

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

TÉCNICAS DE DIAGNÓSTICO NOS MODELOS LINEARES
GENERALIZADOS COM SUPERDISPERSÃO

HELOISA DE MELO RODRIGUES

Orientadora: Prof^a Dr^a Audrey Helen Mariz de Aquino Cysneiros

Área de concentração: Estatística Aplicada

Recife, fevereiro de 2013

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

TÉCNICAS DE DIAGNÓSTICO NOS MODELOS LINEARES
GENERALIZADOS COM SUPERDISPERSÃO

HELOISA DE MELO RODRIGUES

Orientadora: Prof^a Dr^a Audrey Helen Mariz de Aquino Cysneiros

Área de concentração: Estatística Aplicada

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

Recife, fevereiro de 2013

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

Rodrigues, Heloisa de Melo

Técnicas de diagnóstico nos modelos lineares generalizados com superdispersão / Heloisa de Melo Rodrigues. - Recife: O Autor, 2013.

x, 64f.: il., fig., tab.

Orientadora: Audrey Helen Mariz de Aquino Cysneiros.

Dissertação (mestrado) - Universidade Federal de Pernambuco. CCEN. Estatística, 2013.

Inclui bibliografia e apêndice.

1. Estatística aplicada - Modelagem. 2. Análise de regressão. 3. Modelos lineares (estatística) I. Cysneiros, Audrey Helen Mariz de Aquino (orientadora). II. Título.

310

CDD (23. ed.)

MEI2013 – 027

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

Recife, 07 de fevereiro de 2013.

Nós recomendamos que a dissertação de mestrado de autoria de

Heloísa de Melo Rodrigues

Intitulada

“Técnicas de Diagnóstico 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

Lizandra Castilho Fabio USP

Este documento será anexado à versão final da tese.

Dedico este trabalho à minha família e à todos que me ajudaram a chegar até aqui.

Agradecimentos

À professora Audrey Cysneiros, pela orientação, confiança, paciência e dedicação que teve comigo durante a elaboração deste trabalho. Valeu o desafio! Agradeço-a também pelos ensinamentos em todo meu percurso na UFPE até o momento, desde a época da graduação, em que foi minha professora logo no primeiro período do curso, na disciplina Análise Exploratória de Dados, em 2005, e em seguida nas disciplinas de Probabilidade 2 e Inferência Estatística. De lá pra cá, a amizade e relação profissional só fez crescer.

Aos meus pais Fátima e Paulo Cezar, por sempre acreditarem em mim e por me dar muito amor e carinho, essencial para enfrentar qualquer desafio que a vida impõe.

Aos meus avós Eremita e José Rodrigues (*in memoriam*) pelo carinho e apoio.

Aos meus tios Ana, Anália e Manoel e aos meus primos Luciana, Luan, Marília, Letícia e Antônio por sempre estarem presente em minha vida, acompanhando de perto minha trajetória.

Ao meu namorado Felipe por sempre estar ao meu lado, apoiando minhas decisões e muitas vezes me ajudando a fazer escolhas. Obrigada pelo incentivo, amor e companheirismo.

À Socorro, Rafael, Natália, Bruno, Fernanda e Rodrigo pelo convívio e torcida.

Aos amigos e colegas de graduação, especialmente Nathália, Simara, Hélio, Berg, Samuel, Léo e Polly pela amizade e boa convivência.

Aos amigos do mestrado Gustavo, Mariana, Vitor, Sérgio, Marcos, Mileno e um agradecimento especial para Ana Hermínia, Sadraque, Thiago e Diêgo, pelas incansáveis horas de estudo e boas risadas.

À secretária da pós-graduação Valéria Bittencourt, pela simpatia, amizade e disponibilidade de sempre.

Aos amigos Dállete, Fabiana, Rafaella, Yuri, Duda e Annick pelo apoio e amizade.

Aos professores do Departamento de Estatística da UFPE Alex, Andre Toom, Carla Monteiro, Claudia Lima, Cristina Raposo, Cristiano Ferraz, Francisco Cribari, Getúlio Amaral, Isaac Xavier, Leandro Rêgo, Raydonal, Renato Cintra e Sylvio. Obrigada pela contribuição na minha formação acadêmica, tanto na graduação quanto no mestrado. Um agradecimento especial para o professor Francisco Cysneiros, pelos conselhos e sermões, desde a graduação, e pela disposição em compartilhar seus conhecimentos nas cinco disciplinas em que fui sua aluna.

À CAPES pelo apoio financeiro.

*"Falar sem aspas, amar sem interrogação, sonhar com reticências, viver sem
ponto final."
Charles Chaplin*

Resumo

No contexto de modelos de regressão, em alguns casos é comum o fenômeno da superdispersão, que ocorre quando a variância observada dos dados excede aquela prevista por um modelo. Assim, Dey et al. (1997) desenvolveram os modelos lineares generalizados com superdispersão (MLGSs), considerando um modelo de regressão adicional para o parâmetro de dispersão, que é incorporado na função de variância. Desta forma, os MLGSs permitem modelar, simultaneamente, a média e a dispersão no contexto dos modelos lineares generalizados (MLGs) de Nelder e Wedderburn, 1972. Além disso, os MLGSs caracterizam-se por ser uma classe de modelos mais geral que os modelos lineares generalizados duplos (Smyth, 1989). Nesta dissertação são propostas técnicas de diagnósticos para os MLGSs, sendo desenvolvidas as técnicas de alavancagem generalizada, análise de resíduos, influência global, como também o método de influência local, este avaliado sob três esquemas de perturbação. Por fim, é apresentada uma análise gráfica por meio de dados simulados.

Palavras chave: Alavancagem generalizada, Influência global, Influência local, Modelos lineares generalizados com superdispersão, Resíduos.

Abstract

The phenomenon of overdispersion, which is common in some cases of regression models, occurs when the variance of the response variable exceeds the nominal variance predicted by standard generalized linear models (GLMs). With this problem in mind, Dey et al. (1997) developed the overdispersed generalized linear models (OGLMs), considering an additional regression model for a scale parameter which is incorporated in the variance function. That is, the OGLMs allow the mean and dispersion to be modeled simultaneously in a generalized linear model context developed by Nelder and Wedderburn, 1972. Furthermore, the OGLMs are considered a more general model class than the double generalized linear models (Smyth, 1989). In this dissertation, diagnostic methods for the OGLMs are proposed, along with the development of techniques of generalized leverage, analysis of residuals, global influence and local influence, on which three perturbation schemes are used. At last, a graphical analysis through simulated data is used to illustrate the theories presented.

Keywords: Generalized leverage, Global influence, Local influence, Overdispersed generalized linear models, Residuals.

Lista de Figuras

4.1	Gráficos de pontos de alavanca generalizados contra o índice das observações e os valores ajustados para os dados simulados da distribuição passeio aleatório para a média ((a) e (b)) e a dispersão ((c) e (d)).	42
4.2	Gráficos de resíduos componente do desvio contra o índice das observações e os valores ajustados para os dados simulados da distribuição passeio aleatório.	43
4.3	Gráficos de resíduos componente do desvio padronizado contra o índice das observações e os valores ajustados para os dados simulados da distribuição passeio aleatório.	43
4.4	Gráficos de Influência global para os dados simulados da distribuição passeio aleatório para a média ((a) e (b)) e a dispersão ((c) e (d)).	44
4.5	Gráficos de C_i contra o índice das observações para os esquemas de perturbação (a) perturbação aditiva na variável resposta, (b) perturbação aditiva no preditor para o ajuste da média (matriz \mathbf{X}), (c) perturbação aditiva no preditor para o ajuste da dispersão (matriz \mathbf{S}) e (d) perturbação de casos ponderados, referente aos dados simulados da distribuição passeio aleatório.	45

Sumário

1	Introdução	1
1.1	Introdução	1
1.2	Objetivos	2
1.3	Estrutura da Dissertação	2
1.4	Suporte Computacional	3
2	Modelos Lineares Generalizados com Superdispersão	4
2.1	Introdução	4
2.2	Superdispersão	4
2.3	Definição	5
2.4	Algumas distribuições pertencentes à classe dos MLGSs	7
2.5	Aspectos inferenciais	11
2.5.1	Função Escore	11
2.5.2	Informação de Fisher	12
2.5.3	Estimação dos parâmetros	13
3	Técnicas de Diagnóstico nos MLGSs	16
3.1	Alavancagem Generalizada	17
3.2	Resíduos	21
3.3	Influência Global	25
3.4	Influência Local	28
3.4.1	Perturbação aditiva na resposta	30
3.4.2	Perturbação de casos ponderados	31
3.4.3	Perturbação aditiva no preditor	33

4	Resultados de Simulação	39
4.1	Análise Gráfica	40
5	Considerações Finais	46
	Bibliografia	47
A	Cálculo dos Momentos	50
A.1	Função Escore	50
A.2	Informação de Fisher	51
B	Passeio Aleatório	54
B.1	Derivadas da função $\psi(\mu, \phi)$	54
B.2	Obtenção das matrizes $\mathbf{M}_1, \mathbf{M}_2, \Phi_1$ e Φ_2	55
C	Alavancagem Generalizada	57
C.1	Cálculos para $\mathbf{GL}^\mu(\hat{\theta})$	57
C.2	Cálculos para $\mathbf{GL}^\phi(\hat{\theta})$	59
D	Influência Global	61

Capítulo 1

Introdução

1.1 Introdução

Durante muito tempo, os modelos normais lineares foram utilizados para descrever a maioria dos fenômenos aleatórios. Com o avanço da teoria dos modelos de regressão, Nelder e Wedderburn (1972) propuseram os modelos lineares generalizados (MLGs), em que a ideia básica destes modelos consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear η , por meio da função de ligação, esta caracterizada por ser uma função conhecida, monótona e diferenciável.

No contexto de modelos de regressão, em alguns casos é comum o fenômeno da superdispersão, cuja ocorrência é caracterizada quando a variação observada dos dados excede aquela prevista por um MLG, particularmente em relação às distribuições binomial e de Poisson. Assim, Dey et al. (1997) definiram a classe dos modelos lineares generalizados com superdispersão (MLGSs), adicionando um modelo de regressão para o parâmetro de dispersão, que é incorporado na função de variância, de modo que a superdispersão é modelada por meio de uma componente sistemática. Ou seja, os MLGSs permitem que a média e a dispersão sejam modeladas simultaneamente em um contexto dos modelos lineares generalizados. Desta forma, os MLGSs possuem duas componentes sistemáticas, uma para modelar a média e outra para a dispersão.

Na literatura, alguns estudos encontrados acerca dos MLGSs são: Cordeiro e Botter (2001), que apresentaram as fórmulas para os cálculos dos vieses para os estimadores de máxima verossimilhança dos parâmetros, e Cordeiro

et al. (2006), que derivaram correção de Bartlett para os MLGSs. Ainda deve ser ressaltado que não há estudos sobre técnicas de diagnóstico dos MLGSs na literatura. Sendo assim, o presente trabalho vem preencher esta lacuna, tornando-se pioneiro em explorar tal área nessa classe de modelos.

1.2 Objetivos

Nesta dissertação são apresentados os aspectos inferenciais dos modelos lineares generalizados com superdispersão, assim como técnicas de diagnóstico do modelo estudado. Como objetivos específicos, pode-se destacar:

- Aplicar a metodologia de alavancagem generalizada aos MLGSs;
- Propor resíduos para a verificação das suposições do modelo, como também para o estudo de observações atípicas;
- Apresentar medidas de influência global para os MLGSs;
- Desenvolver a metodologia de influência local nos MLGSs para avaliar as estimativas de máxima verossimilhança, considerando diferentes esquemas de perturbação.

1.3 Estrutura da Dissertação

Esta dissertação encontra-se dividida em cinco capítulos. No Capítulo 2 é apresentado um conceito geral de superdispersão, com suas possíveis causas e consequências. Além disso, são definidos os modelos lineares generalizados com superdispersão desenvolvidos por Dey et al. (1997), algumas distribuições pertencentes à esta classe de modelos, assim como seus aspectos inferenciais, tais como função escore, informação de Fisher e estimação dos parâmetros. No Capítulo 3 encontram-se as principais contribuições teóricas desta dissertação, em que são desenvolvidas técnicas de diagnósticos para os MLGSs, tais como: alavancagem generalizada, resíduos, influência global e influência local. No Capítulo 4 é realizada uma análise gráfica com dados simulados com o intuito de apresentar as técnicas de diagnóstico desenvolvidas nesta dissertação. Por último, no Capítulo 5 são feitas as considerações finais deste trabalho.

1.4 Suporte Computacional

No Capítulo 4, referente à análise gráfica via dados simulados, todos os cálculos foram realizados a partir da linguagem matricial de programação `Ox` (versão 6.20) para o sistema operacional Windows. `Ox` é uma linguagem com orientação a objetos, desenvolvida por Jurgen Doornick. Uma das maiores vantagens dessa linguagem é possuir ampla biblioteca numérica, tornando-se muito útil para a área de computação científica. Para fins acadêmicos, `Ox` pode ser obtido gratuitamente, em sua versão Console, por meio do link <http://www.doornick.com>, em que é possível encontrá-lo disponível para download em diferentes plataformas computacionais. Mais informações, ver Doornick (2009).

Para a confecção de gráficos foi utilizado o software gratuito `R` em sua versão 2.15.1 para o sistema operacional Windows, que pode ser obtido no endereço <http://www.r-project.org>. O software `R` é um ambiente para computação científica e gráfica baseado no software `S`, sendo amplamente utilizado por pesquisadores de diversas áreas do conhecimento.

Por fim, acrescenta-se que toda a dissertação foi redigida por meio do `TeXnicCenter` do sistema tipográfico `LATEX`, que consiste em uma série de macros e rotinas baseadas no sistema `TEX`, criado por Donald Knuth. Detalhes sobre o sistema de tipografia `LATEX` podem ser encontrados em Mittelbach et al. (2004), como também por meio do site <http://www.tex.ac.uk/CTAN/latex>.

Capítulo 2

Modelos Lineares Generalizados com Superdispersão

2.1 Introdução

Os modelos lineares generalizados (MLGs) têm sido uma classe padrão de modelos para análise de dados. Contudo, em algumas aplicações, a presença de heterogeneidade em amostras é muito grande para ser explicada pela função de variância simples implícita nos modelos examinados. Utilizando uma família exponencial biparamétrica, que é superdispersada em relação a uma específica família exponencial uniparamétrica, é possível a criação de classes de modelos lineares generalizados com superdispersão (MLGSs), que são analiticamente atraentes, estes definidos por Dey et al. (1997). Cordeiro e Botter (2001) derivaram fórmulas gerais para os vieses de segunda ordem das estimativas de máxima verossimilhança dos MLGSs e Cordeiro et al. (2006) desenvolveram correção de Bartlett para tal classe de modelos.

2.2 Superdispersão

Na aplicação de modelos lineares generalizados, principalmente em modelagem de dados binários agrupados e dados de contagem, como por exemplo as distribuições binomial e Poisson em que $\phi = 1$, para um bom ajuste do modelo é esperado que o desvio seja aproximadamente igual aos graus de liberdade. Se o desvio é maior que os graus de liberdade, há dois cenários possíveis para se considerar:

- (i) Há simplesmente um mau ajuste do modelo, dentre algumas razões como: termos ou variáveis omitidas no preditor linear; relação incorreta entre

a média e as variáveis explicativas, isto é, pode haver um erro na função de ligação ou uma necessidade de transformar uma ou mais variáveis explicativas; e ainda, a presença de outliers.

- (ii) A variação observada pode exceder aquela predita pelo modelo e este fenômeno é denominado **superdispersão**.

Segundo Paula (2012), diferentes circunstâncias podem causar um valor alto para o desvio. Algumas delas representam uma superdispersão aparente. Por exemplo, alguns pontos aberrantes podem aumentar substancialmente o valor do desvio e a simples eliminação desses pontos pode reduzir as evidências de superdispersão. Outra causa aparente de superdispersão é a ausência de algum termo extra na componente sistemática do modelo.

Hinde e Demétrio (1998) apontam diferentes causas possíveis de superdispersão, tais como: variabilidade do experimento, correlação entre respostas individuais, amostragem por cluster, agregação em dados de nível, ou até mesmo a omissão de variáveis não observadas. Em algumas circunstâncias, a causa da superdispersão pode ser aparente a partir da natureza do processo de coleta dos dados, embora diferentes causas do fenômeno da superdispersão possam levar a um mesmo modelo, tornando-se difícil inferir a causa precisa que conduza à superdispersão. Com isso, torna-se necessário estudar medidas de diagnóstico, que são ferramentas importantes para detectar o fenômeno.

Caso seja identificada a presença de superdispersão e esta não seja levada em consideração nas análises, tal fato poderá acarretar em consequências prejudiciais ao modelo. Inicialmente, os erros padrões obtidos do modelo estarão incorretos e podem ser seriamente subestimados e conseqüentemente pode-se avaliar incorretamente a significância individual dos parâmetros de regressão. Ainda, as mudanças no desvio associado com os termos do modelo irão também ser muito grande e isto conduzirá à seleção de modelos extremamente complexos. Finalmente, a interpretação do modelo será incorreta, podendo prejudicar quaisquer previsões.

2.3 Definição

Dey et al. (1997) definiram uma classe de modelos lineares generalizados com superdispersão (MLGSs) em que as variáveis aleatórias Y_1, \dots, Y_n são

independentes e cada Y_ℓ tem função densidade (ou de probabilidade) na família exponencial bidimensional de distribuições dada por

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

em que $A(\cdot)$, $T(\cdot)$ e $\psi(\cdot, \cdot)$ são funções conhecidas. Além disso, seja $\psi^{(r,s)}(\mu, \phi) = \partial^{r+s}\psi(\mu, \phi)/\partial\mu^r\partial\phi^s$ ($r, s \geq 0$). A média e a variância de Y são dadas, respectivamente, por

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \psi^{(2,0)^{-1}}.$$

E, ainda,

$$E\{T(Y)\} = -\psi^{(0,1)} \quad \text{e} \quad \text{Var}\{T(Y)\} = \psi^{(1,1)^2}\psi^{(2,0)^{-1}} - \psi^{(0,2)}. \quad (2.2)$$

Tem-se também que $E\{(Y - \mu)T(Y)\} = -\psi^{(1,1)}\psi^{(2,0)^{-1}}$.

Para esclarecimento de notação, vale ressaltar que para $r = 1$ e $s = 0$ tem-se $\psi^{(1,0)}$, que corresponde à primeira derivada da função $\psi(\mu, \phi)$ em relação a μ e $\psi^{(0,1)}$ ($r = 0$ e $s = 1$) equivale à primeira derivada da função $\psi(\mu, \phi)$ em relação a ϕ , seguindo o mesmo raciocínio para as demais ordens de derivação de μ e ϕ .

É fácil mostrar que $E\left(\frac{\partial^2 \log \pi}{\partial \mu \partial \phi}\right) = 0$, isto é, que os parâmetros μ e ϕ são ortogonais, conforme Cox e Reid (1987). Gelfand e Dalal (1990) mostram que, se $T(Y)$ é convexa e (2.1) é integrável em y , mantendo a média μ fixa, então a variância de Y será acrescida em ϕ . Ou seja, se ϕ cresce, a $\text{Var}(Y)$ também sofrerá um aumento. Portanto, ϕ pode ser considerado um parâmetro de dispersão.

Os MLGSs são definidos por (2.1) e pelas componentes sistemáticas $f(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ e $g(\boldsymbol{\phi}) = \boldsymbol{\tau} = \mathbf{S}\boldsymbol{\gamma}$, em que \mathbf{X} e \mathbf{S} são matrizes $n \times p$ e $n \times q$ de postos completos p e q , respectivamente, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top$ são vetores de parâmetros desconhecidos a serem estimados, $f(\cdot)$ e $g(\cdot)$ são funções monótonas contínuas conhecidas, pelo menos duas vezes diferenciáveis, denominadas funções de ligação da média e dispersão, respectivamente. Para ajustar um MLGS, é necessário uma escolha adequada destas funções de ligação. Cordeiro e Botter (2001) e Cordeiro et al. (2006) sugerem, por exemplo, escolher $g(\boldsymbol{\phi}) = \log \boldsymbol{\phi}$.

Para um dado MLGS há o interesse de se estimar, simultaneamente, os parâmetros β e γ , desde que eles representem os efeitos das variáveis explicativas sobre a resposta média e o parâmetro de dispersão, respectivamente. Seja y_1, \dots, y_n valores amostrais das respectivas variáveis aleatórias Y_1, \dots, Y_n , então o logaritmo da função de verossimilhança é dado por

$$\ell(\beta, \gamma) = \sum_{\ell=1}^n \{(y_\ell - \mu_\ell)\psi^{(1,0)}(\mu_\ell, \phi_\ell) + \phi_\ell T(y_\ell) + \psi(\mu_\ell, \phi_\ell)\} + \sum_{\ell=1}^n \log A(y_\ell). \quad (2.3)$$

2.4 Algumas distribuições pertencentes à classe dos MLGSs

Nesta Seção são apresentadas algumas distribuições de probabilidade pertencentes à classe dos modelos lineares generalizados com superdispersão. Na sequência, a distribuição passeio aleatório foi descrita com maiores detalhes por ser utilizada para ajustar dados simulados no Capítulo 4.

