C. T. S.; GODOI, C. R. M., *Geração de amostras aleatórias multidimensionais com estrutura de dependência para algumas distribuições de probabilidade*, 5ª ESCOLA DE MODELOS DE REGRESSÃO - ABE, CAMPOS DE JORDÃO, 1997.
- [10] LEITE, J. G.; SINGER, J. M., *Métodos assintóticos em estatística*, 9º SINAPE - ABE, SÃO PAULO, 1990.
- [11] GODAMBE, V. P., *An optimum property of regular maximum likelihood estimation*, ANNALS OF MATHEMATICAL STATISTICS, 31, 1208-1211, 1960.
- [12] GUMBEL, E. J., *Statistics of Extremes*, NEW YORK: COLUMBIA UNIVERSITY PRESS, 1958.
- [13] LIANG, K. Y.; ZEGER, S. L., *Longitudinal analysis using generalized linear models*, BIOMETRIKA 73, 13-22, 1986.
- [14] KOTZ, S.; NADAJARAH, S., *Extreme Value Distributions: Theory and Applications*, LONDON: IMPERIAL COLLEGE PRESS, 2000.
- [15] McCULLAGH, P.; NELDER, J. A., *Generalized linear models*, CHAPMAN AND HALL, 2ND ED., 1989.
- [16] NELDER, J. A., WEDDERBURN, R. W. M., *Generalized Linear Models*, JOURNAL OF THE ROYAL STATISTICAL SOCIETY A, 135, 370-384, 1972.

- [17] PAN, W., *Akaike's information criterion in generalized estimating equations*, BIOMETRIKA 57, 120-125, 2001.
- [18] PAULA, G. A., *Modelos de regressão com apoio computacional*, NOTAS DE AULAS, DEPARTAMENTO DE ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, URL [HTTP://WWW.IME.USP.BR/ GIAPAU/MLGS.HTML](http://www.ime.usp.br/giapaula/mlgs.html), 2012.
- [19] PREGIBON, D., *Logistic regression diagnostics*, ANNALS OF STATISTICS, 9, 705-724, 1981.
- [20] R DEVELOPMENT CORE TEAM., *R: A Language and Environment for Statistical Computing*, R FOUNDATION FOR STATISTICAL COMPUTING, VIENNA, AUSTRIA, ISBN 3-900051-07-0, URL [HTTP://WWW.R-PROJECT.ORG/](http://www.R-project.org/), 2012.
- [21] RAMOS, E. M.; CORDEIRO, G. M., *Vieses e correções de Bartlett e tipo-Bartlett em modelos de regressão de valores extremos*, REVISTA DE MATEMÁTICA E ESTATÍSTICA, 24, 77-93, 2006.
- [22] RODRIGUES, J. T. G., *Análise de dados longitudinais para variáveis binárias*, DISSERTAÇÃO DE MESTRADO, DE-UFSCAR, SÃO CARLOS, 89P., 2009.
- [23] VENEZUELA, M. K., *Equação de estimação generalizada e influência local para modelos de regressão beta com medidas repetidas*, TESE DE DOUTORADO, IME-USP, SÃO PAULO, 153P., 2008.
- [24] VENEZUELA, M. K., SANDOVAL, M.C., BOTTER, D.A., *Local influence in estimating equations*, COMPUTATIONAL STATISTICS AND DATA ANALYSIS 55, 1867-1883, 2011.

- [25] VIOLA, M. L. L., *Teoria de valores extremos e cópulas: distribuição valor extremo generalizada e cópulas arquimedianas generalizadas trivariadas*, DISSERTAÇÃO DE MESTRADO, IMECC-UNICAMP, CAMPINAS, 2007.