UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PÓS-GRADUAÇÃO EM ESTATÍSTICA

VEROSSIMILHANÇA PERFILADA NOS MODELOS LINEARES GENERALIZADOS COM SUPERDISPERSÃO

THIAGO ALEXANDRO NASCIMENTO DE ANDRADE

Orientadora: Profª. Dra. Audrey Helen Mariz de Aquino Cysneiros

Área de concentração: Estatística Matemática

Dissertação submetida como requerimento parcial para obtenção do grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, fevereiro de 2013

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da Silva, CRB4-1217

Andrade, Thiago Alexandro Nascimento de
Verossimilhança perfilada nos modelos lineares generalizados com superdispersão / Thiago Alexandro Nascimento de Andrade. - Recife: O Autor, 2013.
x, 49 folhas: il., tab.

Orientadora: Audrey Helen Mariz de Aquino Cysneiros.
Dissertação (Mestrado) - Universidade Federal de Pernambuco. CCEN. Estatística, 2013.

Inclui bibliografia e apêndice.

1. Estatística matemática. 2. Teoria assintótica. I. Cysneiros, Audrey Helen Mariz de Aquino (orientadora). II. Título.

519.5 CDD (23. ed.) MEI2013 – 036

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

Recife, 08 de fevereiro de 2013.

Nós recomendamos que a dissertação de mestrado de autoria de

Thiago Alexandro Nascimento de Andrade

Intitulada

"Verossimilhança Perfilada nos Modelos Lineares Generalizados com Superdispersão"

Seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.

Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

Audrey Helen Mariz de Aquino Cysneiros	Orientadora / UFPE
Francisco José de Azevedo Cysneiros	UFPE
Miguel Angel Uribe Opazo	UNIOESTE

Este documento será anexado à versão final da tese.

A meus amados pais
e minha esposa,
dedico com todo meu amor.

Agradecimentos

Primeiramente a Deus, pela minha vida, por iluminar meus caminhos e por me dar forças necessárias para a conclusão de mais esta etapa de minha vida.

À minha esposa Renata, por todo seu amor, companheirismo, cumplicidade e paciência para com minhas ausências.

Aos meus familiares, em especial a minha mãe Célia, meu pai Valdeci, minha irmã Thaysa, meu Tio Mário, minha sogra Maria do Carmo pelo apoio incondicional em todos os momentos.

Um agradecimento especial a minha avó Quitéria, por todas as orações e suas sábias palavras de conforto que sempre tiveram o dom de me acalmar nos momentos mais difíceis.

À professora Audrey Helen Mariz de Aquino Cysneiros, pela orientação, paciência, amizade e ensinamentos, fundamentais para conclusão deste trabalho.

Aos colegas do mestrado, em especial à Ana, Diêgo, Heloisa e Sadraque por todo companheirismo e pelas muitas horas de estudo compartilhadas.

À Valéria Bittencourt, pela competência, dedicação e carinho que tornaramse marcas registradas de seu trabalho com os alunos da pós-graduação.

A todos os professores do Programa de Mestrado em Estatística da UFPE.

Ao CNPq pelo apoio financeiro.

"A cada dia que vivo, mais me convenço de que o desperdício da vida está no amor que não damos, nas forças que não usamos, na prudência egoísta que nada arrisca e que, esquivando-nos do sofrimento, perdemos também a felicidade"

Carlos Drummond de Andrade

Resumo

A classe de Modelos Lineares Generalizados com Superdispersão (MLGSs) proposta por Dey *et al.* (1997), tem sido amplamente utilizada na modelagem de dados cuja variância da variável resposta excede o valor nominal predito no modelo. O principal objetivo da presente dissertação é a obtenção de um fator de correção de Bartlett, segundo a metodologia proposta por DiCiccio e Stern (1994), à estatística da razão de verossimilhanças perfiladas ajustadas proposta por Cox e Reid (1987) para o teste conjunto dos efeitos da dispersão nesta classe de modelos. Estudos de simulação de Monte Carlo foram realizados com o objetivo de avaliar os desempenhos dos testes baseados nas estatísticas da razão de verossimilhanças usual (LR), razão de verossimilhanças perfiladas ajustadas (LR_{pa}) e razão de verossimilhanças perfiladas ajustadas corrigida (LR^c_{pa}), no que se refere a tamanho e poder em amostras finitas. Os resultados numéricos obtidos favorecem o teste proposto nesta dissertação.

Palavras chave: Correção de Bartlett; Modelos Lineares Generalizados com Superdispersão; Razão de verossimilhanças; Verossimilhança perfilada.

Abstract

The class of overdispersed generalized linear models (OGLMs) developed by Dey et al. (1997) has been widely used in data modeling when the variance of the response variable exceeds the nominal variance predicted by the model. The main goal of this dissertation is to obtain a Bartlett correction (DiCiccio and Stern, 1994) for a test of dispersion effects based on the adjusted profiled likelihood ratio statistic proposed by Cox and Reid (1987) in the OGLMs. Monte Carlo simulation was used to compare the performance of the tests based on the likelihood ratio statistic, adjusted profiled likelihood ratio statistic and corrected adjusted profiled likelihood ratio statistic, using different scenarios of size and power in the finite-sample. The numerical results obtained outperforms the test proposed in this dissertation.

Keywords: Bartlett Correction; Likelihood ratio; Profile likelihood; Overdispersed generalized linear model.

Lista de Tabelas

Sumário

CAPÍTULO 1

Introdução

Nelder e Wedderburn (1972) estenderam a metodologia dos modelos clássicos de regressão, definindo uma ampla classe de modelos, qual denominaram modelos lineares generalizados (MLGs). Estes modelos caracterizam-se pelo fato de a variável resposta poder ter qualquer distribuição definida na família exponencial de distribuições, tais como as distribuições Gama, Normal, Normal Inversa, Binomial e Poisson, dentre outras. Além disso, nos MLGs, a função de ligação que relaciona a média da resposta ao preditor linear, não está restrita à função identidade e pode ser qualquer função monótona derivável.

Apesar de bastante flexíveis, os MLGs podem não fornecer ajustes adequados em situações onde há grande variabilidade dos dados. Buscando melhorar tais ajustes, Dey et al. (1997) introduziram um modelo de regressão adicional para o parâmetro de dispersão, além de incorporá-lo na função de variância, definindo assim a classe de Modelos Lineares Generalizados com Superdispersão (MLGSs).

Devido a grande aplicabilidade dos MLGs e MLGSs na modelagem de dados em diversas áreas do conhecimento como Economia, Demografia, Medicina, Engenharia de produção, entre outras, muitos pesquisadores têm concentrado esforços no sentido de melhorar a qualidade de inferências realizadas com base nesses modelos. Em termos práticos, testes mais poderosos, estimadores mais precisos e intervalos de confiança mais estreitos, podem traduzir-se em diversos benefícios nas atividades produtivas, motivando assim estudos nesta área. Em particular, no que se refere à testes de hipóteses, correções de Bartlett em MLGs e MLGSs vêm recebendo considerável atenção na literatura.

A seguir, apresentamos alguns trabalhos que contemplam as classes dos MLGs e MLGSs. Cordeiro e McCullagh (1991) obtiveram fórmulas gerais para o viés de ordem $O(n^{-1})$ dos estimadores de máxima verossimilhança em MLGs. Ferrari *et al.* (2005) tratam das correções de Bartlett em MLGs. Cordeiro *et al.* (2006) obtiveram um fator de correção de Bartlett para a estatística da razão de verossimilhanças em MLGSs. Cordeiro e Barroso (2007), obtiveram expressões matriciais para o viés de ordem $O(n^{-2})$ dos estimadores de máxima verossimilhança em MLGs. Estudos detalhados sobre a classe dos MLGs podem ser encontrados em Cordeiro (1986), Cordeiro e Demétrio (2010) e Paula (2012).

O presente trabalho tem como objetivo principal a obtenção de um fator de correção de Bartlett, segundo a metodologia proposta por DiCiccio e Stern (1994), para a estatística da razão de verossimilhanças perfiladas ajustadas proposta por Cox e Reid (1987) na classe dos MLGSs.

1.1 Organização da Dissertação

A presente dissertação é composta por cinco Capítulos incluindo esta introdução. No segundo Capítulo, a classe de modelos lineares generalizados com superdispersão é apresentada e são discutidos os principais aspectos inferenciais para esta classe de modelos. O terceiro Capítulo trata das correções de Bartlett. São discutidas metodologias para obtenção de fatores de correção de Bartlett propostas por Lawley (1956) e DiCiccio e Stern (1994) para a estatística razão de verossimilhanças usual e estatística da razão de verossimilhanças perfiladas ajustadas, respectivamente. Nesse capítulo encontra-se a principal contribuição teórica do presente trabalho: Uma expressão em forma matricial do fator de correção de Bartlett, segundo a metodologia proposta por DiCiccio e Stern (1994), à estatística da razão de verossimilhanças perfiladas ajustadas proposta por Cox e Reid (1987) para o teste conjunto dos efeitos da dispersão na classe dos MLGSs. No quarto Capítulo são apresentados resultados de simulação, que foram realizados com o propósito de comparar o desempenho de testes de hipóteses baseados nas estatísticas da razão de verossimilhanças usual, estatística da razão de verossimilhanças perfiladas ajustadas e estatística da razão de verossimilhanças perfiladas ajustadas corrigida. O quinto e último Capítulo é reservado às considerações finais.

1.2 Suporte Computacional

A presente dissertação foi integralmente digitada utilizando o sistema de tipografia LaTeX, desenvolvido por Leslie Lamport na década de 80. Foi utilizado o software MikTeX, que é uma implementação do LaTeX para a utilização no Windows.

O estudo de simulação foi realizado utilizando a linguagem matricial de programação `Ox`, desenvolvida por Jurgen Doornik em 1994. Foi utilizado o editor de textos *OxEdit* em sua versão 6.2 para Windows. As maximizações do logaritmo da função de verossimilhança foram realizadas a partir dos algoritmos BFGS e SQP (Nocedal e Wright, 1999, Capítulos 8 e 18), que estão disponíveis em rotinas definidas na linguagem `Ox`. Para maiores detalhes sobre esta linguagem de programação, ver Doornik (2006).

CAPÍTULO 2

Modelos Lineares Generalizados com Superdispersão

2.1 Introdução

Em regressão, o termo superdispersão é usado para designar situações em que a variância da variável resposta excede o valor nominal predito no modelo. Este comportamento é principalmente identificado em modelos com resposta Binomial e Poisson.

Uma vez detectada a superdispersão em um conjunto de dados, o uso de técnicas convencionais de modelagem pode dar origem a modelos imprecisos e, consequentemente, erros de previsão associados à imprecisão de tais modelos.

Ao longo dos anos, diversos trabalhos que contemplam metodologias específicas para dados em que se verifica a superdispersão vêm sendo propostos na literatura acadêmica especializada. Destacam-se os seguintes trabalhos: Efron (1986); Lindsay (1986); Smyth (1989) e Dey *et al.* (1997).

Dey *et al.* (1997) definem a classe de Modelos Lineares Generalizados com Superdispersão (MLGSs). Esta classe de modelos segue a mesma estrutura dos Modelos Lineares Generalizados (MLGs), embora apresentando um modelo de regressão adicional para o parâmetro de dispersão, além de incorporá-lo na função de variância. Assim sendo, os MLGSs são definidos a partir de um componente aleatório e duas componentes sistemáticas (Previdelli, 2005).

Nas demais seções que compõem este Capítulo, a classe dos MLGSs é apresentada. Para uma abordagem mais completa, recomenda-se a leitura dos seguintes trabalhos: Dey *et al.* (1997) apresentam a classe dos MLGSs, Dey e Ravishanker (1998) abordaram aspectos da inferência Bayesiana nos MLGSs, Cordeiro e Botter (2001) obtiveram os vieses de segunda ordem para os estimadores de máxima verossimilhança nos MLGSs. Previdelli (2005) obteve estimadores corrigidos para modelos não lineares generalizados com superdispersão. Cordeiro *et al.* (2006) deduziram um fator de correção de Bartlett em forma matricial para a estatística da razão de verossimilhanças em MLGSs.

2.2 Definição do Modelo

Sejam $Y_1, \ldots, Y_n$, variáveis aleatórias independentes, em que cada Y_i tem função densidade (ou de probabilidade) definida na família exponencial biparamétrica de distribuições dada por

$$\pi(y; \mu, \phi) = \mathrm{A}(y)\exp[(y-\mu)\psi^{(1,0)}(\mu, \phi) + \phi \mathrm{T}(y) + \psi(\mu, \phi)], \tag{2.1}$$

em que $\mathrm{A}(\cdot)$, $\mathrm{T}(\cdot)$ e $\psi(\cdot, \cdot)$ são funções conhecidas, μ e ϕ são os parâmetros de média e dispersão, respectivamente. As derivadas da função $\psi(\mu, \phi)$ com relação a μ e ϕ são denotadas por $\psi^{(r,s)} = \partial\psi^{r+s}(\mu, \phi)/\partial\mu^r\partial\phi^s$ para $(r, s \geq 0)$. Deste modo, $\psi^{(1,0)}$ denota a primeira derivada da função $\psi(\mu, \phi)$ com relação a μ, $\psi^{(0,1)}$ denota a primeira derivada da função $\psi(\mu, \phi)$ com relação a ϕ, e assim por diante.