Passeio Aleatório

A distribuição passeio aleatório (Wise, 1966; Wasan, 1968) trata-se de uma distribuição especial dos MLGSs, porém não é membro dos modelos lineares generalizados e modelos lineares duplos (Cordeiro e Botter, 2001; Cordeiro et al., 2006). É possível construir uma variável aleatória Y com distribuição passeio aleatório, em que $Y = 1/Z$, sendo Z uma variável aleatória com distribuição inversa Gaussiana (Cordeiro e Botter, 2001; Johnson et al., 1994, p.282).

Se Y tem distribuição passeio aleatório com parâmetros θ e δ , sua função densidade de probabilidade é dada por

$$\pi(y; \theta, \delta) = \left(\frac{\delta}{2\pi y}\right)^{1/2} \exp\left\{-\frac{\delta y}{2} + \frac{\delta}{\theta} - \frac{\delta}{2\theta^2 y}\right\}, \quad (2.4)$$

com $y > 0, \theta > 0, \delta > 0$.

A média e variância de Y são dadas, respectivamente, por

$$E(Y) = \mu = \frac{1}{\theta} + \frac{1}{\delta} \quad \text{e} \quad \text{Var}(Y) = \frac{1}{(\theta\delta)} + \frac{2}{\delta^2}.$$

Para escrever (2.4) na forma da equação (2.1) por meio da parametrização da média, define-se $\phi = -\delta/(2\theta^2)$ e

$$\begin{aligned} \psi(\mu, \phi) = & 2(-\phi)^{1/2}\{(2\mu - \phi)^{1/2} - (-\phi)^{1/2}\}^{-1} + \frac{1}{2}\log 2 \\ & - \log\{(2\mu - \phi)^{1/2} - (-\phi)^{1/2}\} - \mu\{(2\mu - \phi)^{1/2} - (-\phi)^{1/2}\}^{-2}, \end{aligned} \quad (2.5)$$

em que μ e ϕ são parâmetros ortogonais.

Substituindo $\mu = 1/\theta + 1/\delta$ e $\phi = \delta/(2\theta^2)$ em (2.1), após algumas manipulações algébricas, é possível obter as seguintes quantidades:

$$\begin{aligned} A(y) &= \left(\frac{1}{2\pi y}\right)^{1/2}, \\ \psi^{(1,0)}(\mu, \phi) &= -\frac{\delta}{2}, \\ \Gamma(y) &= \frac{1}{y}, \\ \psi(\mu, \phi) &= \frac{\delta}{\theta} + \frac{1}{2}\log \delta - \frac{\delta}{2\theta} - \frac{1}{2}. \end{aligned}$$

Assim, determinam-se todas as componentes da distribuição passeio aleatório sob a forma dos MLGSs. As derivadas da função $\psi(\mu, \phi)$ dada em (2.5) utilizadas nessa dissertação encontram-se no Apêndice B.1. Ressalta-se ainda que o cálculo, considerando o modelo passeio aleatório, para a obtenção das matrizes \mathbf{M}_1 , \mathbf{M}_2 , Φ_1 e Φ_2 definidas em (2.10), pode ser visto no Apêndice B.2.

Além da distribuição passeio aleatório, Previdelli (2005) apresentou as distribuições inversa Gaussiana generalizada, família exponencial dupla e Poisson dupla como pertencentes à classe de modelos aqui estudada, sendo estas descritas a seguir.

Inversa Gaussiana Generalizada

Introduzida por Good (1953) e estudada por Wise (1971, 1975), trata-se de uma importante generalização da distribuição inversa Gaussiana. A função densidade de probabilidade da distribuição inversa Gaussiana generalizada usa os parâmetros ψ , χ e λ e é dada por

$$\pi(y; \psi, \chi, \lambda) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} y^{\lambda-1} \exp\left\{-\frac{1}{2}(\chi y^{-1} + \psi y)\right\}, \quad y > 0; \psi, \chi > 0, \quad (2.6)$$

em que $K_\lambda(\cdot)$ é a função Bessel modificada de terceira ordem. São casos especiais de (2.6) as seguintes distribuições:

- Gama ($\chi = 0, \lambda > 0$),
- Recíproca gama ($\psi = 0, \lambda < 0$),
- Inversa Gaussiana ($\lambda = -1/2$),
- Recíproca da inversa Gaussiana ($\lambda = 1/2$),
- Hiperbólica ($\lambda = 0$).

A média e a variância de Y são dadas, respectivamente, por

$$E(Y) = \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\psi\chi})}{K_\lambda(\sqrt{\psi\chi})}$$

e

$$\text{Var}(Y) = \left(\frac{\chi}{\psi}\right) \left[\frac{K_{\lambda+2}(\sqrt{\psi\chi})}{K_\lambda(\sqrt{\psi\chi})} - \left(\frac{K_{\lambda+1}(\sqrt{\psi\chi})}{K_\lambda(\sqrt{\psi\chi})}\right)^2 \right].$$

Mais informações sobre a distribuição inversa Gaussiana generalizada podem ser obtidas de Jørgensen (1982).

Família Exponencial Dupla

Nas famílias exponenciais uniparamétricas tais como a binomial e Poisson, a variância é uma função da média, sendo comum neste caso o fenômeno da superdispersão. Famílias exponenciais duplas permitem a introdução de um segundo parâmetro que controla a variância independentemente da média. Estas famílias ainda compartilham as propriedades da família exponencial simultaneamente para os parâmetros média e dispersão.

Inicialmente, considere a função de densidade da família exponencial uniparamétrica dada por Efron (1986)

$$g(y; \mu, n) = \exp\{n[\eta y - \psi(\mu)]\} [dG(y; n)], \quad (2.7)$$

em que $\mu = E(Y)$; η é uma função monótona de μ ; $\psi(\mu)$ é uma função normalizadora, escolhida de tal forma que a integral da densidade seja 1; $G(y; n)$ é uma medida conhecida; e n é o tamanho da amostra.

A partir da família exponencial dada em (2.7), a família com função de densidade

$$\pi(y; \mu, \theta, n) = c(\mu, \theta, n)\theta^{1/2}\{g(y; \mu, n)\}^\theta\{g(y; n)\}^{1-\theta}[dG(y; n)] \quad (2.8)$$

é chamada família exponencial dupla com parâmetros μ , θ e n . A constante $c(\mu, \theta, n)$ é definida de tal forma que $\int_{-\infty}^{\infty} \pi(y; \mu, \theta, n)dG(y; n) = 1$. Ao usar (2.8) em análise de regressão, os parâmetros desconhecidos μ e θ poderão ser estimados a partir dos dados. É possível destacar ainda que a densidade (2.8) possui média e variância dadas, aproximadamente, por

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{V(\mu)}{n\theta},$$

em que $V(\mu)$ é a função de variância.

Por fim, destaca-se que os MLGSs englobam a família exponencial dupla apresentada por Efron (1986) e, assim, todas as distribuições pertencentes às famílias binomial, gama e Poisson dupla podem ser escritas sob a forma dos MLGSs. Ressalta-se ainda que, embora a distribuição binomial negativa seja bastante utilizada para casos em que há superdispersão, a mesma não pertence à classe dos MLGSs.

Poisson Dupla

Suponha que $g(y; \mu, n)$ em (2.7) é a função densidade de probabilidade

$$g(y; \mu) = \frac{\exp^{-\mu} \mu^y}{y!}, \quad (2.9)$$

com $y = 0, 1, \dots$. Segundo Efron (1986), o tamanho amostral n foi suprimido da equação (2.9) por uma simples razão: a família Poisson é fechada sob convoluções, então $g(y; \mu, n)$ é a mesma família para todos os valores de n . Assim, a função densidade de probabilidade aproximada para a família Poisson dupla é dada por

$$\pi(y; \mu, \theta) = (\theta^{1/2} e^{-\theta\mu}) \left(\frac{e^{-y} y^y}{y!} \right) \left(\frac{e\mu}{y} \right)^{\theta y},$$

$y = 0, 1, 2, \dots$

A média e a variância de Y são dadas, respectivamente, por

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{\mu}{\theta}.$$

2.5 Aspectos inferenciais

2.5.1 Função Escore

A seguir, são introduzidas algumas notações necessárias no cálculo da função escore e matriz de informação de Fisher para os modelos lineares generalizados com superdispersão. Para tanto, seja

$$\begin{aligned} m_{i\ell} &= \frac{\partial^i \mu_\ell}{\partial \eta_\ell^i}, \\ \phi_{i\ell} &= \frac{\partial^i \phi_\ell}{\partial \tau_\ell^i}, \\ \mathbf{\Psi}^{(r,s)} &= \text{diag}\{\psi_1^{(r,s)}, \dots, \psi_n^{(r,s)}\}, \\ \mathbf{M}_r &= \text{diag}\{m_{r1}, \dots, m_{rn}\}, \\ \mathbf{\Phi}_r &= \text{diag}\{\phi_{r1}, \dots, \phi_{rn}\}, \end{aligned} \tag{2.10}$$

em que $m_{i\ell}$ e $\phi_{i\ell}$ denotam, respectivamente, a i -ésima derivada das funções de ligação inversa $\boldsymbol{\mu} = f^{-1}(\boldsymbol{\eta})$ e $\boldsymbol{\phi} = g^{-1}(\boldsymbol{\tau})$, para $i = 1, 2$ e $\ell = 1, \dots, n$. As matrizes $\mathbf{\Psi}^{(r,s)}$, \mathbf{M}_r e $\mathbf{\Phi}_r$ são de ordem $n \times n$, em que, para as duas últimas, $r = 1, 2$. E ainda, $\psi_\ell^{(r,s)} = \partial^{r+s} \psi(\mu_\ell, \phi_\ell) / \partial \mu_\ell^r \partial \phi_\ell^s$, $r, s \geq 0$, $\ell = 1, \dots, n$.

A função escore para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ é particionada em duas submatrizes, uma para cada parâmetro aqui considerado, e é dada por

$$\mathbf{U} = \mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{pmatrix} \partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) / \partial \boldsymbol{\beta} \\ \partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\top \mathbf{\Psi}^{(2,0)} \mathbf{M}_1 (\mathbf{y} - \boldsymbol{\mu}) \\ \mathbf{S}^\top \mathbf{\Phi}_1 \boldsymbol{\nu} \end{pmatrix}, \tag{2.11}$$

em que $(\mathbf{y} - \boldsymbol{\mu}) = ((y_1 - \mu_1), \dots, (y_n - \mu_n))^\top$ e $\boldsymbol{\nu} = (v_1, \dots, v_n)^\top$, com $v_\ell = (y_\ell - \mu_\ell) \psi_\ell^{(1,1)} + T(y_\ell) + \psi_\ell^{(0,1)}$, $\ell = 1, \dots, n$.

2.5.2 Informação de Fisher

Matriz de Informação Observada de Fisher

Seja $-\ddot{\mathbf{L}}_{\theta\theta}$ a matriz de informação observada de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Após algumas manipulações algébricas envolvendo diferenciação de matrizes, é obtido

$$\ddot{\mathbf{L}}_{\theta\theta} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\gamma} \\ \ddot{\mathbf{L}}_{\gamma\beta} & \ddot{\mathbf{L}}_{\gamma\gamma} \end{pmatrix}, \text{ em que}$$

$$\begin{aligned} \ddot{\mathbf{L}}_{\beta\beta} &= \mathbf{X}^\top \left[\boldsymbol{\Psi}^{(3,0)} \mathbf{M}_1^2 \mathbf{D}(\mathbf{r}) - \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1^2 + \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_2 \mathbf{D}(\mathbf{r}) \right] \mathbf{X}, \\ \ddot{\mathbf{L}}_{\beta\gamma} &= \mathbf{X}^\top \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(2,1)} \boldsymbol{\Phi}_1 \mathbf{M}_1 \mathbf{S}, \\ \ddot{\mathbf{L}}_{\gamma\beta} &= \mathbf{S}^\top \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(2,1)} \boldsymbol{\Phi}_1 \mathbf{M}_1 \mathbf{X} = (\ddot{\mathbf{L}}_{\beta\gamma})^{-1}, \\ \ddot{\mathbf{L}}_{\gamma\gamma} &= \mathbf{S}^\top \left[\boldsymbol{\Phi}_2 \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(1,1)} + \boldsymbol{\Phi}_2 \mathbf{D}(\mathbf{T}) + \boldsymbol{\Phi}_2 \boldsymbol{\Psi}^{(0,1)} + \boldsymbol{\Phi}_1^2 \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(1,2)} + \boldsymbol{\Phi}_1^2 \boldsymbol{\Psi}^{(0,2)} \right] \mathbf{S}, \end{aligned} \quad (2.12)$$

sendo $\mathbf{D}(\mathbf{r}) = \text{diag}\{(y_1 - \mu_1), \dots, (y_n - \mu_n)\}$ e $\mathbf{D}(\mathbf{T}) = \text{diag}\{T(y_1), \dots, T(y_n)\}$. As matrizes $\ddot{\mathbf{L}}_{\beta\beta}$, $\ddot{\mathbf{L}}_{\beta\gamma}$, $\ddot{\mathbf{L}}_{\gamma\beta}$ e $\ddot{\mathbf{L}}_{\gamma\gamma}$ são de ordem $p \times p$, $p \times q$, $q \times p$ e $q \times q$, respectivamente. Assim, a matriz de informação observada, também conhecida como matriz de informação local, possui ordem $(p + q) \times (p + q)$.

Matriz de Informação Esperada de Fisher

A matriz de informação esperada de Fisher é bloco-diagonal dada por

$$\mathbf{K}_{\theta,\theta} = \mathbf{K}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbb{E}(-\ddot{\mathbf{L}}_{\theta\theta}) = \begin{pmatrix} \mathbf{K}_{\beta,\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{\gamma,\gamma} \end{pmatrix},$$

em que $\mathbf{K}_{\beta,\beta} = \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1^2 \mathbf{X}$ e $\mathbf{K}_{\gamma,\gamma} = -\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \mathbf{S}$ são as matrizes de informação para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, respectivamente. Logo, os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são globalmente ortogonais e seus estimadores de máxima verossimilhança (EMVs) $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$ são assintoticamente independentes. Ao avaliar $\mathbf{K}_{\theta,\theta}^{-1}$ nestes EMVs, tem-se uma estimativa da matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$.

O cálculo detalhado da função escore, assim como das matrizes de informação observada e esperada de Fisher, encontram-se no Apêndice A.

2.5.3 Estimação dos parâmetros

Nos MLGSs, a obtenção dos EMVs $\hat{\beta}$ e $\hat{\gamma}$ é feita por meio da maximização numérica do logaritmo da função de verossimilhança da distribuição de probabilidade dada em (2.3). Assim, pode-se utilizar um método de otimização não-linear, que consiste em encontrar o máximo de uma função por meio de algoritmos iterativos, via métodos gradientes, tais como: Newton-Raphson, scoring de Fisher, métodos *quasi*-Newton (DFP e BFGS), entre outros. Mais detalhes sobre métodos de otimização podem ser encontrados em Nocedal e Wright (1999, Capítulos 8 e 19) e Frery e Cribari-Neto (2005).

O processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ é definido expandindo-se a função escore $\mathbf{U}(\boldsymbol{\theta})$ em série de Taylor, a partir de um valor inicial $\boldsymbol{\theta}^{(0)}$, tal que

$$\mathbf{U}(\boldsymbol{\theta}) \cong \mathbf{U}(\boldsymbol{\theta}^{(0)}) + \mathbf{U}'(\boldsymbol{\theta}^{(0)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}),$$

em que $\mathbf{U}'(\boldsymbol{\theta})$ denota a primeira derivada de $\mathbf{U}(\boldsymbol{\theta})$ com respeito a $\boldsymbol{\theta}$. Assim, repetindo-se o procedimento acima, chega-se ao processo iterativo no $(m+1)$ -ésimo passo, que fica dado por

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \{-\mathbf{U}'(\boldsymbol{\theta}^{(m)})\}^{-1}\mathbf{U}(\boldsymbol{\theta}^{(m)}),$$

com $m = 0, 1, \dots$. Como a matriz $-\mathbf{U}'(\boldsymbol{\theta})$ pode não ser positiva definida, a aplicação do método scoring de Fisher substituindo a matriz $-\mathbf{U}'(\boldsymbol{\theta})$ pelo correspondente valor esperado, pode ser mais conveniente. Isso resulta no seguinte processo iterativo:

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

$m = 0, 1, \dots$.

Em notação matricial, o processo iterativo acima corresponde às seguintes equações para estimar $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ iterativamente

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

ou seja,

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \mathbf{K}_{\boldsymbol{\beta},\boldsymbol{\beta}}^{-1(m)} \mathbf{U}(\boldsymbol{\beta}^{(m)}) \quad (2.13)$$

e

$$\boldsymbol{\gamma}^{(m+1)} = \boldsymbol{\gamma}^{(m)} + \mathbf{K}_{\boldsymbol{\gamma},\boldsymbol{\gamma}}^{-1(m)} \mathbf{U}(\boldsymbol{\gamma}^{(m)}), \quad (2.14)$$

$m = 0, 1, \dots$

Substituindo

$$\mathbf{K}_{\boldsymbol{\beta},\boldsymbol{\beta}}^{-1(m)} = (\mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)2} \mathbf{X})^{-1} \quad \text{e} \quad \mathbf{U}(\boldsymbol{\beta}^{(m)}) = \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)} (\mathbf{y} - \boldsymbol{\mu})$$

na equação (2.13) acima, segue que

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} + (\mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)2} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)} (\mathbf{y} - \boldsymbol{\mu}) \\ &= (\mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)2} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)2} [\mathbf{X}\boldsymbol{\beta}^{(m)} + \mathbf{M}_1^{(m)-1} (\mathbf{y} - \boldsymbol{\mu})] \\ &= (\mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)2} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)2} \boldsymbol{\xi}_1^{(m)}, \end{aligned}$$

em que $\boldsymbol{\xi}_1 = \boldsymbol{\eta} + \mathbf{M}_1^{-1}(\mathbf{y} - \boldsymbol{\mu})$, sendo $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ e $m = 0, 1, \dots$

Analogamente para $\boldsymbol{\gamma}$, substituindo-se

$$\mathbf{K}_{\boldsymbol{\gamma},\boldsymbol{\gamma}}^{-1(m)} = (-\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)(m)} \boldsymbol{\Phi}_1^{(m)2} \mathbf{S})^{-1} \quad \text{e} \quad \mathbf{U}(\boldsymbol{\gamma}^{(m)}) = \mathbf{S}^\top \boldsymbol{\Phi}_1^{(m)} \boldsymbol{\nu}$$

na equação (2.14), tem-se que

$$\begin{aligned} \boldsymbol{\gamma}^{(m+1)} &= \boldsymbol{\gamma}^{(m)} + (-\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)(m)} \boldsymbol{\Phi}_1^{(m)2} \mathbf{S})^{-1} \mathbf{S}^\top \boldsymbol{\Phi}_1^{(m)} \boldsymbol{\nu} \\ &= (-\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)(m)} \boldsymbol{\Phi}_1^{(m)2} \mathbf{S})^{-1} (-\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)(m)} \boldsymbol{\Phi}_1^{(m)2}) [\mathbf{S}\boldsymbol{\gamma}^{(m)} - \boldsymbol{\Psi}^{(0,2)-1} \boldsymbol{\Phi}_1^{-1} \boldsymbol{\nu}] \\ &= (-\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)(m)} \boldsymbol{\Phi}_1^{(m)2} \mathbf{S})^{-1} (-\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)(m)} \boldsymbol{\Phi}_1^{(m)2}) \boldsymbol{\xi}_2^{(m)}, \end{aligned}$$

em que $\boldsymbol{\xi}_2 = \boldsymbol{\tau} - \boldsymbol{\Psi}^{(0,2)-1} \boldsymbol{\Phi}_1^{-1} \boldsymbol{\nu}$, sendo $\boldsymbol{\tau} = \mathbf{S}\boldsymbol{\gamma}$ e $\boldsymbol{\nu} = (v_1, \dots, v_n)^\top$, com $v_\ell = (y_\ell - \mu_\ell)\psi_\ell^{(1,1)} + T(y_\ell) + \psi_\ell^{(0,1)}$, $m = 0, 1, \dots$ e $\ell = 1, \dots, n$.

Considerando $\mathbf{W}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(m)} = \boldsymbol{\Psi}^{(2,0)(m)} \mathbf{M}_1^{(m)2}$ e $\mathbf{W}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{(m)} = -\boldsymbol{\Psi}^{(0,2)(m)} \boldsymbol{\Phi}_1^{(m)2}$, para $m = 0, 1, \dots$ chega-se a um processo iterativo de mínimos quadrados reponderados, que fica dado por

$$\begin{aligned}\boldsymbol{\beta}^{(m+1)} &= (\mathbf{X}^\top \mathbf{W}_{\beta\beta}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_{\beta\beta}^{(m)} \boldsymbol{\xi}_1^{(m)}, \\ \boldsymbol{\gamma}^{(m+1)} &= (\mathbf{S}^\top \mathbf{W}_{\gamma\gamma}^{(m)} \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{W}_{\gamma\gamma}^{(m)} \boldsymbol{\xi}_2^{(m)},\end{aligned}\tag{2.15}$$

$m = 0, 1, \dots$, $\boldsymbol{\xi}_1 = \boldsymbol{\eta} + \mathbf{M}_1^{-1}(\mathbf{y} - \boldsymbol{\mu})$ e $\boldsymbol{\xi}_2 = \boldsymbol{\tau} - \boldsymbol{\Psi}^{(0,2)^{-1}} \boldsymbol{\Phi}_1^{-1} \boldsymbol{\nu}$, com $\boldsymbol{\nu} = (v_1, \dots, v_n)^\top$, sendo $v_\ell = (y_\ell - \mu_\ell) \psi_\ell^{(1,1)} + T(y_\ell) + \psi_\ell^{(0,1)}$, $\ell = 1, \dots, n$.

Note que $\boldsymbol{\xi}_1$ e $\boldsymbol{\xi}_2$ em (2.15) desempenham o papel de uma variável dependente modificada, enquanto $\mathbf{W}_{\beta\beta}$ e $\mathbf{W}_{\gamma\gamma}$ correspondem a uma matriz de pesos que muda a cada passo do processo iterativo.

Em termos gerais, tem-se que fazer a regressão da variável dependente modificada $(\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)^\top$ sob a matriz modelo (\mathbf{X}, \mathbf{S}) com os pesos modificados definidos por

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{\beta\beta} & \mathbf{W}_{\beta\gamma} \\ \mathbf{W}_{\gamma\beta} & \mathbf{W}_{\gamma\gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1^2 & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \end{pmatrix}.$$

Capítulo 3

Técnicas de Diagnóstico nos MLGSs

Um ajuste de regressão é uma representação de alguns aspectos dos dados. No entanto, uma única observação pode provocar distorções na interpretação dos coeficientes do modelo, assim como exercer um efeito desproporcional no ajuste, levando muitas vezes a conclusões errôneas quanto aos resultados inferenciais. Na área de modelagem, uma etapa importante é a análise de diagnóstico, que consiste em verificar possíveis afastamentos das suposições feitas para o modelo, assim como a existência de observações extremas que exerçam alguma interferência desproporcional nos resultados do ajuste. Esta etapa teve seu início com a análise de resíduos para avaliar a adequação da distribuição proposta da variável resposta, assim como detectar pontos aberrantes entre o conjunto de dados.

Neste Capítulo são apresentadas algumas técnicas de diagnósticos para os modelos lineares generalizados com superdispersão, sendo esta a principal contribuição desta dissertação por não haver estudos anteriores que aborde este tema quanto à referida classe de modelos. Inicialmente, na Seção 3.1, é tratado o conceito de alavancagem generalizada, sob a abordagem de Wei, Hu e Fung (1998), em que eles generalizam a definição de pontos de alavanca para modelos gerais. Na Seção 3.2 é apresentada a análise de resíduos para os MLGSs e em seguida, na Seção 3.3, a técnica de deleção de pontos (influência global), esta inicialmente introduzida por Cook (1977) para modelos normais lineares e posteriormente estendida para diversas classes de modelos. Por fim, na Seção 3.4, é estudada a metodologia de influência local proposta por Cook (1986), em que a ideia básica consiste em avaliar a influência conjunta

das observações sob pequenas perturbações nos dados ou no próprio modelo. Mais referências sobre métodos de diagnóstico podem ser encontradas em Paula (2012).

3.1 Alavancagem Generalizada

Pontos de alavanca são aqueles que têm grande influência no próprio valor ajustado e a presença destes causa aumento na variabilidade da previsão. Tais pontos se caracterizam também por ter um perfil diferente dos demais no que diz respeito aos valores das variáveis explicativas do modelo, em que, dependendo da localização, esses pontos podem exercer forte influência nas estimativas da variância dos coeficientes de regressão. Assim, o conceito de alavancagem é um dos componentes chaves da análise de diagnóstico em modelos de regressão linear, em que a ideia principal é avaliar a influência de y_i sobre o próprio valor predito. Uma forma de medir essa influência é por meio da derivada $\partial \hat{y}_i / \partial y_i$, que coincide com h_{ii} no caso normal linear, em que h_{ii} representa o i -ésimo elemento da diagonal principal da matriz de projeção (também conhecida como matriz hat ou matriz chapéu) $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ e \mathbf{X} é a matriz modelo. A matriz \mathbf{H} caracteriza-se por ser simétrica e idempotente, em que esta última propriedade resulta na seguinte relação: $\text{posto}(\mathbf{H}) = \text{tr}(\mathbf{H})$.

A alavancagem generalizada de um estimador é definida em modelos de regressão como uma medida da importância de observações avaliadas individualmente. Wei, Hu e Fung (1998) derivaram um simples mas poderoso resultado, desenvolvendo uma expressão explícita para alavancagem, partindo do ponto de vista de que $h_{ii} = \partial \hat{y}_i / \partial y_i$ reflete mais diretamente a influência de y_i no ajuste.

Seja $\mathbf{y} = (y_1, \dots, y_n)^\top$ um vetor de respostas observadas com função densidade de probabilidade $\pi(\mathbf{y}; \boldsymbol{\theta})$, onde $\boldsymbol{\theta}$ é um vetor de parâmetros desconhecidos. A esperança de \mathbf{y} , $\boldsymbol{\mu} = E(\mathbf{y})$, pode ser expressa como $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$. Seja $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{y})$ o estimador de máxima verossimilhança de $\boldsymbol{\theta}$, então $\hat{\mathbf{y}} = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$ é o vetor de respostas preditas. Segundo Wei, Hu e Fung (1998), a alavancagem generalizada de $\hat{\boldsymbol{\theta}}$, $\partial \hat{\mathbf{y}} / \partial \mathbf{y}^\top$ é definida como

$$\text{GL}(\hat{\boldsymbol{\theta}}) = \{(\mathbf{D}_\theta)(-\ddot{\mathbf{L}}_{\theta\theta})^{-1}(\ddot{\mathbf{L}}_{\theta\mathbf{y}})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (3.1)$$

em que $\mathbf{D}_\theta = \partial \boldsymbol{\mu} / \partial \boldsymbol{\theta}^\top$, $\ddot{\mathbf{L}}_{\theta\theta} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ e $\ddot{\mathbf{L}}_{\theta\mathbf{y}} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \mathbf{y}^\top$, com

$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ definida em (2.3).

Na Seção 2.3 encontra-se definido $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$. Sendo $\mathbf{D}_{\boldsymbol{\theta}}^{\mu} = (\mathbf{D}_{\boldsymbol{\beta}}^{\mu}, \mathbf{D}_{\boldsymbol{\gamma}}^{\mu})$, tem-se que

$$\frac{\partial \mu_{\ell}}{\partial \beta_r} = \frac{\partial \mu_{\ell}}{\partial \eta_{\ell}} \cdot \frac{\partial \eta_{\ell}}{\partial \beta_r} = m_{1\ell}(r)_{\ell} \quad \text{e} \quad \frac{\partial \mu_{\ell}}{\partial \gamma_R} = \frac{\partial \mu_{\ell}}{\partial \eta_{\ell}} \cdot \frac{\partial \eta_{\ell}}{\partial \gamma_R} = 0,$$

em que $r = 1, \dots, p$, $R = 1, \dots, q$ e $\ell = 1, \dots, n$.

Assim, em forma matricial,

$$\mathbf{D}_{\boldsymbol{\theta}}^{\mu} = (\mathbf{M}_1 \mathbf{X}, \mathbf{0}), \quad (3.2)$$

em que $\mathbf{M}_1 = \text{diag}\{m_{11}, \dots, m_{1n}\}$, com $m_{1\ell} = \partial \mu_{\ell} / \partial \eta_{\ell}$, $\ell = 1, \dots, n$.

A matriz $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\mathbf{y}}$ é definida como

$$\ddot{\mathbf{L}}_{\boldsymbol{\theta}\mathbf{y}} = \begin{bmatrix} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \\ \ddot{\mathbf{L}}_{\gamma\mathbf{y}} \end{bmatrix},$$

sendo $\ddot{\mathbf{L}}_{\beta\mathbf{y}} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \mathbf{y}^{\top}$ e $\ddot{\mathbf{L}}_{\gamma\mathbf{y}} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\gamma} \partial \mathbf{y}^{\top}$.

A derivada de $\ell(\boldsymbol{\theta})$ em relação a β_r é

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_r} = \sum_{\ell=1}^n \left\{ \psi_{\ell}^{(2,0)}(y_{\ell} - \mu_{\ell}) m_{1\ell}(r)_{\ell} \right\},$$

em que $\psi_{\ell}^{(2,0)} = \partial^2 \psi(\mu_{\ell}, \phi_{\ell}) / \partial \mu_{\ell}^2$, $m_{1\ell} = \partial \mu_{\ell} / \partial \eta_{\ell}$ e $(r)_{\ell} = \partial \eta_{\ell} / \partial \beta_r$, $r = 1, \dots, p$ e $\ell = 1, \dots, n$.

Assim, o $(r\ell)$ -ésimo elemento de $\ddot{\mathbf{L}}_{\beta\mathbf{y}}$ é dado por

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_r \partial y_{\ell}} = \sum_{\ell=1}^n \left\{ \psi_{\ell}^{(2,0)} m_{1\ell}(r)_{\ell} \right\},$$

$r = 1, \dots, p$ e $\ell = 1, \dots, n$.

Analogamente, tem-se que a primeira derivada de $\ell(\boldsymbol{\theta})$ em relação a γ_R é

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_R} = \sum_{\ell=1}^n \phi_{1\ell}(R)_{\ell} v_{\ell},$$

em que $\phi_{1\ell} = \partial \phi_{\ell} / \partial \tau_{\ell}$, $(R)_{\ell} = \partial \tau_{\ell} / \partial \gamma_R$ e $v_{\ell} = \psi_{\ell}^{(1,1)}(y_{\ell} - \mu_{\ell}) + T(y_{\ell}) + \psi_{\ell}^{(0,1)}$, com $\psi_{\ell}^{(1,1)} = \partial^2 \psi(\mu_{\ell}, \phi_{\ell}) / \partial \mu_{\ell} \partial \phi_{\ell}$ e $\psi_{\ell}^{(0,1)} = \partial \psi(\mu_{\ell}, \phi_{\ell}) / \partial \phi_{\ell}$, $\ell = 1, \dots, n$.

Então, o $(R\ell)$ -ésimo elemento de $\ddot{\mathbf{L}}_{\gamma\mathbf{y}}$ pode ser expresso como

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_R \partial y_\ell} = \sum_{\ell=1}^n \left\{ \psi_\ell^{(1,1)} \phi_{1\ell}(R)_\ell + \frac{\partial T(y_\ell)}{\partial y_\ell} \phi_{1\ell}(R)_\ell \right\},$$

$R = 1, \dots, q$ e $\ell = 1, \dots, n$.

Em notação matricial,

$$\ddot{\mathbf{L}}_{\theta\mathbf{y}} = \begin{bmatrix} \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1 \\ \mathbf{S}^\top \left(\boldsymbol{\Psi}^{(1,1)} + \mathbf{D}(\mathbf{T}'_{\mathbf{y}}) \right) \boldsymbol{\Phi}_1 \end{bmatrix}, \quad (3.3)$$

em que $\mathbf{D}(\mathbf{T}'_{\mathbf{y}}) = \text{diag}\{T'(y_1), \dots, T'(y_n)\}$, com $T'(y_\ell) = \frac{\partial T(y_\ell)}{\partial y_\ell}$, $\ell = 1, \dots, n$. Os componentes das matrizes diagonais $\boldsymbol{\Psi}^{(2,0)}$, $\boldsymbol{\Psi}^{(1,1)}$, \mathbf{M}_1 e $\boldsymbol{\Phi}_1$ estão definidos em (2.10).

Substituindo (2.12), (3.2) e (3.3) em (3.1), tem-se que a matriz generalizada de pontos de alavanca para $\boldsymbol{\mu}$ fica dada por

$$GL^\mu(\hat{\boldsymbol{\theta}}) = \mathbf{M}_1 \mathbf{X} \left\{ \ddot{\mathbf{L}}_{\beta\beta}^{-1} + \mathbf{F} \mathbf{E}^{-1} \mathbf{F}^\top \right\} \ddot{\mathbf{L}}_{\beta\mathbf{y}} - \mathbf{M}_1 \mathbf{X} \mathbf{F} \mathbf{E}^{-1} \ddot{\mathbf{L}}_{\gamma\mathbf{y}} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (3.4)$$

em que $\mathbf{F} = \ddot{\mathbf{L}}_{\beta\beta}^{-1} \ddot{\mathbf{L}}_{\beta\gamma}$, $\mathbf{F}^\top = \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1}$, $\mathbf{E} = -\ddot{\mathbf{L}}_{\gamma\gamma} + \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1} \ddot{\mathbf{L}}_{\beta\gamma}$, e $\ddot{\mathbf{L}}_{\beta\beta}$, $\ddot{\mathbf{L}}_{\beta\gamma}$, $\ddot{\mathbf{L}}_{\gamma\beta}$ e $\ddot{\mathbf{L}}_{\gamma\gamma}$ são elementos da matriz de informação observada dada em (2.12).

É possível ainda reescrever (3.4) da seguinte maneira

$$GL^\mu(\hat{\boldsymbol{\theta}}) = GL_\beta^\mu(\hat{\boldsymbol{\theta}}) + GL_\gamma^\mu(\hat{\boldsymbol{\theta}}),$$

em que

$$GL_\beta^\mu(\hat{\boldsymbol{\theta}}) = -\mathbf{M}_1 \mathbf{X} \left\{ \mathbf{X}^\top \left[\boldsymbol{\Psi}^{(3,0)} \mathbf{M}_1^2 \mathbf{D}(\mathbf{r}) - \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1^2 + \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_2 \mathbf{D}(\mathbf{r}) \right] \mathbf{X} \right\}^{-1} \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1 |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

e

$$GL_\gamma^\mu(\hat{\boldsymbol{\theta}}) = \mathbf{M}_1 \mathbf{X} \mathbf{F} \mathbf{E}^{-1} \mathbf{F}^\top \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1 - \mathbf{M}_1 \mathbf{X} \mathbf{F} \mathbf{E}^{-1} \left\{ \mathbf{S}^\top \left[\boldsymbol{\Psi}^{(1,1)} + \mathbf{D}(\mathbf{T}'_{\mathbf{y}}) \right] \boldsymbol{\Phi}_1 \right\} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

sendo $\mathbf{D}(\mathbf{T}'_{\mathbf{y}}) = \text{diag}\{T'(y_1), \dots, T'(y_n)\}$, com $T'(y_\ell) = \frac{\partial T(y_\ell)}{\partial y_\ell}$ e $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$, em que $r_\ell = (y_\ell - \hat{\mu}_\ell)$, $\ell = 1, \dots, n$. Os componentes das matrizes diagonais $\boldsymbol{\Psi}^{(3,0)}$, $\boldsymbol{\Psi}^{(2,0)}$, $\boldsymbol{\Psi}^{(1,1)}$, \mathbf{M}_1 , \mathbf{M}_2 e $\boldsymbol{\Phi}_1$ encontram-se definidos em (2.10).

Também é possível observar a influência de y_ℓ em $\hat{\phi}_\ell$, denotada por $\partial\hat{\phi}_\ell/\partial y_\ell$, em que $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^\top$. Sendo $\mathbf{D}_\theta^\phi = (\mathbf{D}_\beta^\phi, \mathbf{D}_\gamma^\phi)$, tem-se que

$$\frac{\partial\phi_\ell}{\partial\beta_r} = \frac{\partial\phi_\ell}{\partial\tau_\ell} \cdot \frac{\partial\tau_\ell}{\partial\beta_r} = 0 \quad \text{e} \quad \frac{\partial\phi_\ell}{\partial\gamma_R} = \frac{\partial\phi_\ell}{\partial\tau_\ell} \cdot \frac{\partial\tau_\ell}{\partial\gamma_R} = \phi_{1\ell}(R)_\ell,$$

em que $r = 1, \dots, p$, $R = 1, \dots, q$ e $\ell = 1, \dots, n$.

Então, em forma matricial,

$$\mathbf{D}_\theta^\phi = (\mathbf{0}, \boldsymbol{\Phi}_1 \mathbf{S}), \quad (3.5)$$

em que $\boldsymbol{\Phi}_1 = \text{diag}\{\phi_{11}, \dots, \phi_{1n}\}$, com $\phi_{1\ell} = \partial\phi_\ell/\partial\tau_\ell$, $\ell = 1, \dots, n$.

As demais quantidades $\ddot{\mathbf{L}}_{\theta\theta}$ e $\ddot{\mathbf{L}}_{\theta\mathbf{y}}$ são dadas em (2.12) e (3.3), respectivamente.

Assim, analogamente a $GL^\mu(\hat{\boldsymbol{\theta}})$, substituindo (2.12), (3.5) e (3.3) em (3.1), a matriz generalizada de pontos de alavanca $\partial\hat{\boldsymbol{\phi}}/\partial\mathbf{y}$ pode ser expressa como

$$GL^\phi(\hat{\boldsymbol{\theta}}) = -\boldsymbol{\Phi}_1 \mathbf{S} \mathbf{E}^{-1} \mathbf{F}^\top \ddot{\mathbf{L}}_{\beta\mathbf{y}} + \boldsymbol{\Phi}_1 \mathbf{S} \mathbf{E}^{-1} \ddot{\mathbf{L}}_{\gamma\mathbf{y}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (3.6)$$

sendo $\mathbf{E} = -\ddot{\mathbf{L}}_{\gamma\gamma} + \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1} \ddot{\mathbf{L}}_{\beta\gamma}$, $\mathbf{F}^\top = \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1}$. A matriz generalizada de pontos de alavanca para $\boldsymbol{\phi}$ dada em (3.6) pode ainda ser reescrita na forma

$$GL^\phi(\hat{\boldsymbol{\theta}}) = GL_\beta^\phi(\hat{\boldsymbol{\theta}}) + GL_\gamma^\phi(\hat{\boldsymbol{\theta}}),$$

em que

$$GL_\beta^\phi(\hat{\boldsymbol{\theta}}) = -\boldsymbol{\Phi}_1 \mathbf{S} \mathbf{E}^{-1} \mathbf{F}^\top \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1 \\ + \boldsymbol{\Phi}_1 \mathbf{S} \mathbf{V} \mathbf{W}^{-1} \mathbf{V}^\top \left\{ \mathbf{S}^\top \left[(\boldsymbol{\Psi}^{(1,1)} + \mathbf{D}(\mathbf{T}'_y)) \boldsymbol{\Phi}_1 \right] \right\} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

e

$$GL_\gamma^\phi(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Phi}_1 \mathbf{S} \left(-\ddot{\mathbf{L}}_{\gamma\gamma} \right)^{-1} \ddot{\mathbf{L}}_{\gamma\mathbf{y}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ = \boldsymbol{\Phi}_1 \mathbf{S} \left(-\ddot{\mathbf{L}}_{\gamma\gamma} \right)^{-1} \left[\mathbf{S}^\top \left(\boldsymbol{\Psi}^{(1,1)} + \mathbf{D}(\mathbf{T}'_y) \right) \boldsymbol{\Phi}_1 \right] |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

com $\mathbf{E} = -\ddot{\mathbf{L}}_{\gamma\gamma} + \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1} \ddot{\mathbf{L}}_{\beta\gamma}$, $\mathbf{F}^\top = \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1}$, $\mathbf{V} = (-\ddot{\mathbf{L}}_{\gamma\gamma})^{-1} \ddot{\mathbf{L}}_{\gamma\beta}$, $\mathbf{V}^\top = \ddot{\mathbf{L}}_{\beta\gamma} (-\ddot{\mathbf{L}}_{\gamma\gamma})^{-1}$, $\mathbf{W} = \ddot{\mathbf{L}}_{\beta\beta} - \ddot{\mathbf{L}}_{\beta\gamma} \ddot{\mathbf{L}}_{\gamma\gamma}^{-1} \ddot{\mathbf{L}}_{\gamma\beta}$, e $\ddot{\mathbf{L}}_{\beta\beta}$, $\ddot{\mathbf{L}}_{\beta\gamma}$, $\ddot{\mathbf{L}}_{\gamma\beta}$ e $\ddot{\mathbf{L}}_{\gamma\gamma}$ são elementos da matriz de informação observada dada em (2.12). Os componentes das

matrizes diagonais $\Psi^{(2,0)}$, $\Psi^{(1,1)}$, \mathbf{M}_1 e Φ_1 estão definidos em (2.10) e $\mathbf{D}(\mathbf{T}'_y) = \text{diag}\{T'(y_1), \dots, T'(y_n)\}$, com $T'(y_\ell) = \frac{\partial T(y_\ell)}{\partial y_\ell}$, $\ell = 1, \dots, n$.

Os gráficos de $GL^\mu(\hat{\theta})_{\ell\ell}$ e $GL^\phi(\hat{\theta})_{\ell\ell}$ contra a ordem das observações são sugeridos para detectar possíveis pontos com alta alavancagem. É comum também utilizar os gráficos de $GL^\mu(\hat{\theta})_{\ell\ell}$ e $GL^\phi(\hat{\theta})_{\ell\ell}$ contra os valores ajustados.

3.2 Resíduos

A análise de diagnóstico, uma etapa muito importante na análise de regressão, teve início com a análise de resíduos para detectar possíveis pontos mal ajustados e avaliar a adequação da distribuição proposta para a variável resposta. Assim, a análise de resíduos, sob a abordagem dos modelos lineares, constitui uma ferramenta amplamente utilizada para verificar a adequacidade do ajuste, presença de observações extremas (aberrantes) ou que exerçam alguma influência desproporcional nos resultados do ajuste, como também detectar possíveis afastamentos das suposições inicialmente feitas para o modelo, entre as quais a distribuição de probabilidade dos dados, além de verificar indícios de heteroscedasticidade. Uma definição mais geral de resíduos pode ser vista no artigo apresentado por Cox e Snell (1968).