A variável aleatória Y_i tem suporte nos reais e sua média e variância são dadas, respectivamente, por $\mathrm{E}(Y) = \mu$ e $\mathrm{Var}(Y) = \psi^{(2,0)^{-1}}$. Além disso, tem-se que $\mathrm{E}[\mathrm{T}(Y)] = -\psi^{(0,1)}$, $\mathrm{Var}[\mathrm{T}(Y)] = \psi^{(1,1)^2}\psi^{(2,0)^{-1}} - \psi^{(0,2)}$, e ainda $\mathrm{E}[(Y-\mu)\mathrm{T}(Y)] = -\psi^{(1,1)}\psi^{(2,0)^{-1}}$.

Além do componente aleatório (2.1), os MLGSs são definidos pelos modelos de regressão para média e dispersão, dados, respectivamente, por

$$\mathrm{f}(\mu) = \eta = \boldsymbol{X}\boldsymbol{\beta} \tag{2.2}$$

e,

$$\mathrm{g}(\phi) = \tau = \boldsymbol{S}\boldsymbol{\gamma}, \tag{2.3}$$

em que $\mathbf{X}$ e $\mathbf{S}$ são matrizes de covariadas, de dimensão $n \times p$ e $n \times q$, com posto completo p e q, respectivamente. Tem-se ainda que $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ são vetores de parâmetros desconhecidos a serem estimados, que representam, respectivamente, os efeitos das covariadas na média da resposta e no parâmetro de dispersão. As funções $f(\cdot)$ e $g(\cdot)$ são denotadas, respectivamente, por função de ligação da média e função de ligação da dispersão. Essas funções devem ser conhecidas, monótonas, contínuas e pelo menos duas vezes diferenciáveis.

2.3 Estimação dos Parâmetros

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ um vetor de $p+q$ parâmetros desconhecidos do modelo definido em (2.1)-(2.3). Seja $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ um vetor de observações da variável aleatória $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ também associada à (2.1)-(2.3). Então, o logaritmo da função de verossimilhança associado ao vetor $\boldsymbol{\theta}$ dado o vetor de observações $\boldsymbol{y}$ pode ser expresso na forma

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{l=1}^{n} [(y_l - \mu_l)\psi^{(1,0)}(\mu_l, \phi_l) + \phi_l \mathrm{T}(y_l) + \psi(\mu_l, \phi_l)] + \sum_{l=1}^{n} \log \mathrm{A}(y_l). \tag{2.4}$$

Assume-se que (2.4) é regular e derivável até quarta ordem com relação aos parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ (Cox e Hinkley, 1974).

Para definir algumas quantidades de interesse, será introduzida uma notação baseada na proposta por Cordeiro *et al.* (2006). Os índices r, s, t, u, ..., etc., referem-se ao vetor de parâmetros $\boldsymbol{\beta}$, enquanto que R, S, T, U, ..., etc., são indexadores das componentes do vetor $\boldsymbol{\gamma}$ de parâmetros. As derivadas do logaritmo da função de verossimilhança são representadas por $U_r = \partial\ell(\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\beta_r$, $U_{rS} = \partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\beta_r\partial\gamma_S$, etc. As derivadas do inverso das funções de ligação da média $\mu = f^{-1}(\eta)$ e da dispersão $\phi = g^{-1}(\tau)$ são dadas, respectivamente, por $m_{il} = \partial^i\mu_l/\partial\eta_l^i$, $\phi_{il} = \partial^i\phi_l/\partial\tau_l^i$, com $i = 1,2,3$ e $l = 1,\ldots,n$. Define-se ainda $(S)_l = \partial\tau_l/\partial\gamma_S$, $(R)_l = \partial\tau_l/\partial\gamma_R$, $(S,R)_l = (\partial\tau_l/\partial\gamma_s)(\partial\tau_l/\partial\gamma_R)$, etc., e $(s)_l = \partial\eta_l/\partial\beta_s$, $(r)_l = \partial\eta_l/\partial\beta_r$, $(s,r)_l = (\partial\eta_l/\partial\beta_s)(\partial\eta_l/\partial\beta_r)$, etc., com $l = 1,\ldots,n$. Por fim, definem-se as matrizes diagonais de ordem n dadas por $\Psi^{(2,0)} = \mathrm{diag}\{\psi_1^{(2,0)},\ldots,\psi_n^{(2,0)}\}$, $\Psi^{(0,2)} = \mathrm{diag}\{\psi_1^{(0,2)},\ldots,\psi_n^{(0,2)}\}$, $M_r = \mathrm{diag}\{m_{r1},\ldots,m_{rn}\}$ e $\Phi_r = \mathrm{diag}\{\phi_{r1},\ldots,\phi_{rn}\}$, para $r = 1,2,3$.

As primeiras derivadas do logaritmo da função de verossimilhança com relação ao r-ésimo componente de $\boldsymbol{\beta}$ e R-ésimo componente de $\boldsymbol{\gamma}$ são dadas por

$$\mathrm{U}_r = \frac{\partial\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\beta_r} = \sum_{l=1}^{n}[\psi_l^{(2,0)}(\mu_l,\phi_l)m_{1l}(r)_l(y_l-\mu_l)]$$

e

$$U_R = \frac{\partial\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\gamma_R} = \sum_{l=1}^{n}[\psi_l^{(1,1)}(\mu_l,\phi_l)(y_l-\mu_l) + \mathrm{T}(y_l) + \psi_l^{(0,1)}(\mu_l,\phi_l)]\phi_{1l}(R)_l,$$

com $r = 1,\ldots,p$ e $R = 1,\ldots,q$, respectivamente.

A função escore total para o vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top,\boldsymbol{\gamma}^\top)^\top$ de parâmetros do modelo é dada por

$$\boldsymbol{U}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{pmatrix} \partial\ell(\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\boldsymbol{\beta} \\ \partial\ell(\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} X^\top\Psi^{(2,0)}M_1(y-\mu) \\ S^\top\Phi_1\upsilon \end{pmatrix},$$

em que $(y-\mu) = ((y_1-\mu_1),\ldots,(y_n-\mu_n))^\top$ e $\upsilon = (\upsilon_1,\ldots,\upsilon_n)^\top$, com $\upsilon_l = \psi_l^{(1,1)}(y_l-\mu_l) + \mathrm{T}(y_l) + \psi_l^{(0,1)}$.

As segundas derivadas do logaritmo da função de verossimilhança com relação ao vetor de parâmetros são dadas por

$$\begin{aligned}\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\beta_r\partial\beta_s} &= \sum_{l=1}^{n}[\psi_l^{(3,0)}(\mu_l,\phi_l)m_{1l}^2(y_l-\mu_l) + \psi_l^{(2,0)}(\mu_l,\phi_l)m_{2l}(y_l-\mu_l) \\ &- \psi_l^{(2,0)}(\mu_l,\phi_l)m_{1l}^2](r)_l(s)_l,\end{aligned}$$

$$\frac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial \gamma_R \partial \gamma_S} = \sum_{l=1}^{n} [\psi_l^{(1,2)}(\mu_l,\phi_l)\phi_{1l}^2(y_l-\mu_l) + \psi_l^{(1,1)}(\mu_l,\phi_l)\phi_{2l}(y_l-\mu_l) + \phi_{2l}\mathrm{T}(y_l) + \psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}^2 + \psi_l^{(0,1)}(\mu_l,\phi_l)\phi_{2l}](R)_l(S)_l$$

e

$$\frac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial \beta_r \partial \gamma_S} = \sum_{l=1}^{n} [\psi_l^{(2,1)}(\mu_l,\phi_l)m_{1l}\phi_{1l}(y_l-\mu_l)](r)_l(S)_l,$$

com $r, s = 1, \ldots, p$ e $R, S = 1, \ldots, q$.

A decomposição do espaço paramétrico $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ induz a uma semelhante decomposição da matriz de informação de Fisher $\boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\gamma})$ e da matriz inversa associada $\boldsymbol{K}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})$. Além disso, tem-se que $\boldsymbol{K}_{\beta,\gamma} = \boldsymbol{K}_{\gamma,\beta} = \mathbf{0}$, de modo que essas matrizes possuem uma estrutura bloco-diagonal e podem ser expressas por

$$\boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{pmatrix} \boldsymbol{K}_{\beta,\beta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{K}_{\gamma,\gamma} \end{pmatrix}, \quad \boldsymbol{K}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \begin{pmatrix} \boldsymbol{K}^{\beta,\beta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{K}^{\gamma,\gamma} \end{pmatrix},$$

em que $\mathbf{K}_{\beta,\beta} = X^\top \Psi^{(2,0)} M_1^2 X$ e $\mathbf{K}_{\gamma,\gamma} = -S^\top \Psi^{(0,2)} \Phi_1^2 S$ são as matrizes de informação para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, respectivamente, enquanto que $\mathbf{K}^{\beta,\beta} = (X^\top \Psi^{(2,0)} M_1^2 X)^{-1}$ e $\mathbf{K}^{\gamma,\gamma} = (-S^\top \Psi^{(0,2)} \Phi_1^2 S)^{-1}$ são as correspondentes matrizes inversas.

Os estimadores de máxima verossimilhança para os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, são as quantidades $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$ que satisfazem $U(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = 0$. No caso dos MLGSs, trata-se de uma equação não linear, cuja solução pode ser obtida por meio de algoritmos de otimização, tais como Newton-Raphson, scoring de Fisher e BFGS. Maiores detalhes sobre os métodos de otimização podem ser encontrados em Nocedal e Wright (1999).

Apresentamos a seguir o esquema iterativo scoring de Fisher para a obtenção das $(m+1)$-ésimas estimativas de máxima verossimilhança para β e γ, respectivamente.

Para o vetor de parâmetros $\boldsymbol{\beta}$, temos:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \boldsymbol{K}_{\beta,\beta}^{-1(m)}\, \boldsymbol{U}_{\beta}^{(m)}.$$

Substituindo $\boldsymbol{K}_{\beta,\beta}^{-1}$ e $\boldsymbol{U}_{\beta}$ pelas quantidades correspondentes e com alguma álgebra, tem-se:

$$\begin{aligned}
\boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} + (X^{\top}\Psi^{(2,0)(m)}M_1^{2(m)}X)^{-1}X^{\top}\Psi^{(2,0)(m)}M_1^{(m)}(y-\mu) \\
&= (X^{\top}\Psi^{(2,0)(m)}M_1^{2(m)}X)^{-1}X^{\top}\Psi^{(2,0)(m)}M_1^{2(m)}\zeta_1^{(m)} \\
&= (X^{\top}W_{\beta\beta}^{(m)}X)^{-1}X^{\top}W_{\beta\beta}^{(m)}\zeta_1^{(m)},
\end{aligned}$$

em que $W_{\beta\beta} = \Psi^{(2,0)}M_1^2$ e $\zeta_1 = \eta + M_1^{-1}(y-\mu)$, com $\eta = \boldsymbol{X\beta}$.

Procedendo de maneira análoga para o vetor de parâmetros $\boldsymbol{\gamma}$, temos:

$$\boldsymbol{\gamma}^{(m+1)} = \boldsymbol{\gamma}^{(m)} + \boldsymbol{K}_{\gamma,\gamma}^{-1(m)}\boldsymbol{U}_{\gamma}^{(m)}.$$

Ao substituir $\boldsymbol{K}_{\gamma,\gamma}^{-1}$ e $\boldsymbol{U}_{\gamma}$ pelas quantidades correspondentes e com algumas manipulações algébricas, tem-se:

$$\begin{aligned}
\boldsymbol{\gamma}^{(m+1)} &= \boldsymbol{\gamma}^{(m)} + (-S^{\top}\Psi^{(0,2)(m)}\Phi_1^{2(m)}S)^{-1}S^{\top}\Phi_1^{(m)}v^{(m)} \\
&= (-S^{\top}\Psi^{(0,2)(m)}\Phi_1^{2(m)}S)^{-1}(-S^{\top}\Psi^{(0,2)(m)}\Phi_1^{2(m)})\zeta_2^{(m)} \\
&= (S^{\top}W_{\gamma\gamma}^{(m)}S)^{-1}S^{\top}W_{\gamma\gamma}^{(m)}\zeta_2^{(m)},
\end{aligned}$$

em que $W_{\gamma\gamma} = -\Psi^{(0,2)}\Phi_1^2$ e $\zeta_2 = \tau - {\Psi^{(0,2)}}^{-1}\Phi_1^{-1}v$, enquanto $\tau = \boldsymbol{S\gamma}$ e $v = (v_1, \ldots, v_n)^{\top}$, com $v_l = \psi_l^{(1,1)}(y_l - \mu_l) + \mathrm{T}(y_l) + \psi_l^{(0,1)}$.

Deste modo, tem-se as seguintes equações para estimar β e γ iterativamente

$$\begin{aligned}
\boldsymbol{\beta}^{(m+1)} &= (X^{\top}W_{\beta\beta}^{(m)}X)^{-1}X^{\top}W_{\beta\beta}^{(m)}\zeta_1^{(m)}, \\
\boldsymbol{\gamma}^{(m+1)} &= (S^{\top}W_{\gamma\gamma}^{(m)}S)^{-1}S^{\top}W_{\gamma\gamma}^{(m)}\zeta_2^{(m)}.
\end{aligned}$$

Cabe destacar que, nas equações acima, as quantidades ζ_1 e ζ_2 exercem o papel de variáveis dependentes modificadas, enquanto que $W_{\beta\beta}^{(m)}$ e $W_{\gamma\gamma}^{(m)}$ são as matrizes de pesos que mudam a cada passo do processo iterativo.

2.4 Testes de Hipóteses

Consideremos novamente que o modelo definido em (2.1)-(2.3) tenha vetor de parâmetros particionado como $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, em que $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são os vetores de parâmetros de perturbação e interesse, de dimensões p e q, respectivamente. O interesse é testar as hipóteses

$$H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^{(\mathbf{0})} \quad \text{Vs.} \quad H_1 : \boldsymbol{\gamma} \neq \boldsymbol{\gamma}^{(\mathbf{0})}, \tag{2.5}$$

em que $\boldsymbol{\gamma}^{(\mathbf{0})}$ é um vetor q-dimensional de constantes especificado.

Alguns testes para as hipóteses acima são apresentados a seguir.

2.4.1 Testes da Razão de Verossimilhanças

A estatística da razão de verossimilhanças usual para o teste H_0 em questão é definida por

$$LR = 2\{\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) - \ell(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})\},$$

em que $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$ e $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ são os estimadores de máxima verossimilhança de $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, restritos e irrestritos, respectivamente.

Nos casos em que o modelo em estudo envolve parâmetros de perturbação, inferências baseadas na função de verossimilhança usual conduzem a resultados que podem ser menos adequados do que os alcançados na utilização de inferências baseadas na função de verossimilhança perfilada. Observa-se que a função de verossimilhança perfilada é uma adaptação da verossimilhança usual, em que se substituem os parâmetros de perturbação por suas respectivas estimativas de máxima verossimilhança, sendo assim função apenas dos parâmetros de interesse.

No caso do modelo (2.1)-(2.3), o logaritmo da função de verossimilhança perfilada é dado por

$$\ell_p(\boldsymbol{\gamma}) = \ell(\boldsymbol{\gamma}, \hat{\boldsymbol{\beta}}_\gamma), \tag{2.6}$$

em que $\hat{\boldsymbol{\beta}}_\gamma$ é o valor que maximiza (2.4) supondo $\boldsymbol{\gamma}$ fixo.

Observa-se que a obtenção de um estimador de máxima verossimilhança para $\boldsymbol{\gamma}$ a partir de (2.6) se faz em dois momentos de maximização subsequentes. O primeiro consiste em encontrar a quantidade $\hat{\boldsymbol{\beta}}$ que maximiza

$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ supondo $\boldsymbol{\gamma}$ fixo, enquanto o segundo consiste em maximizar $\ell_p(\boldsymbol{\gamma})$ com relação à $\boldsymbol{\gamma}$.

Embora em muitas situações práticas o uso da verossimilhança perfilada seja uma boa alternativa para inferência quando na presença de parâmetros de perturbação, cabe destacar que, em alguns casos, conforme já apontado por Araújo Júnior (2006), tal método conduz a estimadores que não possuem propriedades de otimalidade desejáveis, tais como consistência e eficiência. Este problema decorre, como já referimos anteriormente, do fato de tal função ser uma adaptação da função de verossimilhança usual, não tendo necessariamente todas as propriedades desta. Outro aspecto a ser destacado é que a qualidade das aproximações envolvidas nas inferências decorrentes de resultados assintóticos é prejudicada quando é grande o número parâmetros de perturbação (Cysneiros, 2004).

Com o intuito de atenuar os problemas anteriormente mencionados, diversos ajustes à verossimilhança perfilada vêm sendo propostos na literatura, tais como os propostos por Barndorff-Nielsen (1983, 1994), Cox e Reid (1987, 1992), McCullagh e Tibishirani (1990) e Stern (1997). Uma revisão dessas propostas pode ser encontrada em Silva (2005).

Neste trabalho, será considerado o ajuste de Cox e Reid (1987). O logaritmo da função de verossimilhança perfilada ajustada para o modelo definido em (2.1)-(2.3) é dado por

$$\ell_{pa}(\boldsymbol{\gamma}) = \ell_p(\boldsymbol{\gamma}) - \frac{1}{2}\log|j_{\beta\beta}(\boldsymbol{\gamma}, \hat{\boldsymbol{\beta}}_\gamma)|, \tag{2.7}$$

em que $j_{\beta\beta}(\boldsymbol{\gamma}, \hat{\boldsymbol{\beta}}_\gamma)$ é a informação observada para $\boldsymbol{\beta}$ quando $\boldsymbol{\gamma}$ é fixado, ou seja

$$\begin{aligned} j_{\beta\beta} = -\frac{\partial^2\ell(\beta,\gamma)}{\partial\beta_r\partial\beta_s} &= \sum_{l=1}^{n}[-\psi_l^{(3,0)}(\mu_l,\phi_l)m_{1l}^2(y_l-\mu_l) - \psi_l^{(2,0)}(\mu_l,\phi_l)m_{2l}(y_l-\mu_l) \\ &+ \psi_l^{(2,0)}(\mu_l,\phi_l)m_{1l}^2](r)_l(s)_l. \end{aligned}$$

A equação (2.7) é denotada por verossimilhança perfilada ajustada e a estatística para o teste da razão de verossimilhanças dela proveniente é dada por

$$LR_{pa} = 2\{\ell_{pa}(\hat{\boldsymbol{\gamma}}) - \ell_{pa}(\boldsymbol{\gamma}^{(0)})\},$$

em que $\hat{\boldsymbol{\gamma}}$ é o estimador de máxima verossimilhança irrestrito de $\boldsymbol{\gamma}$.

2.4.2 Teste Escore

Considerando as hipóteses definidas em (2.5), a estatística do teste escore pode ser expressa como

$$S_R = \tilde{\boldsymbol{U}}_{\gamma}^{\top} \tilde{\boldsymbol{K}}^{\gamma,\gamma} \tilde{\boldsymbol{U}}_{\gamma},$$

em que $\tilde{\boldsymbol{U}}_{\gamma}$ e $\tilde{\boldsymbol{K}}^{\gamma,\gamma}$ são, respectivamente, a função escore e a inversa da matriz de informação para $\boldsymbol{\gamma}$, estimadas segundo a hipótese nula.

Conforme já observa Cordeiro (1999), uma vantagem da estatística do teste escore é que esta envolve a maximização da função de verossimilhança apenas segundo a hipótese nula.

2.4.3 Teste Wald

Além da razão de verossimilhanças e do teste escore, outro teste comumente usado para hipóteses como as propostas em (2.5) é o teste Wald. Para o teste em questão, a estatística Wald pode ser definida como

$$W = (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^{(0)})^{\top} \hat{\boldsymbol{K}}^{\gamma,\gamma-1} (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^{(0)}),$$

em que $\hat{\boldsymbol{K}}^{\gamma,\gamma-1}$ é a inversa da matriz $\hat{\boldsymbol{K}}^{\gamma,\gamma}$, estimada segundo a hipótese alternativa e $\hat{\boldsymbol{\gamma}}$ é o estimador de máxima verossimilhança irrestrito de $\boldsymbol{\gamma}$.

Para realização de um teste estatístico, além das hipóteses e da estatística de teste, é necessário o valor crítico de teste, o qual é obtido da distribuição da estatística de teste sob a hipótese nula. Em geral, não é tarefa simples obter a distribuição exata das estatísticas da razão de verossimilhança, escore e Wald sob H_0. Entretanto, em problemas regulares, assintoticamente e sob a hipótese nula, essas estatísticas convergem em distribuição para uma

Qui-quadrado com q graus de liberdade, em que q denota o número de restrições impostas na hipótese nula. Deste modo, testes aproximados podem ser realizados utilizando-se valores críticos desta distribuição.

Por fim, vale ressaltar que nesta dissertação será dada ênfase aos testes baseados na estatística da razão de verossimilhanças.

2.5 Distribuições Pertencentes aos MLGSs

A classe dos MLGSs contempla distribuições importantes como a Passeio Aleatório, Inversa Gaussiana Generalizada, Exponencial Dupla e Poisson Dupla (Previdelli, 2005). Na Tabela 2.1 são apresentadas para cada uma dessas distribuições a função densidade, média, variância e suporte. Usando uma reparametrização conveniente, a densidade dessas distribuições podem ser reescritas na forma da equação (2.1).

Assim como mencionado por Cordeiro e Botter (2001) e Cordeiro *et al.* (2006), no caso particular da Passeio Aleatório, definindo $\phi = -\delta/(2\theta^2)$, além de

$$\begin{aligned} \psi(\mu,\phi) &= 2(-\phi)^{1/2}[(2\mu-\phi)^{1/2}-(-\phi)^{1/2}]^{-1}+\frac{1}{2}\log 2 \\ &- \log[(2\mu-\phi)^{1/2}-(-\phi)^{1/2}] \\ &- \mu[(2\mu-\phi)^{1/2}-(-\phi)^{1/2}]^{-2}, \end{aligned}$$

tem-se que a densidade desta distribuição pode ser reescrita na forma da equação (2.1).

Outro ponto importante a ser destacado é a relação existente na literatura entre as distribuições Inversa Gaussiana (IG), que é obtida como caso particular da distribuição Inversa Gaussiana Generalizada (IGG) quando $\lambda = -1/2$, e a Passeio Aleatório (PA). Tal relação, como mostra Previdelli (2005), é estabelecida da seguinte forma: Se uma variável aleatória $Z \sim IG(\theta,\delta)$, em que θ e δ são respectivamente os parâmetros de média e dispersão, então $Y = Z^{-1} \sim PA(\mu,\phi)$, com μ e ϕ denotando, respectivamente, os parâmetros de média e dispersão.

Tabela 2.1: Densidade, média, variância e suporte de algumas distribuições definidas na classe dos MLGSs

Distribuições	$f(y;\boldsymbol{\theta})$	E(Y)	Var(Y)	Suporte
1. PA[1]	$\left(\frac{\delta}{2\pi y}\right)^{\frac{1}{2}}\exp\left(-\frac{\delta y}{2}+\frac{\delta}{\theta}-\frac{\delta}{2\theta^2 y}\right)$	$\frac{1}{\theta}+\frac{1}{\delta}$	$\frac{1}{\theta\delta}+\frac{2}{\delta^2}$	$y>0$
2. ED [2]	$c(\mu,\theta,n)\theta^{1/2}\{g(y;\mu,n)\}^{\theta}\{g(y;n)\}^{1-\theta}[dh(y;n)]$	μ	$\frac{V(\mu)}{n\theta}$	$-\infty<y<\infty$
3. IGG.[3]	$\frac{\left(\frac{\alpha}{\delta}\right)^{\frac{\lambda}{2}}}{2B_\lambda(\sqrt{\alpha\delta})}\, y^{\lambda-1}\exp\left[-\frac{1}{2}(\delta y^{-1}+\alpha y)\right]$	$\frac{\sqrt{\delta}B_{\lambda+1}(\sqrt{\alpha\delta})}{\sqrt{\alpha}B_\lambda(\sqrt{\alpha\delta})}$	$\left(\frac{\delta}{\alpha}\right)\left[\frac{B_{\lambda+2}(\sqrt{\alpha\delta})}{B_\lambda(\sqrt{\alpha\delta})}-\left(\frac{B_{\lambda+1}(\sqrt{\alpha\delta})}{B_\lambda(\sqrt{\alpha\delta})}\right)^2\right]$	$y>0$
4. PD[4]	$(\theta^{\frac{1}{2}}e^{-\theta\mu})\left(\frac{e^{-y}y^y}{y!}\right)\left(\frac{e\mu}{y}\right)^{\theta y}$	μ	$\frac{\mu}{\theta}$	$y=0,1,2,3,\ldots$

[1]Passeio Aleatório.
[2]Exponencial Dupla.
[3]Inversa Gaussiana Generalizada - $B_\lambda(.)$ é a função de Bessel modificada de terceira ordem; Considerando $\lambda=-1/2$, tem-se a Inversa Gaussiana como caso particular da Inversa Gaussiana Generalizada.
[4]Poisson Dupla.

CAPÍTULO 3

Correção de Bartlett em MLGSs

3.1 Introdução

Em muitas situações práticas é comum fazer uso de testes estatísticos aproximados para grandes amostras, uma vez que testes exatos nem sempre existem. Os testes aproximados caracterizam-se pelo uso de valores críticos obtidos de uma distribuição limite conhecida, sendo por isso denominados de assintóticos de primeira ordem. São exemplos deste tipo de teste os baseados nas estatísticas da razão verossimilhança, escore e Wald. Estes têm, sob a hipótese nula (H_0), distribuição limite comum Qui-quadrado com q graus de liberdade (χ^2_q), em que q denota o número de restrições impostas em H_0, sendo, portanto, assintoticamente equivalentes (Cordeiro, 1999).

Apesar de conduzir a resultados satisfatórios em muitos casos, as aproximações de primeira ordem podem ser pobres se o tamanho da amostra é pequeno ou moderado, dando origem a testes cujas taxas de rejeição em muito diferem dos níveis nominais desejados.

Visando melhorar a qualidade da aproximação de primeira ordem em amostras finitas, Bartlett (1937) propôs um fator de correção para a estatística da razão de verossimilhanças (LR). Este autor observou que, sob a hipótese nula, a esperança da estatística LR tem expansão na forma $q+b+O(n^{-2})$, em que b é uma constante de ordem $O(n^{-1})$. Deste modo, seria possível obter uma nova estatística de teste, cujo primeiro momento fosse mais próximo de seu correspondente na distribuição χ^2_q de referência, aplicando sobre a estatística usual um fator de correção da magnitude de seu primeiro momento. De fato, a nova estatística corrigida LR^c tem, sob a hipótese nula, distribuição limite χ^2_q com erro de aproximação de ordem $O(n^{-2})$.