A seguir são apresentados os resíduos padronizado, componente do desvio e componente do desvio padronizado para os modelos lineares generalizados com superdispersão. Vale ressaltar que toda a análise de resíduos apresentada nesta Seção refere-se somente à construção de expressões para resíduos com respeito ao ajuste da média. Analogamente, é possível obter equações para os resíduos aqui propostos quanto ao ajuste da dispersão, sendo esta uma abordagem para trabalhos futuros.

Resíduo padronizado

A definição mais usual de resíduo é a diferença entre o valor observado e o valor ajustado da ℓ -ésima observação, denominado resíduo ordinário, dado por $r_\ell = (y_\ell - \hat{\mu}_\ell)$, $\ell = 1, \dots, n$, em que o sinal de r_ℓ indica a direção dessa diferença. A análise de resíduos pode se basear nestes resíduos e suas possíveis padronizações, assim como nos resíduos construídos a partir dos componentes da função desvio, frequentemente utilizados em modelos lineares generalizados.

Seja o vetor de resíduos ordinários definido por $\mathbf{r} = (r_1, \dots, r_n)^\top$. Logo, da regressão normal linear segue que $\mathbf{r} = \mathbf{y} - \hat{\boldsymbol{\mu}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$, em que \mathbf{H} é a matriz de projeção ortogonal definida na Seção 3.1. Desta forma, tem-se que $E(\mathbf{r}) = (\mathbf{I} - \mathbf{H})E(\mathbf{Y}) = \mathbf{0}$ e $\text{Var}(\mathbf{r}) = \sigma^2(\mathbf{I} - \mathbf{H})$, com $\sigma^2 = \text{Var}(Y)$. Ou seja, r_ℓ tem distribuição normal de média zero, variância $\sigma^2(1 - h_{\ell\ell})$ e ainda a covariância de r_ℓ e r_j , com $\ell \neq j$, $\ell, j = 1, \dots, n$, fica dada por $-\sigma^2 h_{\ell j}$. Como os r_ℓ 's têm variâncias distintas, é comum expressá-los em forma padronizada, dividindo-se r_ℓ por seu respectivo desvio padrão, permitindo assim uma comparabilidade entre os resíduos. Logo, é obtido o resíduo studentizado dado por

$$t_\ell = \frac{r_\ell}{s(1 - h_{\ell\ell})^{1/2}}, \ell = 1, \dots, n,$$

em que $s^2 = \sum_{\ell=1}^n r_\ell^2 / (n - p)$.

Porém, como r_ℓ não é independente de s^2 , t_ℓ não segue uma distribuição t de Student. Contudo, este problema pode ser contornado substituindo s^2 por $s_{(\ell)}^2$, que corresponde ao erro quadrático médio do modelo sem a ℓ -ésima observação. Assim, é obtido um novo resíduo studentizado, que fica dado por

$$t_\ell^* = \frac{r_\ell}{s_{(\ell)}(1 - h_{\ell\ell})^{1/2}},$$

em que $\ell = 1, \dots, n$, e t_ℓ^* segue uma distribuição central t_{n-p-1} .

A definição de um resíduo studentizado para os MLGSs pode ser feita analogamente à regressão normal linear. Porém, não necessariamente esse novo resíduo possui as mesmas propriedades de t_ℓ^* . Então, torna-se importante a definição de novos resíduos que tenham propriedades conhecidas ou que estejam próximas das propriedades de t_ℓ^* .

Pregibon (1981) definiu um resíduo derivado do processo iterativo de estimação dos parâmetros do modelo. Nos MLGSs, tal processo é dado por (2.15) e a partir da convergência do mesmo tem-se que a expressão do estimador de $\boldsymbol{\beta}$ para a $(m + 1)$ -ésima iteração toma a seguinte forma

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{W}_{\beta\beta}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_{\beta\beta}^{(m)} \boldsymbol{\xi}_1^{(m)}, m = 0, 1, \dots \quad (3.7)$$

em que $\mathbf{W}_{\beta\beta}^{(m)} = \boldsymbol{\Psi}^{(2,0)^{(m)}} \mathbf{M}_1^{(m)^2}$ e $\boldsymbol{\xi}_1 = \boldsymbol{\eta} + \mathbf{M}_1^{-1}(\mathbf{y} - \boldsymbol{\mu})$.

A partir da convergência do processo iterativo dado em (3.7) é obtido

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\boldsymbol{\xi}}_1,$$

em que $\hat{\boldsymbol{\xi}}_1 = \hat{\boldsymbol{\eta}} + \hat{\mathbf{M}}_1^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}) = \hat{\boldsymbol{\eta}} + \hat{\boldsymbol{\Psi}}^{(2,0)1/2} \hat{\mathbf{W}}_{\beta\beta}^{-1/2}(\mathbf{y} - \hat{\boldsymbol{\mu}})$, com $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{W}}_{\beta\beta} = \hat{\boldsymbol{\Psi}}^{(2,0)} \hat{\mathbf{M}}_1^2$ e os componentes das matrizes diagonais \mathbf{M}_1 e $\boldsymbol{\Psi}^{(2,0)}$ estão definidos em (2.10). Assim, $\hat{\boldsymbol{\beta}}$ pode ser interpretado como a solução de mínimos quadrados da regressão linear de $\hat{\mathbf{W}}_{\beta\beta}^{1/2} \hat{\boldsymbol{\xi}}_1$ contra as colunas de $\hat{\mathbf{W}}_{\beta\beta}^{1/2} \mathbf{X}$.

Uma proposta então é considerar o resíduo ordinário da solução de mínimos quadrados da regressão linear ponderada de $\hat{\boldsymbol{\xi}}_1$ contra \mathbf{X} , que é definido por

$$\begin{aligned} \mathbf{r}^* &= \hat{\mathbf{W}}_{\beta\beta}^{1/2} [\hat{\boldsymbol{\xi}}_1 - \hat{\boldsymbol{\eta}}] \\ &= \hat{\mathbf{W}}_{\beta\beta}^{1/2} [\hat{\boldsymbol{\eta}} + \hat{\boldsymbol{\Psi}}^{(2,0)1/2} \hat{\mathbf{W}}_{\beta\beta}^{-1/2}(\mathbf{y} - \hat{\boldsymbol{\mu}}) - \hat{\boldsymbol{\eta}}] \\ &= \hat{\boldsymbol{\Psi}}^{(2,0)1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}}) \\ &= (\hat{\text{Var}}(\mathbf{Y}))^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}}), \end{aligned}$$

em que $\hat{\text{Var}}(\mathbf{Y}) = \hat{\boldsymbol{\Psi}}^{(2,0)-1}$.

Assumindo que $\text{Var}(\hat{\boldsymbol{\xi}}_1) \cong \hat{\mathbf{W}}_{\beta\beta}^{-1}$, tem-se aproximadamente que $\text{Var}(\mathbf{r}^*) \cong (\mathbf{I} - \hat{\mathbf{H}})$, com $\hat{\mathbf{H}} = \hat{\mathbf{W}}_{\beta\beta}^{1/2} \mathbf{X}(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta}^{1/2}$. Ainda é fácil ver que $\text{E}(\mathbf{r}^*) = \mathbf{0}$. Logo, é possível definir o resíduo padronizado (ou resíduo de Pearson) como

$$\begin{aligned} t_{S_\ell} &= \frac{\mathbf{r}^* - \text{E}(\mathbf{r}^*)}{\sqrt{\text{Var}(\mathbf{r}^*)}} \\ &= \frac{\hat{\psi}_\ell^{(2,0)1/2} (y_\ell - \hat{\mu}_\ell)}{\sqrt{(1 - \hat{h}_{\ell\ell})}} \\ &= \frac{y_\ell - \hat{\mu}_\ell}{\sqrt{\hat{\psi}_\ell^{(2,0)-1} (1 - \hat{h}_{\ell\ell})}}, \end{aligned} \tag{3.8}$$

em que $\ell = 1, \dots, n$ e $\hat{h}_{\ell\ell}$ corresponde ao ℓ -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}} = \hat{\mathbf{W}}_{\beta\beta}^{1/2} \mathbf{X}(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta}^{1/2}$.

Resíduo Componente do Desvio

Em modelos lineares generalizados é muito comum avaliar a qualidade do ajuste por meio da função desvio (McCullagh e Nelder, 1989; Paula, 2012),

que corresponde à distância, para cada observação, entre o logaritmo da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo sob investigação (com p parâmetros), avaliado na estimativa de máxima verossimilhança $\hat{\boldsymbol{\mu}}$. A função desvio nos MLGs é dada então por

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\{\ell(\mathbf{y}; \mathbf{y}) - \ell(\hat{\boldsymbol{\mu}}; \mathbf{y})\},$$

em que $\ell(\mathbf{y}; \mathbf{y})$ e $\ell(\hat{\boldsymbol{\mu}}; \mathbf{y})$ correspondem aos logaritmos da função de verossimilhança do modelo saturado e investigado, respectivamente.

A função desvio na classe dos modelos lineares generalizados com superdispersão assume a mesma expressão da função desvio para os MLGs, onde somente a média é ajustada, com ϕ_ℓ no lugar de ϕ . O desvio fica agora definido por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \boldsymbol{\phi}) = \sum_{\ell=1}^n d^{*2}(y_\ell; \hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell)$, em que $d^{*2}(y_\ell; \hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) = 2\{\ell_\ell(y_\ell, \hat{\phi}_\ell) - \ell_\ell(\hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell)\}$. Na prática, deve-se substituir ϕ_ℓ por $\hat{\phi}_\ell = h^{-1}(\hat{\tau}_\ell) = s_\ell^\top \hat{\boldsymbol{\gamma}}$.

A seguir, são apresentados dois tipos de resíduos construídos a partir dos componentes da função desvio.

O resíduo componente do desvio é definido como a raiz quadrada da diferença entre os logaritmos das funções de verossimilhança sob o modelo saturado e o modelo investigado para cada observação ℓ , com o mesmo sinal da diferença ($y_\ell - \hat{\boldsymbol{\mu}}_\ell$). Assim, o resíduo componente do desvio para os modelos lineares generalizados com superdispersão, utilizado como uma medida de qualidade de ajuste, fica dado por

$$t_{D_\ell} = d^*(y_\ell; \hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell),$$

em que $d^*(y_\ell; \hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) = \pm \sqrt{d^{*2}(y_\ell; \hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell)}$.

Ou seja,

$$\begin{aligned} t_{D_\ell} &= \text{sinal}(y_\ell - \hat{\boldsymbol{\mu}}_\ell) \sqrt{2} \left\{ \ell_\ell(y_\ell, \hat{\phi}_\ell) - \ell_\ell(\hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) \right\}^{1/2} \\ &= \pm \sqrt{2} \left\{ \ell_\ell(y_\ell, \hat{\phi}_\ell) - \ell_\ell(\hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) \right\}^{1/2} \\ &= \pm \sqrt{2} \left\{ \sum_{\ell=1}^n \psi(y_\ell, \hat{\phi}_\ell) - \psi(\hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) - (y_\ell - \hat{\boldsymbol{\mu}}_\ell) \psi^{(1,0)}(\hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) \right\}^{1/2}, \end{aligned} \tag{3.9}$$

em que $\psi^{(1,0)}(\hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) = \partial \psi(\hat{\boldsymbol{\mu}}_\ell, \hat{\phi}_\ell) / \partial \hat{\boldsymbol{\mu}}_\ell$, $\ell = 1, \dots, n$.

Com isso, pode-se afirmar que o resíduo t_{D_ℓ} representa uma distância entre a observação y_ℓ e seu valor ajustado $\hat{\mu}_\ell$, medida na escala do logaritmo da função de verossimilhança.

Resíduo Componente do Desvio Padronizado

O resíduo componente do desvio padronizado, como o nome sugere, nada mais é do que uma padronização do resíduo dado em (3.9). Assim, os resíduos componentes do desvio padronizados são definidos como

$$t'_{D_\ell} = \frac{t_{D_\ell}}{\sqrt{1 - \hat{h}_{\ell\ell}}},$$

em que $\ell = 1, \dots, n$ e $\hat{h}_{\ell\ell}$ corresponde ao ℓ -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}} = \hat{\mathbf{W}}_{\beta\beta}^{1/2} \mathbf{X} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta}^{1/2}$.

É possível também utilizar a padronização por meio de $GL^\mu(\hat{\boldsymbol{\theta}})_{\ell\ell}$, que corresponde ao ℓ -ésimo elemento da diagonal principal da matriz de alavancagem generalizada para $\boldsymbol{\mu}$ dada em (3.4).

Esses resíduos são os mais utilizados nas aplicações dos MLGs, onde diversas técnicas analíticas e gráficas podem ser utilizadas para detectar desvios do modelo pesquisado, uma vez de posse destes resíduos, onde possivelmente foi definida uma distribuição teórica adequada para eles. Gráficos de t'_{D_ℓ} versus índice e t'_{D_ℓ} contra os valores ajustados são recomendados.

3.3 Influência Global

Um tópico importante na análise de diagnóstico é a detecção de pontos influentes, ou seja, aqueles que exercem um peso desproporcional nas estimativas dos parâmetros do modelo, podendo alterar o valor das estimativas quando os retiramos do conjunto de dados ou quando são submetidos a perturbações. Uma técnica bastante conhecida para avaliar o impacto da retirada de uma observação particular nas estimativas dos parâmetros de regressão é a deleção de pontos, introduzida por Cook (1977) e conhecida também como influência global.

Seja $\ell(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Para avaliar a sensibilidade das estimativas dos parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ quando a ℓ -ésima observação é excluída, é comum usar a medida de influência LD_ℓ , que

mede o afastamento da verossimilhança, permitindo detectar possíveis pontos influentes, e é denotada por

$$LD_\ell = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{(\ell)})\},$$

em que $\hat{\boldsymbol{\theta}}_{(\ell)}$ corresponde à estimativa $\hat{\boldsymbol{\theta}}$ quando a ℓ -ésima observação é excluída. A medida LD_ℓ trata-se, portanto, de uma medida de influência para avaliar o impacto em $\ell(\hat{\boldsymbol{\theta}})$ com a retirada da ℓ -ésima observação.

Quando não é possível obter uma forma analítica para LD_ℓ , é comum utilizar a segunda aproximação por série de Taylor em torno de $\hat{\boldsymbol{\theta}}$. Essa expansão fica dada por

$$LD_\ell \cong (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \{-\ddot{\mathbf{L}}_{\theta\theta}(\hat{\boldsymbol{\theta}})\}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Substituindo $-\ddot{\mathbf{L}}_{\theta\theta}(\hat{\boldsymbol{\theta}})$ pelo seu valor esperado $\mathbf{K}_{\theta,\theta}(\hat{\boldsymbol{\theta}})$ e $\boldsymbol{\theta}$ por $\hat{\boldsymbol{\theta}}_{(\ell)}$, tem-se que

$$\begin{aligned} LD_\ell^\beta &\cong (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(\ell)})^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(\ell)}) \\ &\text{e} \\ LD_\ell^\gamma &\cong (\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}_{(\ell)})^\top (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S}) (\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}_{(\ell)}) \end{aligned} \quad (3.10)$$

em que $\hat{\mathbf{W}}_{\beta\beta} = \hat{\boldsymbol{\Psi}}^{(2,0)} \hat{\mathbf{M}}_1^2$ e $\hat{\mathbf{W}}_{\gamma\gamma} = -\hat{\boldsymbol{\Psi}}^{(0,2)} \hat{\boldsymbol{\Phi}}_1^2$.

Em geral, não é possível obter uma forma fechada para $\hat{\boldsymbol{\theta}}_{(\ell)}$ e, então, aproximações têm sido utilizadas. Pregibon(1981) sugere a aproximação de um passo, que consiste em tomar a primeira iteração do processo iterativo pelo método scoring de Fisher, iniciando em $\hat{\boldsymbol{\theta}}$. Tal aproximação é dada por

$$\hat{\boldsymbol{\theta}}_{(\ell)}^1 = \hat{\boldsymbol{\theta}} + \{-\ddot{\mathbf{L}}_{\theta\theta}(\hat{\boldsymbol{\theta}})\}^{-1} \ell_{(\ell)}(\hat{\boldsymbol{\theta}}),$$

em que $\ell_{(\ell)}(\hat{\boldsymbol{\theta}})$ é o logaritmo da função de verossimilhança sem a ℓ -ésima observação. Substituindo novamente $-\ddot{\mathbf{L}}_{\theta\theta}(\hat{\boldsymbol{\theta}})$ por $\mathbf{K}_{\theta,\theta}(\hat{\boldsymbol{\theta}})$ ou, similarmente, de acordo com os cálculos do Apêndice D, é obtido

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{(\ell)} &= \hat{\boldsymbol{\beta}} - \frac{\hat{r}_\ell^\beta}{(1 - \hat{h}_{\ell\ell})} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\boldsymbol{\omega}}_{\beta\beta\ell}^{1/2} \\ &\text{e} \\ \hat{\boldsymbol{\gamma}}_{(\ell)} &= \hat{\boldsymbol{\gamma}} - \frac{\hat{r}_\ell^\gamma}{(1 - \hat{p}_{\ell\ell})} (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\boldsymbol{\omega}}_{\gamma\gamma\ell}^{1/2}, \end{aligned} \quad (3.11)$$

em que $\hat{h}_{\ell\ell}$ corresponde ao ℓ -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}} = \hat{\mathbf{W}}_{\beta\beta}^{1/2} \mathbf{X} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta}^{1/2}$ e $\hat{p}_{\ell\ell}$ corresponde ao ℓ -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{P}} = \hat{\mathbf{W}}_{\gamma\gamma}^{1/2} \mathbf{S} (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma}^{1/2}$. E ainda, define-se

$$\hat{r}_\ell^\beta = \hat{\omega}_{\beta\beta\ell}^{1/2} [\hat{\xi}_{1\ell} - \hat{\eta}_\ell] = \hat{\psi}_\ell^{(2,0)1/2} (y_\ell - \hat{\mu}_\ell) \quad \text{e} \quad \hat{r}_\ell^\gamma = \hat{\omega}_{\gamma\gamma\ell}^{1/2} [\hat{\xi}_{2\ell} - \hat{\tau}_\ell],$$

com $\hat{\xi}_{1\ell} = \hat{\eta}_\ell + \hat{m}_{1\ell}^{-1} (y_\ell - \hat{\mu}_\ell)$, sendo $\hat{\eta}_\ell = \mathbf{x}_\ell^\top \hat{\beta}$ e $\hat{\xi}_{2\ell} = \hat{\tau}_\ell - \hat{\psi}_\ell^{(0,2)-1} \hat{\phi}_{1\ell}^{-1} \hat{v}_\ell$, em que $\hat{\tau}_\ell = \mathbf{s}_\ell^\top \hat{\gamma}$ e $\hat{v}_\ell = (y_\ell - \hat{\mu}_\ell) \hat{\psi}_\ell^{(1,1)} + T(y_\ell) + \hat{\psi}_\ell^{(0,1)}$ e $\ell = 1, \dots, n$.

Substituindo (3.11) em (3.10), é obtido

$$\begin{aligned} LD_\ell^\beta &\cong \left[\frac{\hat{r}_\ell^\beta}{(1 - \hat{h}_{\ell\ell})} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell^\top \hat{\omega}_{\beta\beta\ell}^{1/2} \right] (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X}) \\ &\quad \left[\frac{\hat{r}_\ell^\beta}{(1 - \hat{h}_{\ell\ell})} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta\ell}^{1/2} \right] \\ &= \frac{(y_\ell - \hat{\mu}_\ell)^2}{\hat{\psi}_\ell^{(2,0)-1} (1 - \hat{h}_{\ell\ell})} \cdot \frac{1}{(1 - \hat{h}_{\ell\ell})} \mathbf{x}_\ell \hat{\omega}_{\beta\beta\ell}^{1/2} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell^\top \hat{\omega}_{\beta\beta\ell}^{1/2} \\ &= \frac{(y_\ell - \hat{\mu}_\ell)^2}{\hat{\psi}_\ell^{(2,0)-1} (1 - \hat{h}_{\ell\ell})} \cdot \frac{\hat{h}_{\ell\ell}}{(1 - \hat{h}_{\ell\ell})}, \end{aligned}$$

com $\hat{h}_{\ell\ell} = \hat{\omega}_{\beta\beta\ell}^{1/2} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta\ell}^{1/2}$, e

$$\begin{aligned} LD_\ell^\gamma &\cong \left[\frac{\hat{r}_\ell^\gamma}{(1 - \hat{p}_{\ell\ell})} (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell^\top \hat{\omega}_{\gamma\gamma\ell}^{1/2} \right] (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S}) \\ &\quad \left[\frac{\hat{r}_\ell^\gamma}{(1 - \hat{p}_{\ell\ell})} (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\omega}_{\gamma\gamma\ell}^{1/2} \right] \\ &= \frac{(\hat{r}_\ell^\gamma)^2}{(1 - \hat{p}_{\ell\ell})} \cdot \frac{1}{(1 - \hat{p}_{\ell\ell})} \mathbf{s}_\ell \hat{\omega}_{\gamma\gamma\ell}^{1/2} (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell^\top \hat{\omega}_{\gamma\gamma\ell}^{1/2} \\ &= \frac{(\hat{r}_\ell^\gamma)^2}{(1 - \hat{p}_{\ell\ell})} \cdot \frac{\hat{p}_{\ell\ell}}{(1 - \hat{p}_{\ell\ell})}, \end{aligned}$$

sendo $\hat{p}_{\ell\ell} = \hat{\omega}_{\gamma\gamma\ell}^{1/2} \mathbf{s}_\ell^\top (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\omega}_{\gamma\gamma\ell}^{1/2}$.