DiCiccio e Stern (1994) obtiveram uma expressão geral do fator de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas ajustadas, com a qual é possível obter uma estatística melhorada cuja distribuição sob a hipótese nula seja mais próxima da distribuição χ^2_q de referência, com erro de aproximação que passa de $O(n^{-1})$ para $O(n^{-2})$.

O presente Capítulo apresenta correções de Bartlett para a estatística da razão de verossimilhanças usual e estatística da razão de verossimilhanças perfiladas ajustadas na classe dos MLGSs.

Nos últimos anos, diversos trabalhos considerando correções de Bartlett em diferentes contextos vêm sendo propostos na literatura. Algumas das principais referências são apresentadas a seguir. Cribari-Neto e Cordeiro (1996) apresentam uma detalhada revisão da literatura sobre correção de Bartlett para a estatística da razão de verossimilhanças e correções tipo Bartlett para as estatísticas escore e Wald. Ferrari e Uribe-Opazo (2001) derivam um fator de correção de Bartlett para estatística da razão de verossimilhanças na classe de modelos de regressão lineares simétricos. Ferrari *et al.* (2004) obtiveram um fator de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas ajustadas em modelos de regressão lineares normais. Barroso e Cordeiro (2005) obtiveram um fator de correção de Bartlett para a estatística da razão de verossimilhanças em modelos de regressão t heterocedásticos. Araújo Júnior (2006) obteve verossimilhanças perfiladas ajustadas por Cox e Reid (1987) e Barndorff-Nielsen (1983) na distribuição Birnbaum-Saunders. Cordeiro *et al.* (2006), generalizando resultados obtidos por Botter e Cordeiro (1998) e Cordeiro (1983), apresentaram fatores de correção de Bartlett para estatística da razão de verossimilhanças em MLGSs. Lemonte *et al.* (2010) derivaram um fator de correção de Bartlett para estatística da razão de verossimilhanças em modelos de regressão Birnbaum-Saunders. Araújo (2012) obteve um fator de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas ajustadas da classe de modelos não lineares simétricos heteroscedásticos.

3.2 Fator de Correção de Bartlett

Seja $f(\boldsymbol{y};\boldsymbol{\theta})$ um modelo paramétrico, sendo $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ um vetor de parâmetros desconhecidos, em que $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ tem dimensões p e q, respectivamente. Sejam $\ell(\boldsymbol{\theta})$ logaritmo da função de verossimilhança e $\boldsymbol{K}(\boldsymbol{\theta})$ a matriz de informação de Fisher, cuja matriz inversa correspondente é denotada por $\boldsymbol{K}(\boldsymbol{\theta})^{-1}$.

Introduziremos agora as seguintes notações: Os índices $r, s, t, ...$ variam sobre os parâmetros de perturbação, enquanto $R, S, T, ...$ serão indexadores dos parâmetros de interesse e $a, b, c, ...$ serão índices que variam sobre os $p + q$ parâmetros do modelo. As derivadas do logaritmo da função de verossimilhança com relação aos componentes do vetor $\boldsymbol{\theta}$ são representadas por $U_a = \partial\ell(\boldsymbol{\theta})/\partial\theta_a$, $U_{ab} = \partial^2\ell(\boldsymbol{\theta})/\partial\theta_a\theta_b$, etc. Os cumulantes conjuntos serão denotados por $\kappa_{ab} = E(U_{ab})$, $\kappa_{abc} = E(U_{abc})$ etc. As derivadas dos cumulantes com relação aos componentes do vetor $\boldsymbol{\theta}$ serão representados por $\kappa_{ab}^{(c)} = \partial\kappa_{ab}/\partial\theta_c$. Por fim, destaca-se que a matriz de informação tem elementos $-\kappa_{ab}$, sendo $-\kappa^{ab}$, os elementos da matriz inversa correspondente.

Seja LR a estatística da razão de verossimilhanças para o teste das hipóteses $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^{(0)}$ contra $H_1 : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_2^{(0)}$, em que $\boldsymbol{\theta}_2^{(0)}$ é um vetor de constantes conhecidas. Sabe-se que a estatística LR tem, assintoticamente e sob a hipótese nula, distribuição χ_q^2 com um erro de aproximação da ordem de $O(n^{-1})$. Visando melhorar a qualidade desta aproximação, Bartlett (1937) considerou multiplicar a estatística de teste usual pelo fator de correção

$$c = (1 + b/q)^{-1}, \tag{3.1}$$

em que b é uma constante de ordem $O(n^{-1})$. O fator de correção definido em (3.1) baseia-se na relação $c = q^{-1}E(LR)$ e tornou-se conhecido como *fator de correção de Bartlett*.

Diversas metodologias para obtenção de fatores de correção de Bartlett são descritas na literatura, dentre estas a proposta por Lawley (1956). Este autor, considerando problemas regulares, obteve uma fórmula geral para a constante b em termos de cumulantes e derivadas de cumulantes do logaritmo da função de verossimilhança dada por

$$b = \epsilon_{p+q} - \epsilon_p, \tag{3.2}$$

em que ϵ_{p+q} é um termo de ordem $O(n^{-1})$ definido como

$$\epsilon_{p+q} = \sum (l_{abcd} - l_{abcdef}), \tag{3.3}$$

em que o somatório estende-se a todos os componentes do vetor $\boldsymbol{\theta}$. Os termos l_{abcd} e l_{abcdef} são expressos em função dos cumulantes e derivadas de cumulantes

$$l_{abcd} = \kappa^{ab}\kappa^{cd}\left(\frac{\kappa_{abcd}}{4} - \kappa_{abc}^{(d)} + \kappa_{ac}^{(bd)}\right),$$

$$\begin{aligned} l_{abcdef} &= \kappa^{ab}\kappa^{cd}\kappa^{ef}\big[\kappa_{ace}\left(\frac{\kappa_{bdf}}{6} - \kappa_{bf}^{(d)}\right) + \kappa_{acd}\left(\frac{\kappa_{bef}}{4} - \kappa_{bf}^{(e)}\right) \\ &+ \kappa_{ac}^{(e)}\kappa_{bf}^{(d)} + \kappa_{ac}^{(d)}\kappa_{bf}^{(e)}\big]. \end{aligned}$$

O termo ϵ_p na equação (3.2), também é da ordem $O(n^{-1})$ e pode ser obtido de forma análoga a ϵ_{p+q}, a partir da equação (3.3), com o somatório estendendo-se sobre os p parâmetros de perturbação. Tem-se portanto, um fator de correção de Bartlett com o qual é possível obter, a partir da estatística LR usual, uma nova estatística melhorada

$$LR^c = \frac{LR}{1 + b/q},$$

cuja distribuição sob a hipótese nula seja mais próxima da χ_q^2 de referência. De fato, o erro desta aproximação passa a ser de ordem $O(n^{-2})$.

Uma forma alternativa para a realização do teste $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^{(0)}$ contra $H_1 : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_2^{(0)}$ é considerar a estatística da razão de verossimilhanças perfiladas ajustadas LR_{pa}. Assim como a estatística usual, a aproximação da distribuição de LR_{pa} sob a hipótese nula pela distribuição χ_q^2, tem erro de ordem $O(n^{-1})$. Objetivando reduzir distorções de tamanho do teste em amostras finitas, DiCiccio e Stern (1994) propuseram um fator de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas ajustadas, definido por

$$c_{pa} = (1 + b_{pa}/q)^{-1}, \tag{3.4}$$

em que b_{pa} é definido em função de cumulantes do logaritmo da função de verossimilhança perfilada ajustada e das quantidades $\tau^{ab} = \kappa^{aR}\kappa^{bS}\sigma_{RS}$, com (σ_{RS}) denotando a inversa da matriz (κ^{RS}) e $\nu^{ab} = \kappa^{ab} - \tau^{ab}$. Assim,

$$\begin{aligned}
b_{pa} &= \frac{1}{4}\tau^{ad}\tau^{bc}\kappa_{abcd} - \kappa^{ad}\tau^{bc}\kappa_{abc}^{(d)} + (\kappa^{ad}\kappa^{bc} - \nu^{ad}\nu^{bc})\kappa_{ab}^{(cd)} \\
&- \left(\frac{1}{4}\kappa^{ad}\tau^{bc}\tau^{ef} + \frac{1}{2}\kappa^{ad}\tau^{bf}\tau^{ce} - \frac{1}{3}\tau^{ad}\tau^{bf}\tau^{ce}\right)\kappa_{abc}\kappa_{def} \\
&+ (\kappa^{ad}\tau^{bc}\kappa^{ef} + \kappa^{ad}\kappa^{bf}\kappa^{ce} - \nu^{ad}\kappa^{bf}\nu^{ce})\kappa_{abc}\kappa_{de}^{(f)} \\
&- (\kappa^{ad}\kappa^{bc}\kappa^{ef} - \nu^{ad}\nu^{bc}\nu^{ef} + \kappa^{ad}\kappa^{bf}\kappa^{ce} - \nu^{ad}\nu^{bf}\nu^{ce})\kappa_{ab}^{(c)}\kappa_{de}^{(f)}.
\end{aligned}$$

Com base no fator de correção definido em (3.4), obtem-se uma estatística da razão de verossimilhanças perfiladas ajustadas corrigida

$$LR_{pa}^c = \frac{LR_{pa}}{1 + b_{pa}/q},$$

para qual o erro da aproximação da distribuição desta estatística pela distribuição χ_q^2 passa a ser de ordem $O(n^{-2})$.

Os fatores de correção de Bartlett, apresentados em (3.1) e (3.4), são bastante gerais, podendo, em princípio, ser calculados para qualquer classe de modelos. Como foi mostrado, a obtenção de tais fatores requer basicamente o cálculo de cumulantes e derivadas de cumulantes do modelo. Estas quantidades satisfazem certas identidades, denominadas *identidades de Bartlett*, que, em muitos casos, facilitam os cálculos. Para maiores detalhes sobre identidades de Bartlett, ver Cordeiro (1999, Capítulo 5).

3.3 Fator de Correção de Bartlett nos MLGSs

Considerando uma partição conveniente do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, com $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ da classe dos MLGSs, Cordeiro *et al.* (2006) obtiveram um fator de correção de Bartlett para a estatística da razão de verossimilhanças do teste $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^{(0)}$ contra $H_1 : \boldsymbol{\gamma} \neq \boldsymbol{\gamma}^{(0)}$, em que $\boldsymbol{\gamma}^{(0)}$ é um vetor de constantes conhecidas. Tal fator de correção pode ser expresso como

$$c = \{1 + [\epsilon_q(\gamma) + \epsilon_{p,q}(\beta, \gamma)]/q\}^{-1}, \tag{3.5}$$

em que as quantidades $\epsilon_q(\gamma)$ e $\epsilon_{p,q}(\beta, \gamma)$ são obtidas como função das seguintes matrizes de dimensão n: $\Psi^{(r,s)} = \mathrm{diag}\{\psi_1^{(r,s)}, \ldots, \psi_n^{(r,s)}\}$ para $r, s = 1, \ldots, 4$, $M_r = \mathrm{diag}\{m_{r1}, \ldots, m_{rn}\}$, $\Phi_r = \mathrm{diag}\{\phi_{r1}, \ldots, \phi_{rn}\}$ para $r = 1, 2$, $Z_\beta = X(X^\top \Psi^{(2,0)} M_1^2 X)^{-1} X^\top$ e $Z_\gamma = S(-S^\top \Psi^{(0,2)} \Phi_1^2 S)^{-1} S^\top$. Definem-se ainda $Z_{\gamma d}$ e $Z_{\beta d}$ como matrizes diagonais com elementos correspondentes de Z_γ e Z_β, respectivamente. Adicionalmente, $\mathbf{1}$ é um vetor $n \times 1$ de 1's. Assim temos,

$$\begin{aligned}
\epsilon_q(\gamma) &= \frac{1}{4}\mathbf{1}^\top(\Psi^{(0,4)}\Phi_1^4 + 2\Psi^{(0,3)}\Phi_1^2\Phi_2 - \Psi^{(0,2)}\Phi_2^2)Z_{\gamma d}^{(2)}\mathbf{1} \\
&+ \frac{1}{12}\mathbf{1}^\top\Psi^{(0,3)}\Phi_1^3(2Z_\gamma^{(3)} + 3Z_{\gamma d}Z_\gamma Z_{\gamma d})\Phi_1^3\Psi^{(0,3)}\mathbf{1} \\
&+ \frac{1}{4}\mathbf{1}^\top\Psi^{(0,2)}\Phi_1\Phi_2(Z_{\gamma d}Z_\gamma Z_{\gamma d} + 2Z_\gamma^{(3)})\Phi_1\Phi_2\Psi^{(0,2)}\mathbf{1} \\
&+ \frac{1}{2}\mathbf{1}^\top\Psi^{(0,3)}\Phi_1^3 Z_{\gamma d}Z_\gamma Z_{\gamma d}\Phi_1\Phi_2\Psi^{(0,2)}\mathbf{1},
\end{aligned}$$

$$\begin{aligned}\epsilon_{p,q}(\beta,\gamma) &= \frac{1}{2}\mathbf{1}^\top[(\Psi^{(2,1)}\Phi_2 + \Psi^{(2,2)}\Phi_1^2)M_1^2 Z_{\beta d} Z_{\gamma d} \\ &- \Psi^{(2,1)}\Phi_1 M_1^2 Z_\beta^{(2)} Z_\gamma \Phi_1 M_1^2 \Psi^{(2,1)}]\mathbf{1} \\ &+ \frac{1}{4}\mathbf{1}^\top \Psi^{(2,1)}\Phi_1 M_1^2 Z_{\beta d} Z_\gamma [Z_{\beta d} M_1^2 \Psi^{(2,1)} \\ &+ 2Z_{\gamma d}(\Phi_1^2 \Psi^{(0,3)} + \Phi_2 \Psi^{(0,2)})]\Phi_1 \mathbf{1},\end{aligned}$$

sendo $Z_{\gamma d}^{(2)} = Z_{\gamma d} \odot Z_{\gamma d}$, $Z_\gamma^{(3)} = Z_\gamma^{(2)} \odot Z_\gamma$, $Z_\beta^{(2)} = Z_\beta \odot Z_\beta$ com $\odot$ denotando o produto de Hadamard ou produto direto.

A expressão definida em (3.5) pode ser utilizada para obter uma estatística da razão de verossimilhanças corrigida na classe dos MLGSs, cuja distribuição sob a hipótese nula seja mais próxima da Qui-quadrado de referência (Cordeiro *et al.*, 2006).

Ainda considerando testar as hipóteses $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^{(0)}$ contra $H_1 : \boldsymbol{\gamma} \neq \boldsymbol{\gamma}^{(0)}$, derivamos a seguir um fator de correção de Bartlett, segundo a metodologia proposta por DiCiccio e Stern (1994) apresentada na seção 3.2, para a estatística da razão de verossimilhanças perfiladas ajustadas da classe dos MLGSs. Para tanto, considerou-se o fato de que os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ do modelo são globalmente ortogonais, de modo que $\tau^{Rs} = \tau^{rS} = \tau^{rs} = 0$, $\nu^{RS} = \nu^{Rs} = \nu^{rS} = 0$ e ainda $\tau^{RS} = \kappa^{RS}$, $\nu^{rs} = \kappa^{rs}$. Assim, com alguma álgebra, obteve-se o fator de correção[1]

$$c_{pa} = (1 + b_{pa}/q)^{-1}, \tag{3.6}$$

em que

[1]Assim como antes, letras maiúsculas são usadas como indexadores do parâmetro de interesse enquanto letras minúsculas são usadas para indexar os parâmetros de perturbação.

$$
\begin{aligned}
b_{pa} &= \frac{1}{4}\kappa^{RS}\kappa^{TU}\kappa_{RTUS} - \kappa^{RS}\kappa^{TU}\kappa_{RTU}^{(S)} + \kappa^{RS}\kappa^{TU}\kappa_{RT}^{(US)} \\
&- \kappa^{rs}\kappa^{RS}\kappa_{rRS}^{(s)} + \kappa^{rs}\kappa^{RS}\kappa^{tu}\kappa_{rRS}\kappa_{st}^{(u)} \\
&- (\frac{1}{4}\kappa^{rs}\kappa^{RS}\kappa^{TU} + \frac{1}{2}\kappa^{rs}\kappa^{RU}\kappa^{ST})\kappa_{rRS}\kappa_{sTU} \\
&- (\frac{1}{4}\kappa^{RS}\kappa^{TU}\kappa^{VW} + \frac{1}{6}\kappa^{RS}\kappa^{TW}\kappa^{UV})\kappa_{RTU}\kappa_{SVW} \\
&+ (\kappa^{RS}\kappa^{TU}\kappa^{VW} + \kappa^{RS}\kappa^{TW}\kappa^{UV})\kappa_{RTU}\kappa_{SV}^{(W)} \\
&- (\kappa^{RS}\kappa^{TU}\kappa^{VW} + \kappa^{RS}\kappa^{TW}\kappa^{UV})\kappa_{RT}^{(U)}\kappa_{SV}^{(W)} \\
&+ \kappa^{RS}\kappa^{rs}\kappa^{TU}\kappa_{RrT}\kappa_{SU}^{(s)}. \qquad (3.7)
\end{aligned}
$$

O resultado obtido na expressão (3.7) coincide com o obtido por Araújo (2012), e pode ser utilizado em um contexto geral a fim de obter fatores de correção de Bartlett para a estatística da razão de verossimilhanças perfiladas ajustadas em qualquer classe de modelos cujos parâmetros sejam globalmente ortogonais, além de particionados de forma semelhante a que fora aqui considerada.

A fim de particularizar a expressão definida em (3.7) à classe dos MLGSs, podemos reescrevê-la em termos de cumulantes e derivadas de cumulantes do modelo. Tais cumulantes, bem como suas respectivas derivadas, encontram-se definidas no Apêndice A e podem ser expressos por meio das quantidades ω_{1l}, ω_{2l}, ω_{3l}, ω_{4l} e ω_{5l} definidas, respectivamente, como

$$\omega_{1l} = \psi_l^{(0,4)}\phi_{1l}^4 + 6\psi_l^{(0,3)}\phi_{1l}^2\phi_{2l} + 3\psi_l^{(0,2)}\phi_{2l}^2 + 4\psi_l^{(0,2)}\phi_{1l}\phi_{3l},$$

$$\omega_{2l} = \psi_l^{(0,4)}\phi_{1l}^4 + 6\psi_l^{(0,3)}\phi_{1l}^2\phi_{2l} + 3\psi_l^{(0,2)}\phi_{2l}^2 + 3\psi_l^{(0,2)}\phi_{1l}\phi_{3l},$$

$$\omega_{3l} = \psi_l^{(0,4)}\phi_{1l}^4 + 5\psi_l^{(0,3)}\phi_{1l}^2\phi_{2l} + 2\psi_l^{(0,2)}\phi_{2l}^2 + 2\psi_l^{(0,2)}\phi_{1l}\phi_{3l},$$

$$\omega_{4l} = \psi_l^{(0,3)}\phi_{1l}^3 + 3\psi_l^{(0,2)}\phi_{1l}\phi_{2l}$$

e

$$\omega_{5l} = \psi_l^{(0,3)}\phi_{1l}^3 + 2\psi_l^{(0,2)}\phi_{1l}\phi_{2l},$$

com $l = 1, \ldots, n$.

Assim, após alguma álgebra obtemos:

$$\begin{aligned}
\kappa_{RTUS} &= \sum_{l=1}^{n} \omega_{1l}(R,S,T,U)l, \quad \kappa_{RTU}^{(S)} = \sum_{l=1}^{n} \omega_{2l}(R,S,T,U)l, \\
\kappa_{RT}^{(US)} &= \sum_{l=1}^{n} \omega_{3l}(R,S,T,U)l, \quad \kappa_{RTU} = \sum_{l=1}^{n} \omega_{4l}(R,T,U)l \\
\kappa_{RT}^{(U)} &= \sum_{l=1}^{n} \omega_{5l}(R,T,U)l.
\end{aligned}$$

Adicionalmente, temos que alguns cumulantes e derivadas de cumulantes do modelo são nulos: $\kappa_{RrT} = \kappa_{rRS} = \kappa_{rRS}^{(s)} = 0$. Desse modo, a quantidade b_{pa}, definida em (3.7), reduz-se à

$$\begin{aligned}
b_{pa} &= \frac{1}{4}\kappa^{RS}\kappa^{TU}\kappa_{RTUS} - \kappa^{RS}\kappa^{TU}\kappa_{RTU}^{(S)} + \kappa^{RS}\kappa^{TU}\kappa_{RT}^{(US)} \\
&- (\frac{1}{4}\kappa^{RS}\kappa^{TU}\kappa^{VW} + \frac{1}{6}\kappa^{RS}\kappa^{TW}\kappa^{UV})\kappa_{RTU}\kappa_{SVW} \\
&+ (\kappa^{RS}\kappa^{TU}\kappa^{VW} + \kappa^{RS}\kappa^{TW}\kappa^{UV})\kappa_{RTU}\kappa_{SV}^{(W)} \\
&- (\kappa^{RS}\kappa^{TU}\kappa^{VW} + \kappa^{RS}\kappa^{TW}\kappa^{UV})\kappa_{RT}^{(U)}\kappa_{SV}^{(W)}. \qquad (3.8)
\end{aligned}$$

Substituindo em (3.8) as quantidades obtidas para os cumulantes e derivadas dos cumulantes em termos dos $\omega's$, temos

$$
\begin{aligned}
b_{pa} \;=\; & \frac{1}{4}\sum_{l=1}^{n}(R)_l\kappa^{RS}(S)_l\omega_{1l}(T)_l\kappa^{TU}(U)_l \\
- & \sum_{l=1}^{n}(R)_l\kappa^{RS}(S)_l\omega_{2l}(T)_l\kappa^{TU}(U)_l \\
+ & \sum_{l=1}^{n}(R)_l\kappa^{RS}(S)_l\omega_{3l}(T)_l\kappa^{TU}(U)_l \\
- & \frac{1}{4}\sum_{l=1}^{n}\sum_{i=1}^{n}(T)_l\kappa^{TU}(U)_l\omega_{4l}(R)_l\kappa^{RS}(S)_i\omega_{4i}(V)_i\kappa^{VW}(W)_i \\
- & \frac{1}{6}\sum_{l=1}^{n}\sum_{i=1}^{n}(R)_l\kappa^{RS}(S)_i\omega_{4l}(T)_l\kappa^{TW}(W)_i\omega_{4i}(U)_l\kappa^{UV}(V)_i \\
+ & \sum_{l=1}^{n}\sum_{i=1}^{n}(T)_l\kappa^{TU}(U)_l\omega_{4l}(R)_l\kappa^{RS}(S)_i\omega_{5i}(V)_i\kappa^{VW}(W)_i \\
+ & \sum_{l=1}^{n}\sum_{i=1}^{n}(R)_l\kappa^{RS}(S)_i\omega_{4l}(T)_l\kappa^{TW}(W)_i\omega_{5i}(U)_l\kappa^{UV}(V)_i \\
- & \sum_{l=1}^{n}\sum_{i=1}^{n}(T)_l\kappa^{TU}(U)_l\omega_{5l}(R)_l\kappa^{RS}(S)_i\omega_{5i}(V)_i\kappa^{VW}(W)_i \\
- & \sum_{l=1}^{n}\sum_{i=1}^{n}(R)_l\kappa^{RS}(S)_i\omega_{5l}(T)_l\kappa^{TW}(W)_i\omega_{5i}(U)_l\kappa^{UV}(V)_i.
\end{aligned}
$$

Visando simplificar a implementação do fator de correção, é conveniente reescrevê-lo em notação matricial. Para tanto, definem-se as seguintes matrizes quadradas de ordem n: Ω_1, Ω_2, Ω_3, Ω_4 e Ω_5, com elementos ω_{1l}, ω_{2l}, ω_{3l}, ω_{4l} e ω_{5l}, respectivamente, com $l = 1, \ldots, n$ e $H_\gamma = \{h_{\gamma li}\} = S(S^\top \Psi^{(0,2)} \Phi_1^2 S)^{-1} S^\top$ com elementos $h_{\gamma li} = (R)_l \kappa^{RS} (S)_i$, com $l = 1, \ldots, n$ e $i = 1, \ldots, n$. Adicionalmente, define-se a matriz $H_{\gamma d}$ como a matriz diagonal associada a H_γ. Assim, a forma matricial do fator de correção é dada por

$$
\begin{aligned}
b_{pa} &= \frac{1}{4}\mathbf{1}^\top H_{\gamma d}^{(2)}(\Omega_1 - 4\Omega_2 + 4\Omega_3)\mathbf{1} \\
&- \frac{1}{4}\mathbf{1}^\top H_{\gamma d}\Omega_4 H_\gamma \Omega_4 H_{\gamma d}\mathbf{1} - \frac{1}{6}\mathbf{1}^\top \Omega_4 H_\gamma^{(3)}\Omega_4\mathbf{1} \\
&+ \mathbf{1}^\top H_{\gamma d}\Omega_4 H_\gamma \Omega_5 H_{\gamma d}\mathbf{1} + \mathbf{1}^\top \Omega_4 H_\gamma^{(3)}\Omega_5\mathbf{1} \\
&- \mathbf{1}^\top H_{\gamma d}\Omega_5 H_\gamma \Omega_5 H_{\gamma d}\mathbf{1} - \mathbf{1}^\top \Omega_5 H_\gamma^{(3)}\Omega_5\mathbf{1}, \qquad (3.9)
\end{aligned}
$$

em que $H_{\gamma d} = \mathrm{diag}\{h_{\gamma 11}, \ldots, h_{\gamma nn}\}$, $H_{\gamma d}^{(2)} = \mathrm{diag}\{h_{\gamma 11}^2, \ldots, h_{\gamma nn}^2\}$ e $H_\gamma^{(3)} = (h_{\gamma li}^3)$.

A equação (3.9) acima apresenta uma expressão em notação matricial, com a qual pode-se obter um fator de correção de Bartlett à estatística da razão de verossimilhanças perfiladas ajustadas, para o teste conjunto dos efeitos da dispersão na classe dos MLGSs. Esta expressão envolve apenas operações simples entre matrizes, podendo ser facilmente implementada em linguagens de programação, tais como `C`, `Ox` e `R`.

CAPÍTULO 4

Avaliação Numérica

4.1 Introdução

Neste Capítulo são apresentados alguns resultados de simulação realizados com o intuito de avaliar a eficácia do fator de correção de Bartlett, derivado no Capítulo anterior, para a estatística da razão de verossimilhanças perfiladas ajustadas na classe dos Modelos Lineares Generalizados com Superdispersão.

4.1.1 Procedimentos Metodológicos

Por meio de estudos de simulação de Monte Carlo, comparamos os desempenhos dos testes baseados nas seguintes estatísticas: razão de verossimilhanças usual (LR), razão de verossimilhanças perfiladas ajustadas (LR_{pa}) e razão de verossimilhanças perfiladas ajustadas corrigida (LR_{pa}^{c}), sendo esta última obtida através do fator de correção definido pelas expressões (3.6) e (3.9) do Capítulo anterior.