Logo,

$$LD_\ell^\beta \cong t_{S_\ell}^2 \frac{\hat{h}_{\ell\ell}}{(1 - \hat{h}_{\ell\ell})}$$

e

$$LD_\ell^\gamma \cong t_{T_\ell}^2 \frac{\hat{p}_{\ell\ell}}{(1 - \hat{p}_{\ell\ell})}$$

(3.12)

em que $t_{S_\ell}^2 = (\hat{r}_\ell^\beta)^2 / (1 - \hat{h}_{\ell\ell})$, sendo $\hat{r}_\ell^\beta = \hat{\psi}_\ell^{(2,0)^{1/2}}(y_\ell - \hat{\mu}_\ell)$ e $t_{T_\ell}^2 = (\hat{r}_\ell^\gamma)^2 / (1 - \hat{p}_{\ell\ell})$, com $\hat{r}_\ell^\gamma = \hat{\omega}_{\gamma\ell}^{1/2}[\hat{\xi}_{2\ell} - \hat{\tau}_\ell]$, $\ell = 1, \dots, n$. O resíduo t_{S_ℓ} também encontra-se definido em (3.8). Aqui, é adotada a aproximação definida em (3.12) como a distância de Cook para a classe dos modelos lineares generalizados com superdispersão.

A aproximação de um passo, em geral, subestima o verdadeiro valor de LD_ℓ , no entanto é suficiente para chamar a atenção de pontos aberrantes e influentes. Gráficos de LD_ℓ^β e LD_ℓ^γ contra os índices das observações e de $\hat{h}_{\ell\ell}$ e $\hat{p}_{\ell\ell}$ contra os valores ajustados são recomendados.

3.4 Influência Local

A deleção de pontos é uma das técnicas de diagnóstico mais utilizadas para avaliar a retirada de uma observação individualmente. Contudo, tal método pode deixar de detectar observações que sejam conjuntamente discrepantes. Assim, a metodologia de influência local proposta por Cook (1986) tem como ideia principal avaliar a influência conjunta de pontos sob certas perturbações nos dados ou no modelo, sem com isso necessitar a avaliação da retirada de uma ou um conjunto de observações.

Seja $\ell(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança, onde $\boldsymbol{\theta}$ é um vetor $(p + q) \times 1$ de parâmetros desconhecidos do modelo. É possível introduzir um vetor de perturbação $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$, restrito a algum subconjunto Ω do \mathbb{R}^n , e comparar os estimadores de máxima verossimilhança para se determinar influência local, a fim de avaliar os efeitos das perturbações nas estimativas fornecidas pelo modelo. O logaritmo da função de verossimilhança do modelo perturbado possui então a forma $\ell(\boldsymbol{\theta}_\omega)$.

A fim de avaliar a influência das perturbações na estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$, considere o afastamento da verossimilhança, também conhecido como *desvio local*, denotado aqui por $LD(\boldsymbol{\omega})$ e definido da seguinte forma

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

em que $LD(\boldsymbol{\omega}) \geq 0$, $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança do modelo não perturbado e $\hat{\boldsymbol{\theta}}_\omega$ do perturbado, isto é, sob o modelo $\ell(\boldsymbol{\theta}_\omega)$. Ou seja, $LD(\boldsymbol{\omega})$ permite calcular a diferença das verossimilhanças avaliadas em seus

estimadores ($\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$). Valores altos de $LD(\boldsymbol{\omega})$ indicam que as estimações são altamente sensíveis à perturbação.

O método proposto por Cook (1986) sugere estudar o comportamento do afastamento da verossimilhança $LD(\boldsymbol{\omega})$ em torno do vetor de não perturbação $\boldsymbol{\omega}_0$, em que $\ell(\boldsymbol{\theta}_{\boldsymbol{\omega}_0}) = \ell(\boldsymbol{\theta})$. Em particular, tem-se que $LD(\boldsymbol{\omega}_0) = 0$. Desde que $LD(\boldsymbol{\omega}) \geq 0$, $\boldsymbol{\omega}_0$ é um ponto de mínimo da função $LD(\boldsymbol{\omega})$. Cook (1986) então propõe investigar a curvatura normal da linha projetada $LD(\boldsymbol{\omega}_0 + a\mathbf{d})$, em que $a \in \mathbb{R}$ e \mathbf{d} é um vetor fixo não nulo unitário, ou seja, de comprimento 1 ($\|\mathbf{d}\| = 1$). Cada linha projetada pode ser caracterizada por uma curvatura normal $C_d(\boldsymbol{\theta})$, em torno de $a = 0$, para alguma direção arbitrária \mathbf{d} .

Cook (1986) usa conceitos de geometria diferencial para mostrar que a curvatura normal na direção \mathbf{d} assume a forma

$$C_d(\boldsymbol{\theta}) = 2|\mathbf{d}^\top \boldsymbol{\Delta}^\top \ddot{\mathbf{L}}_{\theta\theta}^{-1} \boldsymbol{\Delta} \mathbf{d}|,$$

em que $-\ddot{\mathbf{L}}_{\theta\theta}$ é a matriz de informação observada de Fisher e $\boldsymbol{\Delta}$ é uma matriz $(p+q) \times n$ que depende do esquema de perturbação, com elementos dados por $\Delta_{ij} = \partial^2 \ell(\boldsymbol{\theta}_{\boldsymbol{\omega}}) / \partial \theta_i \partial \omega_j$, com $i = 1, \dots, p+q$, $j = 1, \dots, n$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

Uma sugestão de Cook (1986) é utilizar \mathbf{d}_{max} , que é o autovetor correspondente ao maior autovalor $C_{d_{max}}$ da matriz $\ddot{\mathbf{F}} = \boldsymbol{\Delta}^\top (-\ddot{\mathbf{L}}_{\theta\theta})^{-1} \boldsymbol{\Delta}$. O gráfico do vetor $|\mathbf{d}_{max}|$ contra a ordem das observações pode revelar quais pontos que, sob pequenas perturbações, exercem uma influência desproporcional em $LD(\boldsymbol{\omega})$, na vizinhança de $\boldsymbol{\omega}_0$. Tais observações podem ser responsáveis por alterações consideráveis nas estimativas dos parâmetros sob pequenas perturbações nos dados ou no modelo.

Escobar e Meeker (1992) sugerem tomar como medida de influência os elementos da diagonal principal da matriz $\ddot{\mathbf{F}} = \boldsymbol{\Delta}^\top (-\ddot{\mathbf{L}}_{\theta\theta})^{-1} \boldsymbol{\Delta}$, enquanto Lesaffre e Verbeke (1998) propõem avaliar a curvatura normal na direção da i -ésima observação, ou seja, avaliar $C_d(\boldsymbol{\theta})$ no vetor d_i de dimensão $n \times 1$, composto de um na i -ésima posição e zero nas demais. Tal curvatura é denotada por C_i , em que $C_i = 2|f_{ii}|$, sendo f_{ii} os elementos da diagonal principal da matriz $\ddot{\mathbf{F}}$. Um gráfico de C_i versus índice é sugerido, em que observações com $C_i > 2\bar{C}$, $\bar{C} = \sum_{j=1}^n C_j/n$, merecem atenção especial.

A seguir, serão apresentados três tipos de esquema de perturbação, considerando o modelo definido em (2.1), com o logaritmo da função de verossi-

milhança dado em (2.3). Para cada esquema de perturbação, é de interesse obter a matriz Δ definida por

$$\Delta = \left(\frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \right).$$

A matriz Δ será então apresentada sob a forma

$$\Delta = \begin{pmatrix} \Delta_\beta \\ \Delta_\gamma \end{pmatrix} = \begin{pmatrix} \partial^2 \ell(\boldsymbol{\theta}_\omega) / \partial \beta \partial \boldsymbol{\omega}^\top \\ \partial^2 \ell(\boldsymbol{\theta}_\omega) / \partial \gamma \partial \boldsymbol{\omega}^\top \end{pmatrix},$$

em que os elementos $\Delta_{ij} = \partial^2 \ell(\boldsymbol{\theta}_\omega) / \partial \theta_i \partial \omega_j$, com $i = 1, \dots, p+q$, $j = 1, \dots, n$, são avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

3.4.1 Perturbação aditiva na resposta

Considere uma perturbação aditiva sobre a ℓ -ésima resposta, ou seja,

$$y_{\ell\omega} = y_\ell + \omega_\ell s(y_\ell),$$

em que $s(y_\ell)$, $\ell = 1, \dots, n$, é um fator de escala, comumente utilizado para padronizar os componentes de $\boldsymbol{\omega}$ quando cada y_ℓ apresenta uma variância diferente. Aqui, o fator de escala é a estimativa do desvio padrão de y_ℓ . Assim, o logaritmo da função de verossimilhança do modelo perturbado assume a forma

$$\begin{aligned} \ell(\boldsymbol{\theta}_\omega) &= \sum_{\ell=1}^n \{ (y_{\ell\omega} - \mu_\ell) \psi^{(1,0)}(\mu_\ell, \phi_\ell) + \phi_\ell T(y_{\ell\omega}) + \psi(\mu_\ell, \phi_\ell) \} \\ &\quad + \sum_{\ell=1}^n \log A(y_{\ell\omega}). \end{aligned}$$

Ao derivar $\ell(\boldsymbol{\theta}_\omega)$ em relação a β_s , tem-se que

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s} = \sum_{\ell=1}^n \psi_\ell^{(2,0)} (y_{\ell\omega} - \mu_\ell) m_{1\ell}(s)_\ell,$$

em que $\psi_\ell^{(2,0)} = \partial^2 \psi(\mu_\ell, \phi_\ell) / \partial \mu_\ell^2$, $m_{1\ell} = \partial \mu_\ell / \partial \eta_\ell$ e $(s)_\ell = \partial \eta_\ell / \partial \beta_s$.

Derivando novamente com respeito a ω_ℓ , obtém-se

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s \partial \omega_\ell} = \sum_{\ell=1}^n \psi_\ell^{(2,0)} m_{1\ell}(s)_\ell s(y_\ell),$$

com $s = 1, \dots, p$ e $\ell = 1, \dots, n$.

Analogamente para γ , ao derivar $\ell(\boldsymbol{\theta}_\omega)$ em relação a γ_S , é obtido

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_S} = \sum_{\ell=1}^n \phi_{1\ell}(S)_\ell v_\ell,$$

em que $\phi_{1\ell} = \partial \phi_\ell / \partial \tau_\ell$, $(S)_\ell = \partial \tau_\ell / \partial \gamma_S$ e $v_\ell = \psi_\ell^{(1,1)}(y_{\ell\omega} - \mu_\ell) + T(y_{\ell\omega}) + \psi_\ell^{(0,1)}$, com $\psi_\ell^{(1,1)} = \partial^2 \psi(\mu_\ell, \phi_\ell) / \partial \mu_\ell \partial \phi_\ell$ e $\psi_\ell^{(0,1)} = \partial \psi(\mu_\ell, \phi_\ell) / \partial \phi_\ell$.

Derivando agora em relação a ω_ℓ , tem-se que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_S \partial \omega_\ell} = \sum_{\ell=1}^n \phi_{1\ell}(S)_\ell z_\ell,$$

em que $S = 1, \dots, q$, $\ell = 1, \dots, n$, $z_\ell = \psi_\ell^{(1,1)} s(y_\ell) + T'(y_{\ell\omega})$, com $T'(y_{\ell\omega}) = \partial T(y_{\ell\omega}) / \partial \omega_\ell$.

Para a perturbação aditiva na resposta, $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$ e a matriz $\hat{\boldsymbol{\Delta}}$ fica da forma

$$\begin{pmatrix} \hat{\boldsymbol{\Delta}}_\beta \\ \hat{\boldsymbol{\Delta}}_\gamma \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\top \hat{\boldsymbol{\Psi}}^{(2,0)} \hat{\mathbf{M}}_1 \mathbf{D}(\mathbf{s}) \\ \mathbf{S}^\top \hat{\boldsymbol{\Phi}}_1 \hat{\mathbf{z}} \end{pmatrix},$$

em que $\hat{\mathbf{z}} = \hat{\boldsymbol{\Psi}}^{(1,1)} \mathbf{D}(\mathbf{s}) + \mathbf{D}(\mathbf{T}'_w)$, sendo $\mathbf{D}(\mathbf{s}) = \text{diag}\{s(y_1), \dots, s(y_n)\}$ e $\mathbf{D}(\mathbf{T}'_w) = \text{diag}\{T'(y_{1\omega}), \dots, T'(y_{n\omega})\}$. Os componentes das matrizes diagonais $\boldsymbol{\Psi}^{(2,0)}$, \mathbf{M}_1 , $\boldsymbol{\Phi}_1$ e $\boldsymbol{\Psi}^{(1,1)}$ encontram-se definidos em (2.10).

Perturbações aditivas na variável resposta estão fortemente relacionadas com o conceito de alavanca apresentado na Seção 3.1. Então, ao impor uma mudança aditiva na resposta, a medida de influência local resultante (\mathbf{d}_{max} ou C_i) pode ser utilizada para identificar observações que exercem forte influência no próprio valor ajustado.

3.4.2 Perturbação de casos ponderados

Com esta perturbação é possível avaliar se a contribuição das observações com ponderações afetam o estimador de máxima verossimilhança. O logaritmo da função de verossimilhança do modelo perturbado é dado por

$$\begin{aligned}
 \ell(\boldsymbol{\theta}_\omega) &= \sum_{\ell=1}^n \omega_\ell \ell_\ell(y_\ell; \boldsymbol{\theta}) \\
 &= \sum_{\ell=1}^n \omega_\ell \{ (y_\ell - \mu_\ell) \psi^{(1,0)}(\mu_\ell, \phi_\ell) + \phi_\ell T(y_\ell) + \psi(\mu_\ell, \phi_\ell) \} \\
 &\quad + \sum_{\ell=1}^n \omega_\ell \log A(y_\ell),
 \end{aligned}$$

em que $0 \leq \omega_\ell \leq 1$.

Derivando $\ell(\boldsymbol{\theta}_\omega)$ com relação a β_s , tem-se que

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s} = \sum_{\ell=1}^n \omega_\ell \{ \psi_\ell^{(2,0)}(y_\ell - \mu_\ell) m_{1\ell}(s)_l \},$$

em que $\psi_\ell^{(2,0)} = \partial^2 \psi(\mu_\ell, \phi_\ell) / \partial \mu_\ell^2$, $m_{1\ell} = \partial \mu_\ell / \partial \eta_\ell$ e $(s)_\ell = \partial \eta_\ell / \partial \beta_s$.

Portanto,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s \partial \omega_\ell} = \sum_{\ell=1}^n \psi_\ell^{(2,0)}(y_\ell - \mu_\ell) m_{1\ell}(s)_l,$$

com $s = 1, \dots, p$ e $\ell = 1, \dots, n$.

Para $\boldsymbol{\Delta}_\gamma$, tem-se que a derivada de $\ell(\boldsymbol{\theta}_\omega)$ em relação a γ_S é dada por

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_S} = \sum_{\ell=1}^n \omega_\ell \phi_{1\ell}(S)_\ell q_\ell,$$

em que $\phi_{1\ell} = \partial \phi_\ell / \partial \tau_\ell$, $(S)_\ell = \partial \tau_\ell / \partial \gamma_S$ e $q_\ell = \psi_\ell^{(1,1)}(y_\ell - \mu_\ell) + T(y_\ell) + \psi_\ell^{(0,1)}$, com $\psi_\ell^{(1,1)} = \partial^2 \psi(\mu_\ell, \phi_\ell) / \partial \mu_\ell \partial \phi_\ell$ e $\psi_\ell^{(0,1)} = \partial \psi(\mu_\ell, \phi_\ell) / \partial \phi_\ell$.

Assim, o $(S\ell)$ -ésimo elemento de $\boldsymbol{\Delta}_\gamma$ fica dado por

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_S \partial \omega_\ell} = \sum_{\ell=1}^n \phi_{1\ell}(S)_\ell q_\ell,$$

$S = 1, \dots, q$ e $\ell = 1, \dots, n$.

Para a perturbação de casos ponderados, $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$ e a matriz $\hat{\boldsymbol{\Delta}}$ assume a forma

$$\begin{pmatrix} \hat{\boldsymbol{\Delta}}_\beta \\ \hat{\boldsymbol{\Delta}}_\gamma \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\top \hat{\boldsymbol{\Psi}}^{(2,0)} \hat{\mathbf{M}}_1 \mathbf{D}(\mathbf{r}) \\ \mathbf{S}^\top \hat{\boldsymbol{\Phi}}_1 \hat{\mathbf{q}} \end{pmatrix},$$

em que $\hat{\mathbf{q}} = \hat{\Psi}^{(1,1)} \mathbf{D}(\mathbf{r}) + \mathbf{D}(\mathbf{T}) + \hat{\Psi}_\ell^{(0,1)}$, com $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$, sendo $r_\ell = (y_\ell - \hat{\mu}_\ell)$, $\ell = 1, \dots, n$, e $\mathbf{D}(\mathbf{T}) = \text{diag}\{T(y_1), \dots, T(y_n)\}$. Os componentes das matrizes diagonais $\Psi^{(2,0)}$, $\Psi^{(1,1)}$, $\Psi^{(0,1)}$, \mathbf{M}_1 e Φ_1 estão definidos em (2.10).

3.4.3 Perturbação aditiva no preditor

Os modelos lineares generalizados com superdispersão possuem duas componentes sistemáticas $f(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ e $g(\boldsymbol{\phi}) = \boldsymbol{\tau} = \mathbf{S}\boldsymbol{\gamma}$, em que \mathbf{X} e \mathbf{S} são matrizes $n \times p$ e $n \times q$, respectivamente. Assim, tanto \mathbf{X} quanto \mathbf{S} sofrem perturbação, que consiste em modificar a p -ésima coluna de \mathbf{X} e a q -ésima coluna de \mathbf{S} , adicionando um vetor de perturbação $\boldsymbol{\omega}$ ponderado por um fator de escala s_{x_p} e s_{s_q} , que pode ser o desvio padrão da coluna modificada, fator este adotado aqui, ou ainda pode ser, por exemplo, a norma da coluna modificada. Ou seja,

$$x_{\ell p\omega} = x_{\ell p} + \omega_\ell s_{x_p} \quad \text{e} \quad s_{\ell q\omega} = s_{\ell q} + \omega_\ell s_{s_q}, \quad \ell = 1, \dots, n.$$

Nos MLGSs os parâmetros média e dispersão são modelados simultaneamente. Neste caso, será considerado nesta dissertação dois cenários para o esquema de perturbação individual de uma covariável. Primeiro, $\mathbf{X} \neq \mathbf{S}$, em que será de interesse investigar separadamente as matrizes de regressores \mathbf{X} para o modelo média e \mathbf{S} para o modelo dispersão. Posteriormente, tanto a matriz para o modelo média \mathbf{X} quanto a matriz de regressores \mathbf{S} para o modelo dispersão são iguais, ou seja, $\mathbf{X} = \mathbf{S}$. A seguir é apresentada a forma da matriz $\hat{\Delta}$ para os casos aqui considerados.

As matrizes \mathbf{X} e \mathbf{S} são diferentes

Caso 1: Matriz \mathbf{X} para o modelo média

Considere para este caso que

$$\begin{aligned} \eta_{\ell\omega} &= \beta_1 + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p\omega} + \dots + \beta_k x_{\ell k} \\ &= \beta_1 + \beta_2 x_{\ell 2} + \dots + \beta_p (x_{\ell p} + \omega_\ell s_{x_p}) + \dots + \beta_k x_{\ell k}, \end{aligned} \quad (3.13)$$

e $\mu_{\ell\omega}$ é tal que $f(\mu_{\ell\omega}) = \eta_{\ell\omega}$ e $\phi_{\ell\omega} = \phi_\ell$, $\ell = 1, \dots, n$.

O logaritmo da função de verossimilhança do modelo perturbado, neste caso, fica da forma

$$\begin{aligned} \ell(\boldsymbol{\theta}_\omega) &= \sum_{\ell=1}^n \{(y_\ell - \mu_{\ell\omega})\psi^{(1,0)}(\mu_{\ell\omega}, \phi_\ell) + \phi_\ell T(y_\ell) + \psi(\mu_{\ell\omega}, \phi_\ell)\} \\ &\quad + \sum_{\ell=1}^n \log A(y_\ell). \end{aligned}$$

Ao derivar $\ell(\boldsymbol{\theta}_\omega)$ em relação à β_s , é obtido

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s} = \sum_{\ell=1}^n \psi_{\ell\omega}^{(2,0)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega}(s)_{\ell\omega},$$

em que $\psi_{\ell\omega}^{(2,0)} = \partial^2 \psi(\mu_{\ell\omega}, \phi_\ell) / \partial \mu_{\ell\omega}^2$, $m_{1\ell\omega} = \partial \mu_{\ell\omega} / \partial \eta_{\ell\omega}$ e $(s)_{\ell\omega} = \partial \eta_{\ell\omega} / \partial \beta_s$.

Assim, derivando novamente em relação a ω_ℓ , tem-se que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s \partial \omega_\ell} &= s_{x_p} \beta_p \sum_{\ell=1}^n \{-\psi_{\ell\omega}^{(2,0)} m_{1\ell\omega}^2(s)_{\ell\omega} + \psi_{\ell\omega}^{(3,0)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega}^2(s)_{\ell\omega} \\ &\quad + \psi_{\ell\omega}^{(2,0)}(y_\ell - \mu_{\ell\omega}) m_{2\ell\omega}(s)_{\ell\omega}\} + s_{x_p} \sum_{\ell=1}^n \{\psi_{\ell\omega}^{(2,0)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega}\}, \end{aligned}$$

em que $s = 1, \dots, p$, $\ell = 1, \dots, n$, $\psi_{\ell\omega}^{(3,0)} = \partial^3 \psi(\mu_{\ell\omega}, \phi_\ell) / \partial \mu_{\ell\omega}^3$ e $m_{2\ell\omega} = \partial^2 \mu_{\ell\omega} / \partial \eta_{\ell\omega}^2$.

Analogamente para γ , ao derivar $\ell(\boldsymbol{\theta}_\omega)$ em relação a γ_S , é obtido

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_S} = \sum_{\ell=1}^n \phi_{1\ell}(S)_\ell v_\ell,$$

em que $\phi_{1\ell} = \partial \phi_\ell / \partial \tau_\ell$, $(S)_\ell = \partial \tau_\ell / \partial \gamma_S$ e $v_\ell = \psi_{\ell\omega}^{(1,1)}(y_\ell - \mu_{\ell\omega}) + T(y_\ell) + \psi_{\ell\omega}^{(0,1)}$, com $\psi_{\ell\omega}^{(1,1)} = \partial^2 \psi(\mu_{\ell\omega}, \phi_\ell) / \partial \mu_{\ell\omega} \partial \phi_\ell$ e $\psi_{\ell\omega}^{(0,1)} = \partial \psi(\mu_{\ell\omega}, \phi_\ell) / \partial \phi_\ell$.

Ao derivar novamente, agora em relação a ω_ℓ , o $(S)_\ell$ -ésimo elemento de $\hat{\Delta}_\gamma$ é dado por

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_S \partial \omega_\ell} &= s_{x_p} \beta_p \sum_{\ell=1}^n \{-\psi_{\ell\omega}^{(1,1)} \phi_{1\ell}(S)_\ell m_{1\ell\omega} + \psi_{\ell\omega}^{(2,1)}(y_\ell - \mu_{\ell\omega}) \phi_{1\ell}(S)_\ell m_{1\ell\omega} \\ &\quad + \psi_{\ell\omega}^{(1,1)} \phi_{1\ell}(S)_\ell m_{1\ell\omega}\}, \end{aligned}$$

em que $S = 1, \dots, q$, $\ell = 1, \dots, n$ e $\psi_{\ell\omega}^{(2,1)} = \partial^3 \psi(\mu_{\ell\omega}, \phi_\ell) / \partial \mu_{\ell\omega}^2 \partial \phi_\ell$.

Aqui, $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$ e a matriz $\hat{\boldsymbol{\Delta}}$ é dada por

$$\begin{pmatrix} \hat{\boldsymbol{\Delta}}_\beta \\ \hat{\boldsymbol{\Delta}}_\gamma \end{pmatrix} = \begin{pmatrix} s_{x_p} \hat{\beta}_p [\mathbf{X}^\top (\hat{\Psi}^{(2,0)} \hat{\mathbf{M}}_1^2 + \hat{\Psi}^{(3,0)} \mathbf{D}(\mathbf{r}) \hat{\mathbf{M}}_1^2 + \hat{\Psi}^{(2,0)} \mathbf{D}(\mathbf{r}) \hat{\mathbf{M}}_2)] \\ \quad \quad \quad + s_{x_p} \mathbf{F}_p \hat{\Psi}^{(2,0)} \mathbf{D}(\mathbf{r}) \hat{\mathbf{M}}_1 \\ s_{x_p} \hat{\beta}_p [\mathbf{S}^\top (-\hat{\Psi}^{(1,1)} \hat{\Phi}_1 \hat{\mathbf{M}}_1 + \hat{\Psi}^{(2,1)}) \mathbf{D}(\mathbf{r}) \hat{\Phi}_1 \hat{\mathbf{M}}_1 + \hat{\Psi}^{(1,1)} \hat{\Phi}_1 \hat{\mathbf{M}}_1] \end{pmatrix},$$

em que \mathbf{F}_p é uma matriz $p \times n$ de zeros, exceto na p -ésima linha, que é composta por uns e $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$, sendo $r_\ell = (y_\ell - \hat{\mu}_\ell)$, $\ell = 1, \dots, n$. Os componentes das matrizes diagonais $\Psi^{(2,0)}$, $\Psi^{(3,0)}$, $\Psi^{(1,1)}$, $\Psi^{(2,1)}$, \mathbf{M}_1 , \mathbf{M}_2 e Φ_1 estão definidos em (2.10).

Caso 2: Matriz S para o modelo dispersão

Considere agora que

$$\begin{aligned} \tau_{\ell\omega} &= \gamma_1 + \gamma_2 s_{\ell 2} + \dots + \gamma_q s_{\ell q\omega} + \dots + \gamma_k s_{\ell k} \\ &= \gamma_1 + \gamma_2 s_{\ell 2} + \dots + \gamma_q (s_{\ell q} + \omega_\ell s_{s_q}) + \dots + \gamma_k s_{\ell k}, \end{aligned}$$

em que $\phi_{\ell\omega}$ é tal que $g(\phi_{\ell\omega}) = \tau_{\ell\omega}$ e $\mu_{\ell\omega} = \mu_\ell$, $\ell = 1, \dots, n$.

O logaritmo da função de verossimilhança do modelo perturbado fica dado por

$$\begin{aligned} \ell(\boldsymbol{\theta}_\omega) &= \sum_{\ell=1}^n \{(y_\ell - \mu_\ell) \psi^{(1,0)}(\mu_\ell, \phi_{\ell\omega}) + \phi_{\ell\omega} T(y_\ell) + \psi(\mu_\ell, \phi_{\ell\omega})\} \\ &\quad + \sum_{\ell=1}^n \log A(y_\ell). \end{aligned}$$

Derivando $\ell(\boldsymbol{\theta}_\omega)$ em relação à β_s , é obtido

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s} = \sum_{\ell=1}^n \psi_{\ell\omega}^{(2,0)}(y_\ell - \mu_\ell) m_{1\ell}(s)_\ell,$$

em que $\psi_{\ell\omega}^{(2,0)} = \partial^2 \psi(\mu_\ell, \phi_{\ell\omega}) / \partial \mu_\ell^2$, $m_{1\ell} = \partial \mu_\ell / \partial \eta_\ell$ e $(s)_\ell = \partial \eta_\ell / \partial \beta_s$.

Ao derivar novamente em relação a ω_ℓ , é obtido

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s \partial \omega_\ell} = s_{s_q} \gamma_q \sum_{\ell=1}^n \{\psi_{\ell\omega}^{(2,1)}(y_\ell - \mu_\ell) m_{1\ell}(s)_\ell\},$$

em que $s = 1, \dots, p$, $\ell = 1, \dots, n$, $\psi_{\ell\omega}^{(2,1)} = \partial^3\psi(\mu_\ell, \phi_{\ell\omega})/\partial\mu_\ell^2\partial\phi_{\ell\omega}$.

Agora para γ , derivando $\ell(\boldsymbol{\theta}_\omega)$ em relação a γ_S , tem-se que

$$\frac{\partial\ell(\boldsymbol{\theta}_\omega)}{\partial\gamma_S} = \sum_{\ell=1}^n \phi_{1\ell\omega}(S)_{\ell\omega} v_\ell,$$

em que $\phi_{1\ell\omega} = \partial\phi_{\ell\omega}/\partial\tau_{\ell\omega}$, $(S)_{\ell\omega} = \partial\tau_{\ell\omega}/\partial\gamma_S$ e $v_\ell = \psi_{\ell\omega}^{(1,1)}(y_\ell - \mu_\ell) + T(y_\ell) + \psi_{\ell\omega}^{(0,1)}$, com $\psi_{\ell\omega}^{(1,1)} = \partial^2\psi(\mu_\ell, \phi_{\ell\omega})/\partial\mu_\ell\partial\phi_{\ell\omega}$ e $\psi_{\ell\omega}^{(0,1)} = \partial\psi(\mu_\ell, \phi_{\ell\omega})/\partial\phi_{\ell\omega}$.

Derivando novamente em relação a ω_ℓ , tem-se que o $(S\ell)$ elemento de $\hat{\Delta}_\gamma$ fica da seguinte forma

$$\begin{aligned} \frac{\partial^2\ell(\boldsymbol{\theta}_\omega)}{\partial\gamma_S\partial\omega_\ell} &= s_{s_q}\gamma_q \sum_{\ell=1}^n \{ \psi_{\ell\omega}^{(1,2)}(y_\ell - \mu_\ell)\phi_{1\ell\omega}^2(S)_{\ell\omega} \\ &\quad + \psi_{\ell\omega}^{(1,1)}(y_\ell - \mu_\ell)\phi_{2\ell\omega}(S)_{\ell\omega} + T(y_\ell)\phi_{2\ell\omega}(S)_{\ell\omega} \\ &\quad + \psi_{\ell\omega}^{(0,2)}\phi_{1\ell\omega}^2(S)_{\ell\omega} + \psi_{\ell\omega}^{(0,1)}\phi_{2\ell\omega}(S)_{\ell\omega} \} \\ &\quad + s_{s_q} \sum_{\ell=1}^n \{ \psi_{\ell\omega}^{(1,1)}(y_\ell - \mu_\ell)\phi_{1\ell\omega} + T(y_\ell)\phi_{1\ell\omega} + \psi_{\ell\omega}^{(0,1)}\phi_{1\ell\omega} \}, \end{aligned}$$

em que $S = 1, \dots, q$, $\ell = 1, \dots, n$, $\psi_{\ell\omega}^{(1,2)} = \partial^3\psi(\mu_\ell, \phi_{\ell\omega})/\partial\mu_\ell\partial\phi_{\ell\omega}^2$ e $\psi_{\ell\omega}^{(0,2)} = \partial^2\psi(\mu_\ell, \phi_{\ell\omega})/\partial\phi_{\ell\omega}^2$.

Para este caso, $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$ e a matriz $\hat{\Delta}$ é dada por

$$\begin{pmatrix} \hat{\Delta}_\beta \\ \hat{\Delta}_\gamma \end{pmatrix} = \begin{pmatrix} s_{s_q}\hat{\gamma}_q[\mathbf{X}^\top\hat{\mathbf{M}}_1\hat{\boldsymbol{\Phi}}_1\hat{\boldsymbol{\Psi}}^{(2,1)}\mathbf{D}(\mathbf{r})] \\ s_{s_q}\hat{\gamma}_q[\mathbf{S}^\top(\hat{\boldsymbol{\Psi}}^{(1,2)}\mathbf{D}(\mathbf{r})\hat{\boldsymbol{\Phi}}_1^2 + \hat{\boldsymbol{\Psi}}^{(1,1)}\mathbf{D}(\mathbf{r})\hat{\boldsymbol{\Phi}}_2 \\ + \mathbf{D}(\mathbf{T})\hat{\boldsymbol{\Phi}}_2 + \hat{\boldsymbol{\Psi}}^{(0,2)}\hat{\boldsymbol{\Phi}}_1^2 + \hat{\boldsymbol{\Psi}}^{(0,1)}\hat{\boldsymbol{\Phi}}_2)] \\ + s_{s_q}\mathbf{F}_q\hat{\boldsymbol{\Phi}}_1[\hat{\boldsymbol{\Psi}}^{(1,1)}\mathbf{D}(\mathbf{r}) + \mathbf{D}(\mathbf{T}) + \hat{\boldsymbol{\Psi}}^{(0,1)}] \end{pmatrix},$$

em que \mathbf{F}_q é uma matriz $q \times n$ de zeros, exceto na q -ésima linha, que é composta por uns, $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$, sendo $r_\ell = (y_\ell - \hat{\mu}_\ell)$, $\ell = 1, \dots, n$, e $\mathbf{D}(\mathbf{T}) = \text{diag}\{T(y_1), \dots, T(y_n)\}$. Os componentes das matrizes diagonais $\boldsymbol{\Psi}^{(2,1)}$, $\boldsymbol{\Psi}^{(1,2)}$, $\boldsymbol{\Psi}^{(1,1)}$, $\boldsymbol{\Psi}^{(0,2)}$, $\boldsymbol{\Psi}^{(0,1)}$, \mathbf{M}_1 , $\boldsymbol{\Phi}_1$ e $\boldsymbol{\Phi}_2$ estão definidos em (2.10).

As matrizes \mathbf{X} e \mathbf{S} são iguais

Neste caso, $\eta_{\ell\omega}$ é dado segundo (3.13) e

$$\begin{aligned}\tau_{\ell\omega} &= \gamma_1 + \gamma_2 x_{\ell 2} + \dots + \gamma_p x_{\ell p} + \dots + \gamma_k x_{\ell k} \\ &= \gamma_1 + \gamma_2 x_{\ell 2} + \dots + \gamma_p (x_{\ell p} + \omega_\ell s_{x_p}) + \dots + \gamma_k x_{\ell k},\end{aligned}$$

em que $\mu_{\ell\omega}$ é tal que $f(\mu_{\ell\omega}) = \eta_{\ell\omega}$ e $\phi_{\ell\omega}$ é tal que $g(\phi_{\ell\omega}) = \tau_{\ell\omega}$, $\ell = 1, \dots, n$.

O logaritmo da função de verossimilhança do modelo perturbado para a perturbação aditiva no preditor considerando o caso em que $\mathbf{X} = \mathbf{S}$ é dado por

$$\begin{aligned}\ell(\boldsymbol{\theta}_\omega) &= \sum_{\ell=1}^n \{(y_\ell - \mu_{\ell\omega})\psi^{(1,0)}(\mu_{\ell\omega}, \phi_{\ell\omega}) + \phi_{\ell\omega} T(y_\ell) + \psi(\mu_{\ell\omega}, \phi_{\ell\omega})\} \\ &\quad + \sum_{\ell=1}^n \log A(y_\ell).\end{aligned}$$

A primeira derivada de $\ell(\boldsymbol{\theta}_\omega)$ em relação a β_s é dada por

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s} = \sum_{\ell=1}^n \psi_{\ell\omega}^{(2,0)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega}(s)_{\ell\omega},$$

em que $\psi_{\ell\omega}^{(2,0)} = \partial^2 \psi(\mu_{\ell\omega}, \phi_{\ell\omega}) / \partial \mu_{\ell\omega}^2$, $m_{1\ell\omega} = \partial \mu_{\ell\omega} / \partial \eta_{\ell\omega}$ e $(s)_{\ell\omega} = \partial \eta_{\ell\omega} / \partial \beta_s$.

Derivando novamente em relação a ω_ℓ , é obtido

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \beta_s \partial \omega_\ell} &= s_{x_p} \sum_{\ell=1}^n \{-\beta_p \psi_{\ell\omega}^{(2,0)} m_{1\ell\omega}^2(s)_{\ell\omega} + \psi_{\ell\omega}^{(3,0)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega}^2(s)_{\ell\omega} \\ &\quad + \gamma_p \psi_{\ell\omega}^{(2,1)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega} \phi_{1\ell\omega}(s)_{\ell\omega} + \beta_p \psi_{\ell\omega}^{(2,0)}(y_\ell - \mu_{\ell\omega}) m_{2\ell\omega}(s)_{\ell\omega}\} \\ &\quad + s_{x_p} \sum_{\ell=1}^n \{\psi_{\ell\omega}^{(2,0)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega}\},\end{aligned}$$

em que $s = 1, \dots, p$, $\ell = 1, \dots, n$, $\psi_{\ell\omega}^{(3,0)} = \partial^3 \psi(\mu_{\ell\omega}, \phi_{\ell\omega}) / \partial \mu_{\ell\omega}^3$, $\psi_{\ell\omega}^{(2,1)} = \partial^2 \psi(\mu_{\ell\omega}, \phi_{\ell\omega}) / \partial \mu_{\ell\omega}^2 \partial \phi_{\ell\omega}$, $\phi_{1\ell\omega} = \partial \phi_{\ell\omega} / \partial \tau_{\ell\omega}$ e $m_{2\ell\omega} = \partial^2 \mu_{\ell\omega} / \partial \eta_{\ell\omega}^2$.

Seguindo o mesmo raciocínio para γ , derivando $\ell(\boldsymbol{\theta}_\omega)$ em relação a γ_s , tem-se que

$$\frac{\partial \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_s} = \sum_{\ell=1}^n \phi_{1\ell\omega}(S)_{\ell\omega} v_\ell,$$

em que $\phi_{1\ell\omega} = \partial \phi_{\ell\omega} / \partial \tau_{\ell\omega}$, $(S)_{\ell\omega} = \partial \tau_{\ell\omega} / \partial \gamma_s$ e $v_\ell = \psi_{\ell\omega}^{(1,1)}(y_\ell - \mu_{\ell\omega}) + T(y_\ell) + \psi_{\ell\omega}^{(0,1)}$, com $\psi_{\ell\omega}^{(1,1)} = \partial^2 \psi(\mu_{\ell\omega}, \phi_{\ell\omega}) / \partial \mu_{\ell\omega} \partial \phi_{\ell\omega}$ e $\psi_{\ell\omega}^{(0,1)} = \partial \psi(\mu_{\ell\omega}, \phi_{\ell\omega}) / \partial \phi_{\ell\omega}$.

Assim, o $(S\ell)$ -ésimo elemento de $\hat{\Delta}_\gamma$ fica dado por

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}_\omega)}{\partial \gamma_S \partial \omega_\ell} = & s_{x_p} \sum_{\ell=1}^n \{ \beta_p \psi_{\ell\omega}^{(2,1)}(y_\ell - \mu_{\ell\omega}) m_{1\ell\omega} \phi_{1\ell\omega}(S)_{\ell\omega} \\ & + \gamma_p \psi_{\ell\omega}^{(1,2)}(y_\ell - \mu_{\ell\omega}) \phi_{1\ell\omega}^2(S)_{\ell\omega} \\ & - \beta_p \psi_{\ell\omega}^{(1,1)} m_{1\ell\omega} \phi_{1\ell\omega}(S)_{\ell\omega} + \gamma_p \psi_{\ell\omega}^{(1,1)}(y_\ell - \mu_{\ell\omega}) \phi_{2\ell\omega}(S)_{\ell\omega} \\ & + \psi_{\ell\omega}^{(1,1)}(y_\ell - \mu_{\ell\omega}) \phi_{1\ell\omega} + \gamma_p T(y_\ell) \phi_{2\ell\omega}(S)_{\ell\omega} \\ & + T(y_\ell) \phi_{1\ell\omega} + \beta_p \psi_{\ell\omega}^{(1,1)} m_{1\ell\omega} \phi_{1\ell\omega}(S)_{\ell\omega} \\ & + \gamma_p \psi_{\ell\omega}^{(0,2)} \phi_{1\ell\omega}^2(S)_{\ell\omega} + \gamma_p \psi_{\ell\omega}^{(0,1)} \phi_{2\ell\omega}(S)_{\ell\omega} \\ & + \psi_{\ell\omega}^{(0,1)} \phi_{1\ell\omega} \}, \end{aligned}$$

em que $S = 1, \dots, q$, $\ell = 1, \dots, n$, $\psi_{\ell\omega}^{(1,2)} = \partial^3 \psi(\mu_{\ell\omega}, \phi_{\ell\omega}) / \partial \mu_{\ell\omega} \phi_{\ell\omega}^2$, $\psi_{\ell\omega}^{(0,2)} = \partial^2 \psi(\mu_{\ell\omega}, \phi_{\ell\omega}) / \partial \phi_{\ell\omega}^2$ e $\phi_{2\ell\omega} = \partial^2 \phi_{\ell\omega} / \partial \tau_{\ell\omega}^2$.