Os desempenhos dos testes foram avaliados segundo dois critérios. O primeiro consiste em verificar a proximidade da probabilidade de rejeição da hipótese nula, sendo esta verdadeira (probabilidade de erro tipo I), com relação aos respectivos níveis nominais fixados. Ou seja, estimamos por simulação de Monte Carlo as taxas de rejeição dos testes, isto é, $P(LR \geq x_\alpha)$, $P(LR_{pa} \geq x_\alpha)$ e $P(LR^c_{pa} \geq x_\alpha)$, em que x_α é o quantil $(1-\alpha)$ da distribuição χ^2_q, com q definindo o número de restrições impostas na hipótese nula. Estes testes são conhecidos na literatura como *tamanho do teste*. Neste contexto, foram analisados os impactos nas taxas de rejeição ocasionados pelo aumento no número de parâmetros de interesse e perturbação. O segundo critério consiste em verificar o *poder dos testes*, ou seja, probabilidade de rejeição da hipótese nula, sendo esta falsa.

As simulações baseiam-se na distribuição Passeio Aleatório. Para gerar as ocorrências desta distribuição com média μ e parâmetro de dispersão ϕ, utilizou-se a relação existente com a distribuição Inversa Gaussiana: Seja uma variável aleatória $Z_i \sim IG(\theta_i, \delta_i)$, em que θ_i e δ_i são, respectivamente, os parâmetros de média e dispersão dados por

$$\theta_i = \exp(\beta_1 x_{1i} + \ldots + \beta_p x_{pi})$$

e

$$\delta_i = \exp(\gamma_1 s_{1i} + \ldots + \gamma_q s_{qi}).$$

Então $Y_i = Z_i^{-1} \sim PA(\mu_i, \phi_i)$, em que $\mu_i = 1/\theta_i + 1/\delta_i$ e $\phi_i = -\delta_i/(2\theta_i^2)$.

As matrizes de covariadas $\boldsymbol{X}$ e $\boldsymbol{S}$ foram geradas como ocorrências independentes de $N(0,1)$ e $U(0,1)$, respectivamente. Todos os parâmetros do modelo foram fixados com valor um, exceto aqueles fixados na hipótese nula.

Foram considerados no experimento 10.000 réplicas de Monte Carlo, diferentes tamanhos amostrais, a saber: n = 20, 30, 40 e 50, diferentes níveis de significância: $\alpha = 0,5\%$, $1,0\%$, $5,0\%$ e $10,0\%$, além de diversos valores para os números de parâmetros de perturbação (p) e interesse (q).

Em todos os cenários considerados, as hipóteses testadas foram $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ Vs. $H_1 : \boldsymbol{\gamma} \neq \mathbf{0}$, em que $\boldsymbol{\gamma}$ é um vetor q-dimensional e $\mathbf{0}$ é um vetor de zeros de mesma dimensão.

Por fim, salienta-se que todas as entradas nas tabelas que se seguem correspondem a porcentagens.

4.2 Resultados e Discussões

Nas Tabelas 4.1 à 4.4 encontram-se os resultados das simulações realizadas com o intuito de avaliar os desempenhos dos testes no que se refere à proximidade da probabilidade de rejeição da hipótese nula, sendo esta verdadeira, quando comparada aos respectivos níveis nominais fixados.

Na Tabela 4.1 são apresentadas as taxas de rejeição empíricas dos testes para o modelo Passeio Aleatório com $p = 3, q = 6$ e $p = 4, q = 3$, considerando diferentes tamanhos amostrais, a saber: $n = 20$, 30, 40 e 50, e diferentes níveis de significância, a saber: $\alpha = 0,5\%$, $1,0\%$, $5,0\%$ e $10,0\%$. De modo geral, os resultados indicam que em amostras de tamanho pequeno o teste da razão de verossimilhanças usual, baseado na estatística LR, é notavelmente liberal, ou seja, rejeita a hipótese nula com uma frequência consideravelmente superior aos níveis nominais fixados. Por exemplo, para $p = 3$ e $q = 6$, $n = 20$ e $\alpha = 10\%$, a taxa de rejeição deste teste é de $40,1\%$, ou seja mais que quatro vezes maior que o nível nominal selecionado. Também verifica-se que, mesmo para amostra de tamanho grande, o teste baseado em LR ainda é liberal. Novamente considerando $\alpha = 10\%$, $p = 3$ e $q = 6$, agora $n = 50$, essa taxa de rejeição passa a ser de $16,5\%$. Observa-se ainda que o teste da razão de verossimilhanças perfiladas ajustadas, baseado na estatística LR_{pa}, é levemente liberal, exibindo taxas de rejeição mais próximas dos níveis nominais do que o teste baseado na razão de verossimilhanças usual. Como ilustração, para $p = 3$ e $q = 6$, $n = 30$ e $\alpha = 10\%$, a taxa de rejeição deste teste é $13,8\%$. Ainda como ilustração, considerando o modelo com $p = 4$ e $q = 3$, $n = 40$, além de $\alpha = 5\%$, a taxa de rejeição deste teste é de $6,5\%$. Já a aplicação da correção de Bartlett na estatística da razão de verossimilhanças perfiladas ajustadas, originou um teste com resultados bastante favoráveis, de modo que o teste baseado em LR^c_{pa} exibe taxas de rejeição mais próximas dos níveis nominais correspondentes. Novamente, considerando $p = 3$ e $q = 6$, $n = 30$ e $\alpha = 10\%$, a taxa de rejeição deste teste é $10,3\%$, enquanto que, para $p = 4$ e $q = 3$, $n = 40$ e $\alpha = 5\%$ a taxa de rejeição do teste é de $5,8\%$.

Na Tabela 4.2 são apresentadas as taxas de rejeição empíricas dos testes para o modelo Passeio Aleatório com $q = 4$, $n = 40$, vários níveis de significância e vários valores para p. O intuito é analisar o impacto do aumento no número de parâmetros de perturbação nas taxas de rejeição dos testes considerados. De modo geral, observa-se que o teste baseado na estatística LR apresenta taxas de rejeição bem superiores aos níveis nominais desejados, e que esta tendência liberal acentua-se à medida que o número de parâmetros de perturbação do modelo aumenta. Por outro lado, verifica-se que os testes baseados nas estatísticas LR_{pa} e LR_{pa}^c tendem a ser mais estáveis, além de apresentarem taxas de rejeição mais próximas dos níveis nominais considerados. Verifica-se ainda que o teste baseado na estatística LR_{pa}^c é o que exibe taxas de rejeição mais próximas dos níveis fixados. Como ilustração, para $p = 3$ e $\alpha = 10\%$, as taxas de rejeição estimadas para os testes baseados nas estatísticas LR, LR_{pa} e LR_{pa}^c são, respectivamente, $16,6\%$, $12,4\%$ e $11,0\%$. Ainda considerando $\alpha = 10\%$, agora com $p = 7$, as taxas passam a ser $31,6\%$, $12,1\%$ e $10,9\%$.

A fim de avaliar o impacto do aumento no número de parâmetros de interesse no desempenho dos testes, a Tabela 4.3 apresenta as taxas de rejeição empíricas dos testes em vários níveis de significância para o modelo Passeio Aleatório com valor fixo $p = 3$, $n = 40$ e vários valores para q. Com base nos resultados, verifica-se para os cenários considerados que o aumento no número de parâmetros de interesse tem impacto semelhante ao produzido pelo aumento no número de parâmetros de perturbação. De modo geral, o teste baseado na estatística LR é consideravelmente liberal, rejeitando a hipótese nula com frequência bem superior a desejada, enquanto que os testes baseados em LR_{pa} e LR_{pa}^c apresentaram taxas de rejeição mais próximas dos níveis nominais considerados. Também para este cenário verifica-se que o teste baseado na estatística LR_{pa}^c é o que apresenta taxas de rejeição empíricas mais próximas dos níveis nominais considerados. Por exemplo, para $q = 4$ e nível de significância $\alpha = 5,0\%$, as taxas de rejeição dos testes baseados nas estatísticas LR, LR_{pa} e LR_{pa}^c são, respectivamente, $9,7\%$, $6,8\%$ e $5,8\%$. Para o mesmo nível de significância e considerando $q = 6$ as taxas são $11,5\%$, $7,1\%$ e $5,3\%$.

O comportamento observado nas Tabelas 4.2 e 4.3 também se verifica na Tabela 4.4, onde encontram-se os resultados de simulação realizados com o intuito de investigar o impacto nas taxas de rejeição ocasionado pelo aumento simultâneo do número de parâmetros de interesse e perturbação. Para tanto, considerou-se o modelo Passeio Aleatório com diversos valores de $p = q$, $n = 40$ e diversos níveis de significância. Destaca-se que novamente o teste baseado em LR mostrou-se bastante sensível ao aumento no número de parâmetros, enquanto que os testes baseados em LR_{pa} e LR_{pa}^c apresentaram maior estabilidade e taxas de rejeição empíricas mais próximas dos níveis desejados. Como exemplo, para $p = q = 4$ e $\alpha = 10,0\%$, as taxas de rejeição dos testes baseados nas estatísticas LR, LR_{pa} e LR_{pa}^c são, respectivamente, $19,4\%$, $12,4\%$ e $11,1\%$, enquanto que para um mesmo nível de significância, considerando $p = q = 7$ as taxas passam a ser $42,4\%$, $13,4\%$ e $10,0\%$, respectivamente.

A Tabela 4.5 apresenta os resultados de simulação para as taxas de rejeição dos testes sob hipótese alternativa (teste de poder). Uma vez que os testes de tamanho baseados nas estatísticas LR, LR_{pa} e LR_{pa}^c apresentaram considerável disparidade entre as taxas de rejeição para um dado nível nominal fixado, os testes de poder foram realizados utilizando valores críticos estimados para cada estatística, de modo que todos os testes pudessem ter o mesmo tamanho. Foi considerado o modelo Passeio Aleatório com $p = q = 3$, níveis de significância $\alpha = 5,0\%$ e $\alpha = 10,0\%$, $n = 40$ e a hipótese alternativa $H_1 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma^* \neq 0$, com γ^* variando de $0,1$ à $0,8$. De modo geral, observa-se que o comportamento dos testes baseados nas três estatísticas consideradas foram bastante semelhantes exibindo taxas de rejeição crescentes no sentido do distanciamento da hipótese nula e que o teste baseado na estatística usual apresenta poder superior aos demais testes. Além disso, observa-se que os testes baseados nas estatísticas LR_{pa} e LR_{pa}^c exibem poderes bastante semelhantes, concluindo-se que a introdução do fator de correção não acarretou em perda de poder para este teste. Como ilustração, para $\gamma^* = 0,4$ e $\alpha = 10,0\%$, as taxas de rejeição dos testes baseados nas estatísticas LR, LR_{pa} e LR_{pa}^c são, respectivamente, $74,3\%$, $59,8\%$ e $59,9\%$.

Em linhas gerais, observa-se que o teste baseado na estatística LR é bastante liberal. Esta tendência liberal é acentuada conforme aumentamos o número de parâmetros de interesse e/ou perturbação. Por outro lado, o teste baseado em LR_{pa} apresenta menor distorção de tamanho, com taxas de rejeição bem mais próximas dos níveis nominais que o teste usual e leve tendência liberal. A introdução do fator de correção de Bartlett proposto neste trabalho atenua esta tendência liberal, de modo que o teste baseado na estatística LR_{pa}^c apresenta taxas de rejeição empíricas ainda mais próximas dos níveis nominais fixados.

Tabela 4.1: Taxas de rejeição dos testes LR, LR_{pa}, e LR^c_{pa}, sob H_0, para o modelo Passeio Aleatório, com $p = 3, q = 6$ e $p = 4, q = 3$, vários valores de n e α

		$p = 3$ e $q = 6$			$p = 4$ e $q = 3$		
n	$\alpha(\%)$	LR	LR_{pa}	LR^c_{pa}	LR	LR_{pa}	LR^c_{pa}
	0,5	9,1	1,9	1,2	6,4	2,1	1,8
20	1,0	13,3	2,9	1,7	9,2	2,9	2,4
	5,0	29,0	8,7	5,3	22,1	8,2	7,2
	10,0	40,1	15,1	9,7	32,2	14,1	12,0
	0,5	3,4	1,6	1,1	2,7	1,5	1,3
30	1,0	5,6	2,6	1,7	4,3	2,4	2,1
	5,0	15,8	8,0	5,8	13,6	7,3	6,6
	10,0	25,3	13,8	10,3	22,0	13,3	11,9
	0,5	1,8	1,1	0,8	1,8	1,0	0,9
40	1,0	3,2	1,8	1,4	2,9	1,6	1,4
	5,0	11,5	7,1	5,3	10,7	6,5	5,8
	10,0	19,4	13,0	10,3	18,7	12,0	11,1
	0,5	1,6	1,0	0,7	1,4	0,9	0,8
50	1,0	2,6	1,7	1,3	2,3	1,5	1,4
	5,0	9,6	6,9	5,6	9,2	5,8	5,4
	10,0	16,5	12,4	10,4	16,0	11,2	10,5

Tabela 4.2: Taxas de rejeição dos testes LR, LR_{pa}, e LR_{pa}^c, sob H_0, para o modelo Passeio Aleatório, com $n = 40$, $q = 4$ e vários valores de α e p.