Para este caso, $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$ e a matriz $\hat{\Delta}$ é dada por

$$\begin{pmatrix} \hat{\Delta}_\beta \\ \hat{\Delta}_\gamma \end{pmatrix} = \begin{pmatrix} s_{x_p} \{ \mathbf{X}^\top [-\hat{\beta}_p \hat{\Psi}^{(2,0)} \hat{\mathbf{M}}_1^2 + \hat{\Psi}^{(3,0)} \mathbf{D}(\mathbf{r}) \hat{\mathbf{M}}_1^2 + \hat{\gamma}_p \hat{\Psi}^{(2,1)} \mathbf{D}(\mathbf{r}) \hat{\mathbf{M}}_1 \hat{\Phi}_1 \\ + \hat{\beta}_p \mathbf{D}(\mathbf{r}) \hat{\Psi}^{(2,0)} \hat{\mathbf{M}}_2] + \mathbf{F}_p \mathbf{D}(\mathbf{r}) \hat{\Psi}^{(2,0)} \hat{\mathbf{M}}_1 \} \\ s_{x_p} \{ \mathbf{X}^\top [\hat{\beta}_p \hat{\Psi}^{(2,1)} \mathbf{D}(\mathbf{r}) \hat{\Phi}_1 \hat{\mathbf{M}}_1 + \hat{\gamma}_p \hat{\Psi}^{(1,2)} \mathbf{D}(\mathbf{r}) \hat{\Phi}_1^2 \\ - \hat{\beta}_p \hat{\Psi}^{(1,1)} \hat{\Phi}_1 \hat{\mathbf{M}}_1 + \hat{\gamma}_p \hat{\Psi}^{(1,1)} \mathbf{D}(\mathbf{r}) \hat{\Phi}_2 + \hat{\gamma}_p \mathbf{D}(\mathbf{T}) \hat{\Phi}_2 \\ + \hat{\beta}_p \hat{\Psi}^{(1,1)} \hat{\Phi}_1 \hat{\mathbf{M}}_1 + \hat{\gamma}_p \hat{\Psi}^{(0,2)} \hat{\Phi}_1^2 + \hat{\gamma}_p \hat{\Psi}^{(0,1)} \hat{\Phi}_2] \\ + \mathbf{F}_p [\hat{\Psi}^{(1,1)} \mathbf{D}(\mathbf{r}) \hat{\Phi}_1 + \mathbf{D}(\mathbf{T}) \hat{\Phi}_1 + \hat{\Psi}^{(0,1)} \hat{\Phi}_1] \} \end{pmatrix},$$

em que $\mathbf{D}(\mathbf{T}) = \text{diag}\{T(y_1), \dots, T(y_n)\}$ e $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$, sendo $r_\ell = (y_\ell - \hat{\mu}_\ell)$, $\ell = 1, \dots, n$, e \mathbf{F}_p é uma matriz $p \times n$ de zeros, exceto na p -ésima linha, que é composta por uns. Os componentes das matrizes diagonais $\hat{\Psi}^{(2,0)}$, $\hat{\Psi}^{(3,0)}$, $\hat{\Psi}^{(2,1)}$, $\hat{\Psi}^{(1,2)}$, $\hat{\Psi}^{(1,1)}$, $\hat{\Psi}^{(0,2)}$, $\hat{\Psi}^{(0,1)}$, $\hat{\mathbf{M}}_1$, $\hat{\mathbf{M}}_2$, $\hat{\Phi}_1$ e $\hat{\Phi}_2$ encontram-se definidos em (2.10).

Por meio do esquema de perturbação aditiva no preditor, é possível acessar a influência individual de cada covariável no processo de estimação do modelo, ou seja, ao perturbar cada covariável separadamente, pode-se estabelecer quais covariáveis causam maior mudança no deslocamento pela verossimilhança sob a adição de pequenas alterações nos seus valores originais.

Capítulo 4

Resultados de Simulação

Neste Capítulo é apresentada uma análise gráfica de dados simulados, por meio da simulação de Monte Carlo, com a finalidade de apresentar as técnicas de diagnóstico abordadas nesta dissertação. Assim, foram consideradas 10000 réplicas e tamanho da amostra fixado em $n = 50$, em que foi utilizada a distribuição passeio aleatório, distribuição esta pertencente à classe dos modelos lineares generalizados com superdispersão como visto na Seção 2.4.

Inicialmente foi gerada uma variável aleatória inversa Gaussiana, isto é, $Z_\ell \sim IG(\theta_\ell, \delta_\ell)$, e então foi utilizada sua recíproca como variável resposta (Previdelli, 2005), dado que se Z é uma variável aleatória com distribuição inversa Gaussiana, então sua recíproca $Y = 1/Z$ possui distribuição passeio aleatório (Cordeiro e Botter, 2001; Johnson et al., 1994, p.282). As componentes sistemáticas do modelo foram definidas por duas funções logarítmicas, com $p = 3$ e $q = 2$, ou seja,

$$\begin{aligned}\mu_{IG_\ell} &= \theta_\ell = \exp(\eta_\ell) = \exp(\beta_1 x_{1\ell} + \beta_2 x_{2\ell} + \beta_3 x_{3\ell}) \quad \text{e} \\ \phi_{IG_\ell} &= \delta_\ell = \exp(\tau_\ell) = \exp(\gamma_1 s_{1\ell} + \gamma_2 s_{2\ell}),\end{aligned}$$

em que $\ell = 1, \dots, 50$. Assim, $Y_\ell = 1/Z_\ell \sim \text{PA}(\mu_\ell, \phi_\ell)$, sendo $\mu_\ell = \frac{1}{\theta_\ell} + \frac{1}{\delta_\ell}$ e $\phi_\ell = -\delta_\ell/(2\theta_\ell^2)$, isto é, os parâmetros média e dispersão do modelo passeio aleatório μ_ℓ e ϕ_ℓ são funções da média e dispersão da distribuição inversa Gaussiana θ_ℓ e δ_ℓ , respectivamente (Previdelli, 2005).

Os valores iniciais dos parâmetros foram: $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$, $\gamma_1 = 1$ e $\gamma_2 = 2$ (Cordeiro e Botter, 2001). As covariadas \mathbf{X} e \mathbf{S} foram obtidas de retiradas independentes de uma distribuição uniforme padrão $U(0, 1)$, por meio da função `ranu` do software `0x`, utilizando o gerador GM (George Marsaglia).

Assim, tem-se que $\mathbf{X} \neq \mathbf{S}$. Para avaliar a qualidade da estimação dos parâmetros do modelo passeio aleatório utilizando os dados simulados, foram utilizadas as medidas de viés e viés relativo, sendo este último definido como $100 \times (\text{viés} / \text{valor verdadeiro do parâmetro})\%$, em que o valor verdadeiro do parâmetro é representado aqui pelo chute inicial. Na Tabela 4.1 são apresentadas tais medidas, podendo-se notar que há uma boa aproximação dos verdadeiros valores dos parâmetros.

Tabela 4.1: EMV, viés e viés relativo (VR), em %, dos parâmetros do modelo passeio aleatório usando os dados simulados.

Parâmetros	EMV	Viés	VR (%)
β_1	0,9706	-0,0294	-2,9403
β_2	1,9800	-0,0200	-1,0005
β_3	2,9530	-0,0470	-1,5659
γ_1	1,0799	0,0799	7,9870
γ_2	2,1572	0,1572	7,8607

4.1 Análise Gráfica

A seguir são apresentados gráficos gerados no software R a partir de dados simulados no 0x da distribuição passeio aleatório como definido acima. Tais gráficos referem-se aos conceitos de alavancagem generalizada, resíduos, influência global e influência local abordados no Capítulo 3. Ressalta-se que para as perturbações aditiva no preditor para os ajustes da média e dispersão foram perturbadas as variáveis preditoras X_2 e S_2 , respectivamente.

A partir da Figura 4.1, nota-se que as observações #25, #30, #39 e #50 aparecem com alta alavancagem, com o ponto #30 aparecendo somente nos gráficos relacionados ao ajuste da média (Figuras 4.1(a) e 4.1(b)) e as observações #25 e #50 destacadas somente nos gráficos com respeito à dispersão (Figuras 4.1(c) e 4.1(d)). Os pontos #30 e #39 ainda aparecem como aberrantes nas Figuras 4.2 e 4.3, nos gráficos que apresentam o resíduo componente do desvio e componente do desvio padronizado, respectivamente, contra o índice das observações. Nota-se ainda, pela Figura 4.3, no gráfico dos resíduos componente do desvio padronizado versus valores ajustados, que há indícios de heteroscedasticidade. Por meio da Figura 4.4, que apresenta alguns gráficos para análise de influência global vista na Seção 3.3, percebe-se alguns pontos

mais afastados dos demais. Na Figura 4.5 são apresentados os gráficos de C_i contra a ordem das observações para os esquemas de perturbação estudados, sendo eles: (a) perturbação aditiva na variável resposta, (b) perturbação aditiva no preditor para o ajuste da média (matriz \mathbf{X}), (c) perturbação aditiva no preditor para o ajuste da dispersão (matriz \mathbf{S}) e (d) perturbação de casos ponderados. Em geral, por meio da Figura 4.5, percebe-se algumas observações influentes, em que é possível destacar os pontos #27, #30, #39 e #50, sendo os três últimos também pontos de alavanca. Vale ressaltar ainda que as observações #30 e #39 aparecem tanto nas Figuras 4.1(a) e 4.1(b), que correspondem à alavancagem generalizada para μ , quanto na Figura 4.5(a), sendo esta correspondente à perturbação aditiva na variável resposta. Tal fato pode ocorrer devido à relação existente entre alavancagem generalizada e perturbação aditiva na resposta, como comentado na Seção 3.4.1.

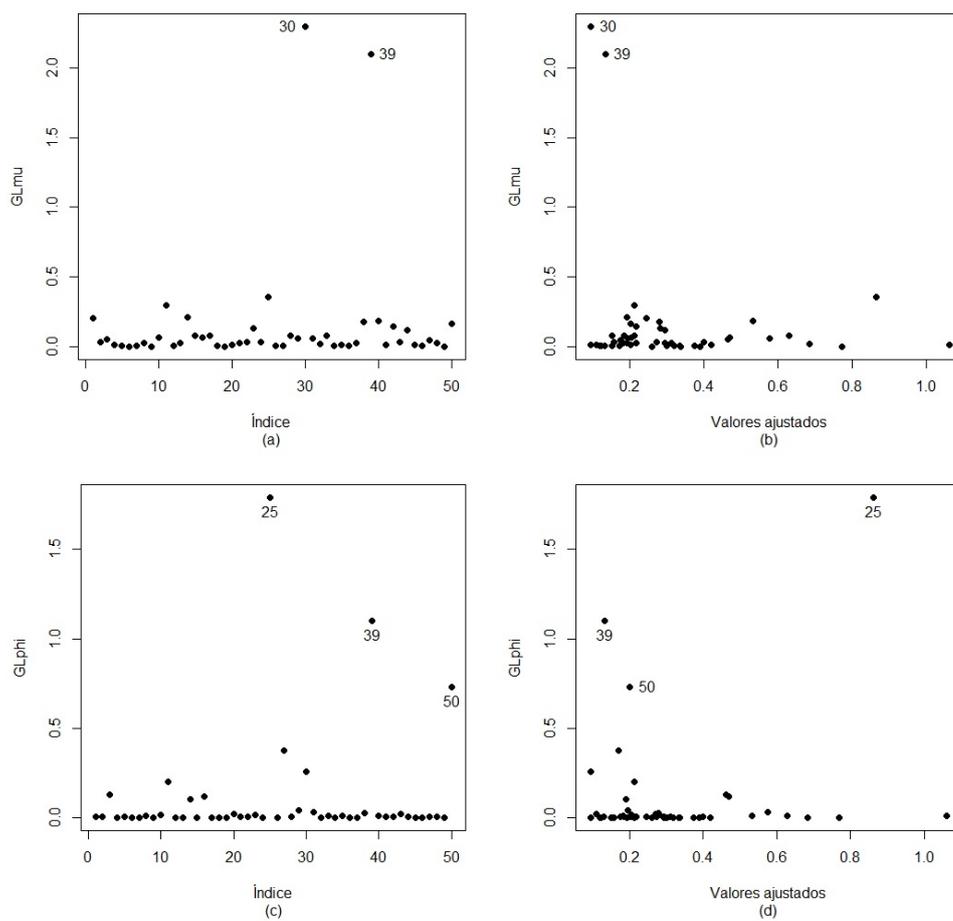


Figura 4.1: Gráficos de pontos de alavanca generalizados contra o índice das observações e os valores ajustados para os dados simulados da distribuição passeio aleatório para a média ((a) e (b)) e a dispersão ((c) e (d)).

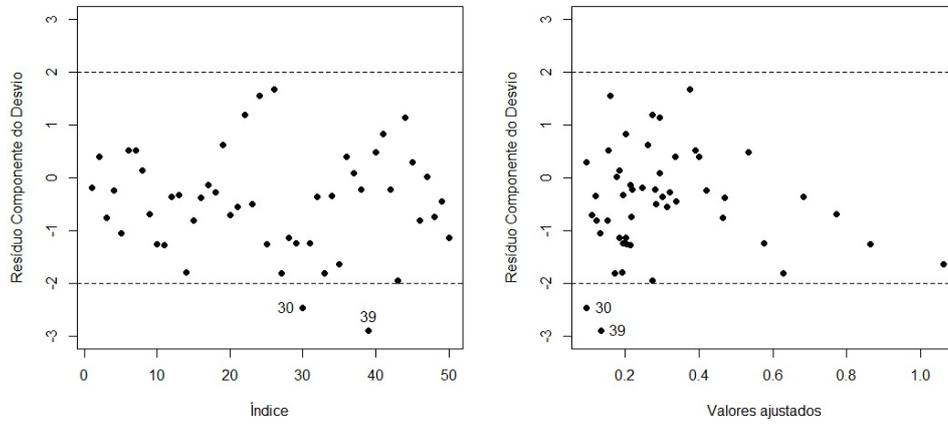


Figura 4.2: Gráficos de resíduos componente do desvio contra o índice das observações e os valores ajustados para os dados simulados da distribuição passeio aleatório.

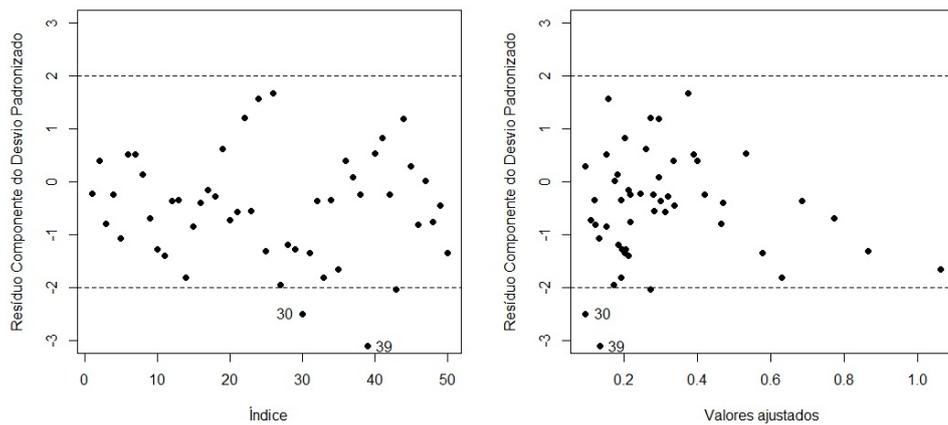


Figura 4.3: Gráficos de resíduos componente do desvio padronizado contra o índice das observações e os valores ajustados para os dados simulados da distribuição passeio aleatório.

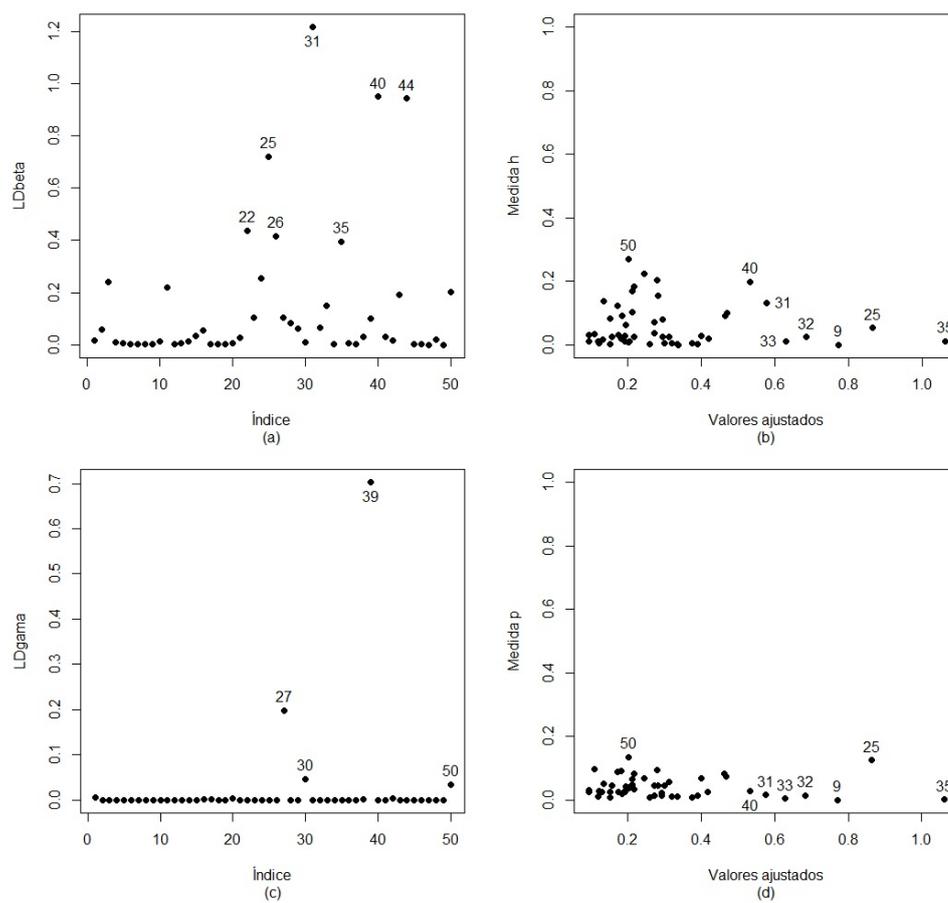


Figura 4.4: Gráficos de Influência global para os dados simulados da distribuição passeio aleatório para a média ((a) e (b)) e a dispersão ((c) e (d)).

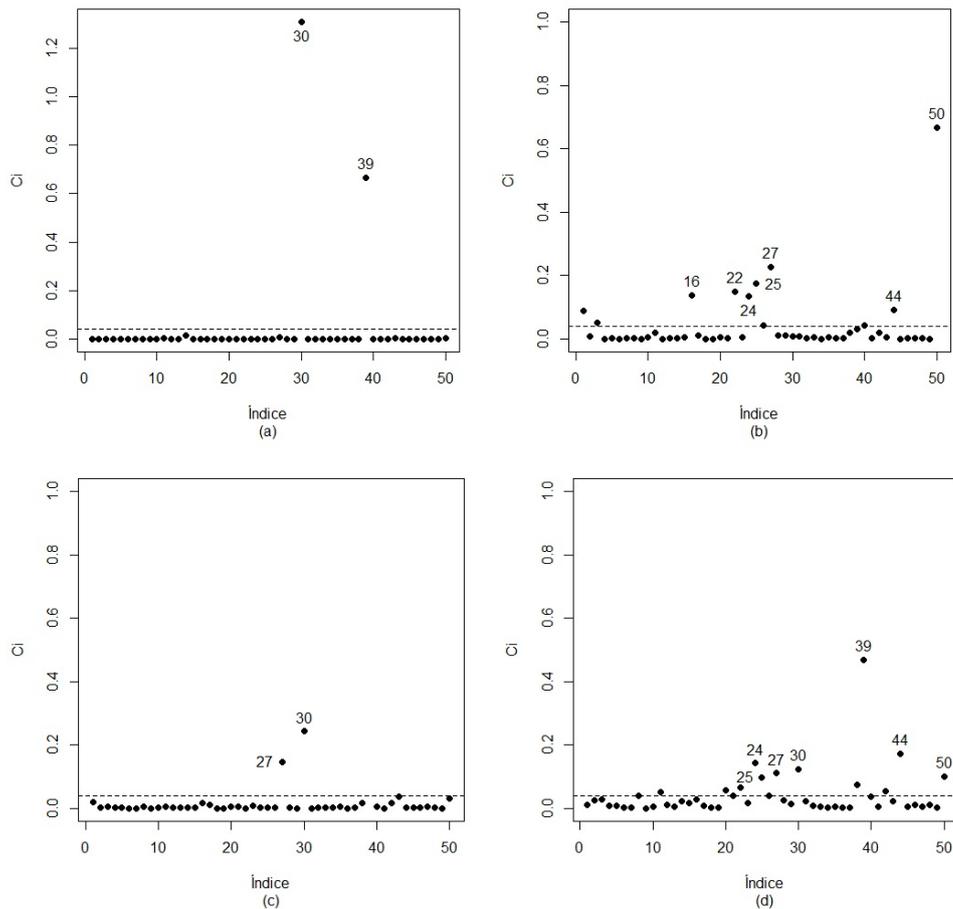


Figura 4.5: Gráficos de C_i contra o índice das observações para os esquemas de perturbação (a) perturbação aditiva na variável resposta, (b) perturbação aditiva no preditor para o ajuste da média (matriz \mathbf{X}), (c) perturbação aditiva no preditor para o ajuste da dispersão (matriz \mathbf{S}) e (d) perturbação de casos ponderados, referente aos dados simulados da distribuição passeio aleatório.

Capítulo 5

Considerações Finais

Nesta dissertação foram estudadas técnicas de diagnóstico nos modelos lineares generalizados com superdispersão (MLGSs), sendo esta a maior contribuição deste trabalho, por não haver na literatura estudos na área para a referida classe de modelos. Assim, este trabalho preencheu uma lacuna, em que o Capítulo 3 constitui a principal contribuição teórica desta dissertação.

Estudos com dados simulados foram utilizados para fins de aplicação das técnicas de diagnóstico apresentadas na classe de modelos estudada. Assim, nota-se que os métodos diagnóstico para os MLGSs desenvolvidos neste trabalho mostraram-se eficientes para detecção de pontos com alta alavancagem, assim como observações aberrantes e influentes entre o conjunto de dados. Toda a análise de verificação de observações mais afastadas das demais foi realizada por meio de métodos gráficos, construídos a partir do software R.

Com isso, conclui-se que as técnicas de diagnóstico na classe dos modelos lineares generalizados com superdispersão desenvolvidas aqui contribuem amplamente na área de modelagem por ser bastante útil em situações práticas em análise de regressão.

Acrescenta-se que uma proposta de continuidade do tema abordado nesta dissertação refere-se ao desenvolvimento dos seguintes trabalhos futuros: obtenção de expressões para os resíduos padronizado e componente do desvio padronizado com respeito ao ajuste da dispersão, assim como a construção de gráfico envelope para os resíduos estudados.