		$n = 40$ e $q = 4$		
$\alpha(\%)$	p	LR	LR_{pa}	LR_{pa}^c
	3	1,5	1,1	0,9
	4	2,3	1,1	0,9
0,5	5	2,9	1,2	1,0
	6	3,7	1,2	1,0
	7	4,9	1,1	0,9
	3	2,6	1,9	1,6
	4	3,7	1,9	1,5
1,0	5	4,4	1,9	1,6
	6	6,0	2,1	1,8
	7	7,7	2,0	1,5
	3	9,7	6,8	5,8
	4	11,5	7,0	6,0
5,0	5	13,5	6,7	5,8
	6	17,4	6,7	5,9
	7	21,1	6,9	6,0
	3	16,6	12,4	11,0
	4	19,4	12,4	11,1
10,0	5	21,9	11,9	10,7
	6	27,1	12,3	10,8
	7	31,6	12,1	10,9

Tabela 4.3: Taxas de rejeição dos testes LR, LR_{pa}, e LR_{pa}^c, sob H_0, para o modelo Passeio Aleatório, com $n = 40$, $p = 3$ vários valores de α e q.

		$n = 40$ e $p = 3$		
$\alpha(\%)$	q	LR	LR_{pa}	LR_{pa}^c
	3	1,3	1,0	0,9
	4	1,5	1,1	0,9
0,5	5	1,8	1,1	0,8
	6	1,8	1,1	0,8
	7	2,4	1,1	0,7
	3	2,3	1,8	1,5
	4	2,6	1,9	1,6
1,0	5	3,1	1,9	1,5
	6	3,2	1,8	1,4
	7	4,0	2,1	1,3
	3	8,5	6,8	6,1
	4	9,7	6,8	5,8
5,0	5	10,6	7,3	6,1
	6	11,5	7,1	5,3
	7	12,6	7,5	5,3
	3	15,0	12,0	11,2
	4	16,6	12,4	11,0
10,0	5	18,1	13,3	11,2
	6	19,4	13,0	10,3
	7	20,9	13,6	10,2

Tabela 4.4: Taxas de rejeição dos testes LR, LR_{pa}, e LR^c_{pa}, sob H_0, para o modelo Passeio Aleatório, com $n = 40$, vários valores de α e $p = q$.

		$n = 40$		
$\alpha(\%)$	$p = q$	LR	LR_{pa}	LR^c_{pa}
	3	1,3	1,0	0,9
	4	2,3	1,1	0,9
0,5	5	3,4	1,3	1,0
	6	6,2	1,3	0,9
	7	10,0	1,2	0,7
	3	2,3	1,8	1,5
	4	3,7	1,9	1,5
1,0	5	5,6	2,1	1,6
	6	9,2	2,0	1,6
	7	14,1	2,0	1,3
	3	8,5	6,8	6,1
	4	11,5	7,0	6,0
5,0	5	16,6	7,5	6,1
	6	22,6	7,3	5,5
	7	31,0	7,3	5,2
	3	15,0	12,0	11,2
	4	19,4	12,4	11,1
10,0	5	25,6	13,3	11,4
	6	32,9	13,3	10,5
	7	42,4	13,4	10,0

Tabela 4.5: Taxas de rejeição não-nulas dos testes LR, LR_{pa}, e LR^{c}_{pa}, para o modelo Passeio Aleatório, com $n = 40$, $\alpha = 5,0\%$ e $\alpha = 10,0\%$, $p = q = 3$.

		LR		LR_{pa}		LR^{c}_{pa}	
n	γ^*	$10,0\%$	$5,0\%$	10%	5%	10%	5%
	$0,1$	$16,7$	$9,0$	$11,4$	$5,6$	$11,5$	$5,7$
	$0,2$	$30,9$	$19,2$	$20,5$	$11,0$	$20,5$	$11,0$
	$0,3$	$51,4$	$35,8$	$36,4$	$23,5$	$36,4$	$23,6$
40	$0,4$	$74,3$	$59,4$	$59,8$	$43,0$	$59,9$	$43,1$
	$0,5$	$90,5$	$80,7$	$80,9$	$66,5$	$80,9$	$66,6$
	$0,6$	$97,5$	$93,8$	$93,8$	$86,2$	$93,8$	$86,3$
	$0,7$	$99,6$	$98,6$	$98,5$	$95,8$	$98,5$	$95,9$
	$0,8$	$100,0$	$99,8$	$99,8$	$99,1$	$99,8$	$99,1$

CAPÍTULO 5

Considerações Finais

Neste trabalho, discutimos o aperfeiçoamento de testes de hipóteses baseados na estatística da razão de verossimilhanças da classe de Modelos Lineares Generalizados com Superdispersão. A principal contribuição teórica encontra-se no Capítulo 3, onde obtivemos uma expressão em forma matricial do fator de correção de Bartlett, segundo a metodologia proposta por DiCiccio e Stern (1994), à estatística da razão de verossimilhanças perfiladas ajustadas proposta por Cox e Reid (1987) para o teste conjunto dos efeitos da dispersão nesta classe de modelos.

Através de estudos de simulação de Monte Carlo, comparamos o desempenho dos testes baseados na estatística da razão de verossimilhanças usual (LR), estatística da razão de verossimilhanças perfiladas ajustadas (LR_{pa}), além da respectiva versão aperfeiçoada com o fator de Bartlett desenvolvido neste trabalho, a saber, razão de verossimilhanças perfiladas ajustadas corrigida (LR_{pa}^c).

Os resultados de simulação de tamanho indicam que, de modo geral, o teste baseado na estatística LR é bastante liberal, conduzindo a taxas de rejeição da hipótese nula, sendo esta verdadeira, bem superiores aos níveis nominais fixados. Esta tendência liberal é acentuada conforme aumentamos o número de parâmetros de interesse e/ou perturbação.

O teste baseado na estatística LR_{pa}, de modo geral, apresenta taxas de rejeição empíricas bem próximas dos níveis nominais fixados, com leve tendência liberal. A introdução do fator de correção de Bartlett proposto neste trabalho melhorou a aproximação da distribuição desta estatística pela distribuição Qui-quadrado de referência, de modo que o teste baseado na estatística LR^c_{pa} apresenta taxas de rejeição empíricas ainda mais próximas dos níveis nominais fixados.

Com relação ao poder dos testes observou-se que o teste baseado na estatística LR apresentou maior poder, o que já era esperado uma vez que as simulações de tamanho evidenciaram que este teste é notavelmente liberal. Verificou-se ainda que testes baseados em LR_{pa} e LR^c_{pa} apresentaram desempenhos bastante semelhantes, o que indica que a introdução do fator de correção de Bartlett proposto não acarretou em perda de poder para este teste.

De modo geral, os resultados de simulação favorecem o uso da estatística corrigida proposta nesta dissertação, podendo, portanto, ser recomendada para testes de hipótese na classe de Modelos Lineares Generalizados com Superdispersão.

Bibliografia

[1] Araújo Júnior, C. A. G. (2006). Ajustes para verossimilhança perfilada na distribuição BirnBaum–Saunders. Dissertação de Mestrado, Programa de pós-graduação em Estatística, Universidade Federal de Pernambuco - *UFPE*.

[2] Araújo, M. C. (2012). Verossimilhança perfilada nos modelos não lineares simétricos heteroscedásticos. Dissertação de Mestrado, Programa de pós-graduação em Estatística, Universidade Federal de Pernambuco - *UFPE*.

[3] Barndorff-Nielsen, O. E. (1983). On a formula to the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

[4] Barndorff-Nielsen, O. E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society B*, **56**, 125–140.

[5] Barroso, L. P. e Cordeiro, G. M. (2005). Bartlett corrections in heteroskedastic t regression models. *Statistics & Probability Letters*, **75**, 86–96.

[6] Bartlett, M. S. (1937). Properties of suficiency and statistical tests. *Proceedings of the Royal Society of London A,* **160**, 268–282.

[7] Botter, D. A. e Cordeiro, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Journal of Statistical Computation and Simulation*, **62**, 91–104.

[8] Cordeiro, G. M. (1983). Improved likelihood ratio statistics for generalized linear models. *Jounal of the Royal Statistical Society B*, **45**, 404–413.

[9] Cordeiro, G. M. (1986). Modelos Lineares Generalizados. Livro texto de minicurso, VII Simpósio Nacional de Probabilidade e Estatística, UNICAMP, Campinas, SP.

[10] Cordeiro, G. M.; McCullagh, P. (1991). Bias correction in generalized linear models. *Jounal of the Royal Statistical Society B*, **53**, 629–643.

[11] Cordeiro, G.M. (1999). Introdução à Teoria Assintótica. *Colóquio Brasileiro de Matemática*, **22**, IMPA, Rio de Janeiro.

[12] Cordeiro, G.M. e Botter, D.A. (2001). Second-Order Biases of Maximun Likelihood Estimates in Overdispersed Generalized Linear Models. *Statistics and Probability Letters*, **55**, 269–280.

[13] Cordeiro, G. M.; Cysneiros, A. H. M. A.; Cysneiros, F. J. A. (2006). Bartlett adjustments for overdispersed generalised linear models. *Communications in Statistics - Theory and Methods* **35**, 937–952.

[14] Cordeiro, G. M. e Barroso, L. P. (2007). A third-order bias corrected estimate in generalized linear models. *Test*, **16**, 76–89.

[15] Cordeiro, G. M. e Demétrio, C. G. B (2010). Modelos Lineares Generalizados e Extensões. Disponível em: http://www4.esalq.usp.br/departamentos/lce/arquivos/aulas/2010/LCE5868/livro.pdf.

[16] Cox, D. R. & Hinkley, D. V. (1974). Theorical Statistics. *Chapman and Hall*, London.

[17] Cox, D.R. e Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**, 1–39.

[18] Cox, D. R. e Reid, N. (1992). A note on the difference between profile and modified profile likelihood. *Biometrika*, **79**, 408–411.

[19] Cribari-Neto, F. e Cordeiro, G. M. (1996). On Bartlett and Bartlett-type corrections. *Econometric reviews*, **15**, 339–367.

[20] Cysneiros, A. H. M. A. (2004) Refinamentos para Testes de Hipóteses em Modelos de Regressão Lineares e Não–Lineares Heteroscedásticos. Tese de Doutorado, Programa de pós-graduação em Estatística, Universidade de São Paulo- *USP*.

[21] Dey, D. K.; Gelfand, A. E.; Peng, F. (1997). Overdispersed generalized linear models.*Journal of Statistical Planning and Inferences* **64**, 93–107.

[22] Dey, D. K. e Ravishanker, N.(1998).Bayesian Approaches for Overdispersion in Generalized Linear Models (unpublished).

[23] DiCiccio, T. J. e Stern, S. E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological),* **56**, 397–408.

[24] Doornik, J. A. (2006). Ox: An Object-Oriented Matrix Programming Language. *Timberlake Consultants Ltd*, London, 4 ed.

[25] Efron, B. (1986). Double Exponencials Families and Their Use in Generalized Linear Regression. *Journal of the American Statistical Association*, **81**, 709–721.

[26] Ferrari, S.L.P. e Uribe–Opazo, M. A. (2001) . Corrected likelihood ratio tests in a class of symmetric linear regression models. *Brazilian Journal of Probability and Statistics*,**v.15**, p. 49–67.

[27] Ferrari, S. L. P.; Cysneiros, A. H. M. A.; Cribari-Neto, F. (2004). An improved test for heteroskedasticity using adjusted modified profile likelihood inference. *Journal of Statistical Planning and Inferences*, **124**, 423–437.

[28] Ferrari, S. L. P.; Lucambio, F.; Cribari-Neto, F. (2005). Improve Profile Likelihood Inference. *Journal of Statistical Planning and Inference,* **134**, 2, 373–391.

[29] Lawley, D. N. (1956). A general method for approximating to the distributionof the likelihood ratio criteria. *Biometrika*, **71**, 233–244.

[30] Lemonte , A.J.; Ferrari, S. L. P.; Cribari-Neto, F. (2010). Improved likelihood inference in Birnbaum-Saunders regressions. *Computational Statistics and Data Analysis*, **54**, 1307–1316.

[31] Lindsay, B. (1986). Exponential Family Mixture Models (with Least-Squares Estimators).*The Annals of Statistics*, **14**, 124–137.

[32] Silva (2005), M. F. Estimação e teste de hipótese baseados em verossimilhanças perfiladas. Tese de Doutorado, Programa de pós-graduação em Estatística, Universidade de São Paulo-*USP*.

[33] McCullagh, P. e Tibishirani, R. (1990). A simple method for the adjustment of profile likelihood. *Jounal of the Royal Statistical Society B*, **52**, 325–344.

[34] Nelder, J.A. e Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A*, **135**, 370–384.

[35] Nocedal, J. e Wright, S. J. (1999). Numerical Optimization. *Springer Verlag*, New York.

[36] Paula, G. A. (2012). *Modelos de Regressão com apoio computacional.* São Paulo: Instituto de Matemática e Estatística - USP.

[37] Previdelli, I.T.S. (2005). Estimadores corrigidos para modelos não-lineares generalizados superdispersados. Tese de Doutorado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina.

[38] Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Jounal of the Royal Statistical Society B* **51**, 47–60.

[39] Stern, S. E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Jounal of the Royal Statistical Society B*, **59**, 653–665.

APÊNDICE A

Cálculo dos Cumulantes

Considere a classe de Modelos Lineares Generalizados com Superdispersão apresentada no Capítulo 2 deste trabalho. Neste Apêndice, serão apresentadas algumas derivadas do logaritmo da função de verossimilhança, cumulantes e derivadas de cumulantes para esta classe de modelos.

A.1 Derivadas do logaritmo da função de verossimilhança

Diferenciando o logaritmo da função de verossimilhança definido em (2.4) com relação aos componentes do vetor de parâmetros $\boldsymbol{\beta}$ do modelo, obtemos:

$$U_r = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r} = \sum_{l=1}^{n} [\psi_l^{(2,0)}(\mu_l, \phi_l) m_{1l}(y_l - \mu_l)(r)_l],$$

$$\begin{aligned} U_{rs} = \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r \partial \beta_s} &= \sum_{l=1}^{n} [\psi_l^{(3,0)}(\mu_l, \phi_l) m_{1l}^2 (y_l - \mu_l) + \psi_l^{(2,0)}(\mu_l, \phi_l) m_{2l}(y_l - \mu_l) \\ &- \psi_l^{(2,0)}(\mu_l, \phi_l) m_{1l}^2](r)_l(s)_l, \end{aligned}$$

$$
\begin{aligned}
U_{rst} &= \frac{\partial^3 \ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\beta_r\partial\beta_s\partial\beta_t}\\
&= \sum_{l=1}^{n}[\psi_l^{(4,0)}(\mu_l,\phi_l)m_{1l}^3(y_l-\mu_l)+3\psi_l^{(3,0)}(\mu_l,\phi_l)m_{1l}m_{2l}(y_l-\mu_l)\\
&- 2\psi_l^{(3,0)}(\mu_l,\phi_l)m_{1l}^3-3\psi_l^{(2,0)}(\mu_l,\phi_l)m_{1l}m_{2l}\\
&+ \psi_l^{(2,0)}(\mu_l,\phi_l)m_{3l}(y_l-\mu_l)](r)_l(s)_l(t)_l
\end{aligned}
$$

e

$$
\begin{aligned}
U_{rstu} &= \frac{\partial^4 \ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\beta_r\partial\beta_s\partial\beta_t\partial\beta_u}\\
&= \sum_{l=1}^{n}[\psi_l^{(5,0)}(\mu_l,\phi_l)m_{1l}^4(y_l-\mu_l)+6\psi_l^{(4,0)}(\mu_l,\phi_l)m_{1l}^2m_{2l}(y_l-\mu_l)\\
&- 3\psi_l^{(4,0)}(\mu_l,\phi_l)m_{1l}^4+3\psi_l^{(3,0)}(\mu_l,\phi_l)m_{2l}^2(y_l-\mu_l)\\
&+ 4\psi_l^{(3,0)}(\mu_l,\phi_l)m_{1l}m_{3l}(y_l-\mu_l)-12\psi_l^{(3,0)}(\mu_l,\phi_l)m_{1l}^2m_{2l}\\
&- 3\psi_l^{(2,0)}(\mu_l,\phi_l)m_{2l}^2-4\psi_l^{(2,0)}(\mu_l,\phi_l)m_{1l}m_{3l}\\
&+ \psi_l^{(2,0)}(\mu_l,\phi_l)m_{4l}(y_l-\mu_l)](r)_l(s)_l(t)_l(u)_l.
\end{aligned}
$$

Diferenciando o logaritmo da função de verossimilhança definido em (2.4) com relação aos componentes do vetor de parâmetros $\boldsymbol{\gamma}$ do modelo, obtemos:

$$
U_R = \frac{\partial \ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\gamma_R} = \sum_{l=1}^{n}[\psi_l^{(1,1)}(\mu_l,\phi_l)(y_l-\mu_l)+\mathrm{T}(y_l)+\psi_l^{(0,1)}(\mu_l,\phi_l)]\phi_{1l}(R)_l,
$$

$$
\begin{aligned}
U_{RS} = \frac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\gamma_R\partial\gamma_S} &= \sum_{l=1}^{n}[\psi_l^{(1,2)}(\mu_l,\phi_l)\phi_{1l}^2(y_l-\mu_l)+\psi_l^{(1,1)}(\mu_l,\phi_l)\phi_{2l}(y_l-\mu_l)\\
&+ \phi_{2l}\mathrm{T}(y_l)+\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}^2+\psi_l^{(0,1)}(\mu_l,\phi_l)\phi_{2l}](R)_l(S)_l,
\end{aligned}
$$

$$
\begin{aligned}
U_{RST} &= \frac{\partial^3 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \gamma_R \partial \gamma_S \partial \gamma_T} \\
&= \sum_{l=1}^{n} [\psi_l^{(1,3)}(\mu_l, \phi_l)\phi_{1l}^3(y_l - \mu_l) + 3\psi_l^{(1,2)}(\mu_l, \phi_l)\phi_{1l}\phi_{2l}(y_l - \mu_l) \\
&+ \psi_l^{(1,1)}(\mu_l, \phi_l)\phi_{3l}(y_l - \mu_l) + \phi_{3l}\mathrm{T}(y_l) + \psi_l^{(0,3)}(\mu_l, \phi_l)\phi_{1l}^3 \\
&+ 3\psi_l^{(0,2)}(\mu_l, \phi_l)\phi_{1l}\phi_{2l} + \psi_l^{(0,1)}(\mu_l, \phi_l)\phi_{3l}](R)_l(S)_l(T)_l
\end{aligned}
$$

e

$$
\begin{aligned}
U_{RSTU} &= \frac{\partial^4 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \gamma_R \partial \gamma_S \partial \gamma_T \partial \gamma_U} \\
&= \sum_{l=1}^{n} [\psi_l^{(1,4)}(\mu_l, \phi_l)\phi_{1l}^4(y_l - \mu_l) + 6\psi_l^{(1,3)}(\mu_l, \phi_l)\phi_{1l}^2\phi_{2l}(y_l - \mu_l) \\
&+ 3\psi_l^{(1,2)}(\mu_l, \phi_l)\phi_{2l}^2(y_l - \mu_l) + 4\psi_l^{(1,2)}(\mu_l, \phi_l)\phi_{1l}\phi_{3l}(y_l - \mu_l) \\
&+ \psi_l^{(1,1)}(\mu_l, \phi_l)\phi_{4l}(y_l - \mu_l) + \phi_{4l}\mathrm{T}(y_l) + \psi_l^{(0,4)}(\mu_l, \phi_l)\phi_{1l}^4 \\
&+ 6\psi_l^{(0,3)}(\mu_l, \phi_l)\phi_{1l}^2\phi_{2l} + 3\psi_l^{(0,2)}(\mu_l, \phi_l)\phi_{2l}^2 + 4\psi_l^{(0,2)}(\mu_l, \phi_l)\phi_{1l}\phi_{3l} \\
&+ \psi_l^{(0,1)}(\mu_l, \phi_l)\phi_{4l}](R)_l(S)_l(T)_l(U)_l.
\end{aligned}
$$

A seguir, são obtidas algumas derivadas mistas, que são necessárias para o cálculo de cumulantes conjuntos dos vetores de parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ do modelo.

$$
U_{rS} = \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r \partial \gamma_S} = \sum_{l=1}^{n} [\psi_l^{(2,1)}(\mu_l, \phi_l)\phi_{1l} m_{1l}(y_l - \mu_l)](r)_l(S)_l,
$$

$$
\begin{aligned}
U_{rST} &= \frac{\partial^3 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r \partial \gamma_S \partial \gamma_T} \\
&= \sum_{l=1}^{n} [\psi_l^{(2,2)}(\mu_l, \phi_l)\phi_{1l}^2 m_{1l}(y_l - \mu_l) \\
&+ \psi_l^{(2,1)}(\mu_l, \phi_l)\phi_{2l} m_{1l}(y_l - \mu_l)](r)_l(S)_l(T)_l,
\end{aligned}
$$

$$\begin{aligned}
U_{rsT} &= \frac{\partial^3 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r \partial \beta_s \partial \gamma_T} \\
&= \sum_{l=1}^{n} [\psi_l^{(3,1)}(\mu_l, \phi_l)\phi_{1l} m_{1l}^2 (y_l - \mu_l) \\
&+ \psi_l^{(2,1)}(\mu_l, \phi_l)\phi_{1l} m_{2l} (y_l - \mu_l) \\
&- \psi_l^{(2,1)}(\mu_l, \phi_l)\phi_{1l} m_{1l}^2](r)_l (s)_l (T)_l
\end{aligned}$$

e

$$\begin{aligned}
U_{rsTU} &= \frac{\partial^4 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r \partial \beta_s \partial \gamma_T \partial \gamma_U} \\
&= \sum_{l=1}^{n} [\psi_l^{(3,2)}(\mu_l, \phi_l)\phi_{1l}^2 m_{1l}^2 (y_l - \mu_l) \\
&+ \psi_l^{(3,1)}(\mu_l, \phi_l)\phi_{2l} m_{1l}^2 (y_l - \mu_l) \\
&+ \psi_l^{(2,2)}(\mu_l, \phi_l)\phi_{1l}^2 m_{2l} (y_l - \mu_l) \\
&+ \psi_l^{(2,1)}(\mu_l, \phi_l)\phi_{2l} m_{2l} (y_l - \mu_l) \\
&- \psi_l^{(2,2)}(\mu_l, \phi_l)\phi_{1l}^2 m_{1l}^2 \\
&- \psi_l^{(2,1)}(\mu_l, \phi_l)\phi_{2l} m_{1l}^2](r)_l (s)_l (T)_l (U)_l.
\end{aligned}$$

A.2 Cálculo dos Cumulantes

Calculando o valor esperado das derivadas acima, obtêm-se os seguintes cumulantes:

$$\kappa_{rs} = -\sum_{l=1}^{n} [\psi_l^{(2,0)}(\mu_l, \phi_l) m_{1l}^2](r)_l (s)_l,$$

$$\kappa_{rst} = -\sum_{l=1}^{n} [2\psi_l^{(3,0)}(\mu_l, \phi_l) m_{1l}^3 + 3\psi_l^{(2,0)}(\mu_l, \phi_l) m_{1l} m_{2l}](r)_l (s)_l (t)_l,$$

$$
\begin{aligned}
\kappa_{rstu} = & - \sum_{l=1}^{n}[3\psi_l^{(4,0)}(\mu_l,\phi_l)m_{1l}^4 + 12\psi_l^{(3,0)}(\mu_l,\phi_l)m_{1l}^2 m_{2l} \\
& + 3\psi_l^{(2,0)}(\mu_l,\phi_l)m_{2l}^2 + 4\psi_l^{(2,0)}(\mu_l,\phi_l)m_{1l}m_{3l}](r)_l(s)_l(t)_l(u)_l,
\end{aligned}
$$

$$
\kappa_{RS} = \sum_{l=1}^{n}[\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}^2](R)_l(S)_l,
$$

$$
\kappa_{RST} = \sum_{l=1}^{n}[\psi_l^{(0,3)}(\mu_l,\phi_l)\phi_{1l}^3 + 3\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}\phi_{2l}](R)_l(S)_l(T)_l,
$$

$$
\begin{aligned}
\kappa_{RSTU} = & \sum_{l=1}^{n}[\psi_l^{(0,4)}(\mu_l,\phi_l)\phi_{1l}^4 + 6\psi_l^{(0,3)}(\mu_l,\phi_l)\phi_{1l}^2\phi_{2l} \\
& + 3\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{2l}^2 + 4\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}\phi_{3l}](R)_l(S)_l(T)_l(U)_l,
\end{aligned}
$$

$$
\kappa_{rS} = 0, \;\; \kappa_{rST} = 0, \;\; \kappa_{rsT} = -\sum_{l=1}^{n}[\psi_l^{(2,1)}(\mu_l,\phi_l)\phi_{1l}m_{1l}^2](r)_l(s)_l(T)_l,
$$

e

$$
\kappa_{rsTU} = -\sum_{l=1}^{n}[\psi_l^{(2,2)}(\mu_l,\phi_l)\phi_{1l}^2 m_{1l}^2 + \psi_l^{(2,1)}(\mu_l,\phi_l)\phi_{2l}m_{1l}^2](r)_l(s)_l(T)_l(U)_l.
$$

A.3 Derivadas de Cumulantes

A seguir, são apresentadas algumas derivadas de cumulantes com relação aos componentes dos vetores de parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ do modelo.

$$
\kappa_{RS}^{(T)} = \sum_{l=1}^{n}[\psi_l^{(0,3)}(\mu_l,\phi_l)\phi_{1l}^3 + 2\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}\phi_{2l}](R)_l(S)_l(T)_l,
$$

$$
\begin{aligned}
\kappa_{RS}^{(TU)} &= \sum_{l=1}^{n}[\psi_l^{(0,4)}(\mu_l,\phi_l)\phi_{1l}^4 + 5\psi_l^{(0,3)}(\mu_l,\phi_l)\phi_{1l}^2\phi_{2l} \\
&+ 2\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{2l}^2 + 2\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}\phi_{3l}](R)_l(S)_l(T)_l(U)_l,
\end{aligned}
$$

$$
\begin{aligned}
\kappa_{RST}^{(U)} &= \sum_{l=1}^{n}[\psi_l^{(0,4)}(\mu_l,\phi_l)\phi_{1l}^4 + 6\psi_l^{(0,3)}(\mu_l,\phi_l)\phi_{1l}^2\phi_{2l} \\
&+ 3\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{2l}^2 + 3\psi_l^{(0,2)}(\mu_l,\phi_l)\phi_{1l}\phi_{3l}](R)_l(S)_l(T)_l(U)_l
\end{aligned}
$$

e

$$\kappa_{rST}^{(u)} = 0.$$