Bibliografia

- [1] Cook, R. D. (1977). Detection of influential observations in linear regressions. *Technometrics*, **19**, 15-18.
- [2] Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B*, **48**, 133-169.
- [3] Cordeiro, G. M., Botter, D. A. (2001). Second-Order biases of maximum likelihood estimates in overdispersed generalized linear models. *Statistics and Probability Letters*, **55**, 269-280.
- [4] Cordeiro, G. M., Cysneiros, A. H. M. A., Cysneiros, F. J. A. (2006). Bartlett Adjustments for Overdispersed Generalized Linear Models. *Communications in Statistics - Theory and Methods*, **35**, 937-952.
- [5] Cox, D. R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B*, **49**, 1-39.
- [6] Cox, D. R., Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B*, **30**, 248-275.
- [7] Dey, D. K., Gelfand, A. E., Peng, F. (1997). Overdispersed generalized linear models. *J. Statist. Plann. Inference*, **64**, 93-107.
- [8] Doornick, J. A. (2009). *Ox: an Object-Oriented Matrix Language*, 4th ed. London: Timberlake Consultants Press and Oxford: <http://www.doornick.com>.
- [9] Efron, B. (1986). Double Exponential Families and Their Use in Generalized Linear Regression. *Journal of the American Statistical Association*, **81**, 709-721.

-
- [10] Escobar, L. A., Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, **48**, 507-528.
- [11] Frery, A. C., Cribari-Neto, F. (2005). *Elementos de Estatística Computacional Usando Plataformas de Software Livre/Gratuito*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada.
- [12] Gelfand, A. E., Dalal, S. R. (1990). A note on overdispersed exponential families. *Biometrika*, **77**, 55-64.
- [13] Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237-260.
- [14] Hinde, J., Demétrio, C. G. B. (1998). Overdispersion: model and estimation. *Computational Statistics and Data Analysis*, **27**, 151-170.
- [15] Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*. 2nd ed. New York: John Wiley and Sons.
- [16] Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Lecture Notes in Statistics. 9. New York: Springer-Verlag.
- [17] Lesaffre, F., Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **38**, 963-974.
- [18] McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.
- [19] Mittelbach, F., Goossens, M., Braams, J., Carlisle, D., Rowley, C. (2004). *The LATEX Companion*. Tools and Techniques for Computer Typesetting. Addison Wesley, Boston.
- [20] Nelder, J. A., Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**, 370-384.
- [21] Nocedal, J., Wright, S. J. (1999). *Numerical Optimization*. New York: Springer-Verlag.
- [22] Paula, G. A. (2012). *Modelos de Regressão com apoio computacional*. São Paulo: Instituto de Matemática e Estatística - USP.

- [23] Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, **9**, 705-724.
- [24] 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.
- [25] Smyth, G. K. (1989). Generalized linear models with varying dispersion. *J. R. Statistic. Soc. B*, **51**, 47-60.
- [26] Thomas, W., Cook, R. D. (1989). Assessing influence on regression coefficients in generalized linear models. *Biometrika*, **76**, 741-749.
- [27] Wasan, M. T. (1968). *First Passage Time Distribution of Brownian Motion*. Monograph, Department of Mathematics, Kingston, Ontario: Queen's University.
- [28] Wei, B. C., Hu, Y. Q., Fung, W. K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, **25**, 25-37.
- [29] Wise, M. E. (1966). Tracer-dilution curves in cardiology and random walk and lognormal distributions. *Acta Physiologica Pharmacologica Neerlandica*, **14**, 175-204.
- [30] Wise, M. E. (1971). Skew probability curves with negative powers of the time and related random walks in series. *Statistica Neerlandica*, **25**, 159-180.
- [31] Wise, M. E. (1975). Skew distributions in biomedicine, including some with negative powers of time. *Statistical Distributions in Scientific Work*, **2**, G. P. Patil, S. Kotz and J. K. Ord (editors), 241-262, Dordrecht: Reidel.

Apêndice A

Cálculo dos Momentos

O logaritmo da função de verossimilhança dos modelos lineares generalizados com superdispersão tem a forma

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{\ell=1}^n \{ (y_\ell - \mu_\ell) \psi^{(1,0)}(\mu_\ell, \phi_\ell) + \phi_\ell T(y_\ell) + \psi(\mu_\ell, \phi_\ell) \} + \sum_{\ell=1}^n \log A(y_\ell),$$

em que $A(\cdot)$, $T(\cdot)$ e $\psi(\cdot, \cdot)$ são funções conhecidas e $\psi^{(r,s)} = \partial \psi^{r+s}(\mu, \phi) / \partial \mu^r \partial \phi^s$, com $r, s \geq 0$. As matrizes \mathbf{M}_r e $\boldsymbol{\Phi}_r$, com $r = 1, 2$ estão definidas na Seção 2.3, equação (2.10).

A.1 Função Escore

A função escore de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ é dada por

$$\mathbf{U} = \mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{pmatrix} \partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) / \partial \boldsymbol{\beta} \\ \partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma} \end{pmatrix},$$

com elementos

$$\begin{aligned} U_r &= \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r} = \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(2,0)} \frac{\partial \mu_\ell}{\partial \beta_r} - \psi_\ell^{(1,0)} \frac{\partial \mu_\ell}{\partial \beta_r} + \psi_\ell^{(1,0)} \frac{\partial \mu_\ell}{\partial \beta_r} \right\} \\ &= \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(2,0)} \frac{\partial \mu_\ell}{\partial \eta_\ell} \frac{\partial \eta_\ell}{\partial \beta_r} \right\} \\ &= \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(2,0)} m_{1\ell}(r) \right\} \end{aligned}$$

e

$$\begin{aligned}
U_R &= \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \gamma_R} = \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(1,1)} \frac{\partial \phi_\ell}{\partial \gamma_R} + T(y_\ell) \frac{\partial \phi_\ell}{\partial \gamma_R} + \psi_\ell^{(0,1)} \frac{\partial \phi_\ell}{\partial \gamma_R} \right\} \\
&= \sum_{\ell=1}^n \left\{ \frac{\partial \phi_\ell}{\partial \gamma_R} \left[(y_\ell - \mu_\ell) \psi_\ell^{(1,1)} + T(y_\ell) + \psi_\ell^{(0,1)} \right] \right\} \\
&= \sum_{\ell=1}^n \left\{ \frac{\partial \phi_\ell}{\partial \tau_\ell} \frac{\partial \tau_\ell}{\partial \gamma_R} \left[(y_\ell - \mu_\ell) \psi_\ell^{(1,1)} + T(y_\ell) + \psi_\ell^{(0,1)} \right] \right\} \\
&= \sum_{\ell=1}^n \{ \phi_{1\ell}(R) v_\ell \},
\end{aligned}$$

sendo

$$\begin{aligned}
m_{1\ell} &= \frac{\partial \mu_\ell}{\partial \eta_\ell}, \quad (r)_\ell = \frac{\partial \eta_\ell}{\partial \beta_r}, \quad \phi_{1\ell} = \frac{\partial \phi_\ell}{\partial \tau_\ell}, \quad (R)_\ell = \frac{\partial \tau_\ell}{\partial \gamma_R}, \\
v_\ell &= (y_\ell - \mu_\ell) \psi_\ell^{(1,1)} + T(y_\ell) + \psi_\ell^{(0,1)}, \\
r &= 1, \dots, p \quad \text{e} \quad R = 1, \dots, q.
\end{aligned}$$

Em notação matricial,

$$\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} = \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1 (\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \quad \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \mathbf{S}^\top \boldsymbol{\Phi}_1 \boldsymbol{\nu},$$

em que $(\mathbf{y} - \boldsymbol{\mu}) = ((y_1 - \mu_1), \dots, (y_n - \mu_n))^\top$ e $\boldsymbol{\nu} = (v_1, \dots, v_n)^\top$, com v_ℓ como visto acima.

A.2 Informação de Fisher

Matriz de Informação Observada de Fisher

Seja $\ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ o logaritmo da função de verossimilhança, assim como U_r e U_R as funções score para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, respectivamente, dadas acima. A matriz de informação observada de Fisher é dada por

$$\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\gamma} \\ \ddot{\mathbf{L}}_{\gamma\beta} & \ddot{\mathbf{L}}_{\gamma\gamma} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} \end{pmatrix},$$

com elementos

$$\begin{aligned}
 U_{rs} &= \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r \partial \beta_s} = \frac{\partial U_r}{\partial \beta_s} = \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(3,0)} m_{1\ell}^2(r)(s)_\ell - \psi_\ell^{(2,0)} m_{1\ell}^2(r)(s)_\ell + \right. \\
 &\quad \left. (y_\ell - \mu_\ell) \psi_\ell^{(2,0)} m_{2\ell}(r)(s)_\ell \right\}, \\
 U_{rS} &= \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \beta_r \partial \gamma_S} = \frac{\partial U_r}{\partial \gamma_S} = \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(2,1)} \phi_{1\ell} m_{1\ell}(r)(S)_\ell \right\}, \\
 U_{Rs} &= \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \gamma_R \partial \beta_s} = \frac{\partial U_R}{\partial \beta_s} = \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(2,1)} \phi_{1\ell} m_{1\ell}(R)_\ell(s)_\ell \right\}, \\
 U_{RS} &= \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \gamma_R \partial \gamma_S} = \frac{\partial U_R}{\partial \gamma_S} = \sum_{\ell=1}^n \left\{ (y_\ell - \mu_\ell) \psi_\ell^{(1,1)} \phi_{2\ell}(R)_\ell(S)_\ell + T(y_\ell) \phi_{2\ell}(R)_\ell(S)_\ell + \right. \\
 &\quad \psi_\ell^{(0,1)} \phi_{2\ell}(R)_\ell(S)_\ell + (y_\ell - \mu_\ell) \psi_\ell^{(1,2)} \phi_{1\ell}^2(R)_\ell(S)_\ell + \\
 &\quad \left. \psi_\ell^{(0,2)} \phi_{1\ell}^2(R)_\ell(S)_\ell \right\},
 \end{aligned}$$

sendo

$$\begin{aligned}
 m_{1\ell} &= \frac{\partial \mu_\ell}{\partial \eta_\ell}, & m_{2\ell} &= \frac{\partial^2 \mu_\ell}{\partial \eta_\ell^2}, & \phi_{1\ell} &= \frac{\partial \phi_\ell}{\partial \tau_\ell}, & \phi_{2\ell} &= \frac{\partial^2 \phi_\ell}{\partial \tau_\ell^2}, \\
 (r)_\ell &= \frac{\partial \eta_\ell}{\partial \beta_r}, & (s)_\ell &= \frac{\partial \eta_\ell}{\partial \beta_s}, & (R)_\ell &= \frac{\partial \tau_\ell}{\partial \gamma_R}, & (S)_\ell &= \frac{\partial \tau_\ell}{\partial \gamma_S}, \\
 & & & & r = s = 1, \dots, p & \text{ e } & R = S = 1, \dots, q.
 \end{aligned}$$

Em notação matricial,

$$\begin{aligned}
 \ddot{\mathbf{L}}_{\beta\beta} &= \mathbf{X}^\top \left[\boldsymbol{\Psi}^{(3,0)} \mathbf{M}_1^2 \mathbf{D}(\mathbf{r}) - \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1^2 + \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_2 \mathbf{D}(\mathbf{r}) \right] \mathbf{X}, \\
 \ddot{\mathbf{L}}_{\beta\gamma} &= \mathbf{X}^\top \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(2,1)} \boldsymbol{\Phi}_1 \mathbf{M}_1 \mathbf{S}, \\
 \ddot{\mathbf{L}}_{\gamma\beta} &= \mathbf{S}^\top \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(2,1)} \boldsymbol{\Phi}_1 \mathbf{M}_1 \mathbf{X} = (\ddot{\mathbf{L}}_{\beta\gamma})^{-1}, \\
 \ddot{\mathbf{L}}_{\gamma\gamma} &= \mathbf{S}^\top \left[\boldsymbol{\Phi}_2 \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(1,1)} + \boldsymbol{\Phi}_2 \mathbf{D}(\mathbf{T}) + \boldsymbol{\Phi}_2 \boldsymbol{\Psi}^{(0,1)} + \boldsymbol{\Phi}_1^2 \mathbf{D}(\mathbf{r}) \boldsymbol{\Psi}^{(1,2)} + \boldsymbol{\Phi}_1^2 \boldsymbol{\Psi}^{(0,2)} \right] \mathbf{S},
 \end{aligned}$$

em que $\mathbf{D}(\mathbf{r}) = \text{diag}\{(y_1 - \mu_1), \dots, (y_n - \mu_n)\}$ e $\mathbf{D}(\mathbf{T}) = \text{diag}\{T(y_1), \dots, T(y_n)\}$.

Matriz de Informação Esperada de Fisher

Considere $\ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ o logaritmo da função de verossimilhança. Além disso, seja U_{rs}, U_{rS}, U_{Rs} e U_{RS} as quantidades correspondentes à matriz de informação observada de Fisher apresentadas acima e $E(y_\ell - \mu_\ell) = 0$. Então, a matriz de informação esperada de Fisher $\mathbf{K}_{\boldsymbol{\theta}, \boldsymbol{\theta}} = E(-\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}})$, com $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, é dada por

$$\mathbf{K}_{\boldsymbol{\theta}, \boldsymbol{\theta}} = \begin{pmatrix} E(-\ddot{\mathbf{L}}_{\beta\beta}) & E(-\ddot{\mathbf{L}}_{\beta\gamma}) \\ E(-\ddot{\mathbf{L}}_{\gamma\beta}) & E(-\ddot{\mathbf{L}}_{\gamma\gamma}) \end{pmatrix} = \begin{pmatrix} \mathbf{K}_{\beta, \beta} & \mathbf{K}_{\beta, \gamma} \\ \mathbf{K}_{\gamma, \beta} & \mathbf{K}_{\gamma, \gamma} \end{pmatrix},$$

com elementos

$$\begin{aligned} E(-U_{rs}) &= \sum_{\ell=1}^n \left\{ \psi_\ell^{(2,0)} m_{1\ell}^2(r)_\ell(s)_\ell \right\}, \\ E(-U_{rS}) &= 0, \\ E(-U_{Rs}) &= 0, \\ E(-U_{RS}) &= \sum_{\ell=1}^n \left\{ -\phi_{2\ell}(R)_\ell(S)_\ell E[T(y_\ell)] - \phi_{2\ell} \psi_\ell^{(0,1)}(R)_\ell(S)_\ell - \phi_{1\ell}^2 \psi_\ell^{(0,2)}(R)_\ell(S)_\ell \right\}, \end{aligned}$$

em que $r = s = 1, \dots, p$ e $R = S = 1, \dots, q$.

Como $E[T(y_\ell)] = -\psi_\ell^{(0,1)}$, conforme equação (2.2), tem-se que

$$E(-U_{RS}) = \sum_{\ell=1}^n -\phi_{1\ell}^2 \psi_\ell^{(0,2)}(R)_\ell(S)_\ell.$$

Em notação matricial,

$$\mathbf{K}_{\boldsymbol{\theta}, \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1^2 \mathbf{X} & \mathbf{0} \\ \mathbf{0} & -\mathbf{S}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \mathbf{S} \end{pmatrix}.$$

Apêndice B

Passeio Aleatório

B.1 Derivadas da função $\psi(\mu, \phi)$

Para os cálculos da função escore, das matrizes de informação observada e esperada de Fisher, assim como das técnicas de diagnóstico para o caso particular do modelo passeio aleatório estudado no Capítulos 4, visto na Seção 2.4 como distribuição pertencente aos modelos lineares generalizados com superdispersão, foram necessárias algumas derivadas da função

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

em que $\psi^{(r,s)} = \partial\psi^{r+s}(\mu, \phi)/\partial\mu^r\partial\phi^s$ ($r, s \geq 0$).

As derivadas da função $\psi(\mu, \phi)$ necessárias para esta dissertação podem ser obtidas usando o sistema de computação algébrica MAPLE ou similar e são apresentadas a seguir

$$\begin{aligned}\psi^{(1,0)} &= \frac{(-\phi)^{1/2}(2\mu - \phi)^{1/2} - 2\mu + \phi}{\{(2\mu - \phi)^{1/2} - (-\phi)^{1/2}\}^3(2\mu - \phi)^{1/2}}, \\ \psi^{(2,0)} &= \frac{2}{\{(2\mu - \phi)^{1/2} - (-\phi)^{1/2}\}^3(2\mu - \phi)^{1/2}}, \\ \psi^{(3,0)} &= \frac{-2[4(2\mu - \phi)^{1/2} - (-\phi)^{1/2}]}{\{(2\mu - \phi)^{1/2} - (-\phi)^{1/2}\}^4(2\mu - \phi)^{3/2}}, \\ \psi^{(0,1)} &= \frac{(-\phi)^{1/2}(2\mu - \phi)^{1/2} - 2\mu + \phi}{(-\phi)^{1/2}(2\mu - \phi)^{1/2}\{(-\phi)^{1/2} - (2\mu - \phi)^{1/2}\}^2},\end{aligned}$$

$$\begin{aligned}\psi^{(0,2)} &= \frac{1}{2\phi(\phi^2 - 2\mu\phi)^{1/2}}, \\ \psi^{(1,1)} &= \frac{1}{\{(2\mu - \phi)^{1/2} - (-\phi)\}^2(-\phi)^{1/2}(2\mu - \phi)^{1/2}}, \\ \psi^{(1,2)} &= \frac{1}{2(2\mu - \phi)^{3/2}(-\phi)^{3/2}}, \\ \psi^{(2,1)} &= -\frac{2\{2\phi - 3\mu + 2(\phi^2 - 2\mu\phi)^{1/2}\}}{(\phi - 2\mu)\{(2\mu - \phi)^{1/2} - (-\phi)^{1/2}\}^4(\phi^2 - 2\mu\phi)^{1/2}},\end{aligned}$$

com $\mu = 1/\theta + 1/\delta$ e $\phi = -\delta/(2\theta^2)$.

B.2 Obtenção das matrizes \mathbf{M}_1 , \mathbf{M}_2 , Φ_1 e Φ_2

Na Seção 2.4 foi visto que a distribuição passeio aleatório possui uma relação com a distribuição inversa Gaussiana, ou seja, se Z é uma variável aleatória com distribuição inversa Gaussiana, então sua recíproca $Y = 1/Z$ segue uma distribuição passeio aleatório. Assim, é usada tal relação entre as duas distribuições para a obtenção das matrizes \mathbf{M}_1 , \mathbf{M}_2 , Φ_1 e Φ_2 , em que é necessário utilizar a regra da cadeia (Previdelli, 2005).

Inicialmente, para o ajuste da média, são apresentados os cálculos tanto para a ligação identidade, quanto para a log. De forma geral, tem-se que

$$\begin{aligned}\mathbf{M}_1 &= \frac{\partial\mu_{PA}}{\partial\eta} = \frac{\partial\mu_{PA}}{\partial\mu_{IG}} \cdot \frac{\partial\mu_{IG}}{\partial\eta} \\ \mathbf{M}_2 &= \frac{\partial^2\mu_{PA}}{\partial\eta^2},\end{aligned}$$

em que $\mu_{PA} = 1/\theta + 1/\delta$ e $\mu_{IG} = \theta = f^{-1}(\eta)$, com $\eta = X\beta$, correspondem aos parâmetros da média para os modelos passeio aleatório e inversa Gaussiana, respectivamente.

Para a ligação identidade, em que $\eta = f(\mu_{IG}) = \mu_{IG}$, é obtido a seguinte forma para \mathbf{M}_1 e \mathbf{M}_2 :

$$\begin{aligned}\mathbf{M}_1 &= \frac{\partial\mu_{PA}}{\partial\mu_{IG}} \cdot \frac{\partial\mu_{IG}}{\partial\eta} = -\frac{1}{\theta^2}, \\ \mathbf{M}_2 &= \frac{\partial^2\mu_{PA}}{\partial\eta^2} = \frac{\partial}{\partial\eta} \left(\frac{-1}{\eta^2} \right) = \frac{2}{\theta^3}.\end{aligned}$$

Já para a ligação log, tem-se que $\eta = f(\mu_{IG}) = \log(\mu_{IG})$. Logo,

$$\mathbf{M}_1 = \frac{\partial \mu_{PA}}{\partial \mu_{IG}} \cdot \frac{\partial \mu_{IG}}{\partial \eta} = -\frac{1}{\theta^2} \cdot \theta = -\frac{1}{\theta},$$

$$\mathbf{M}_2 = \frac{\partial^2 \mu_{PA}}{\partial \eta^2} = \frac{\partial}{\partial \eta} \left(-\frac{1}{\exp(\eta)} \right) = \frac{1}{\theta}.$$

Quanto ao ajuste da dispersão, é considerado apenas a ligação log, em que $\tau = g(\phi_{IG}) = \log(\phi_{IG})$. De forma geral tem-se que

$$\Phi_1 = \frac{\partial \phi_{PA}}{\partial \tau} = \frac{\partial \phi_{PA}}{\partial \phi_{IG}} \cdot \frac{\partial \phi_{IG}}{\partial \tau}$$

$$\Phi_2 = \frac{\partial^2 \phi_{PA}}{\partial \tau^2},$$

em que $\phi_{PA} = -1/2\theta^2$ e $\phi_{IG} = \delta = g^{-1}(\tau)$, com $\tau = S\gamma$, correspondem aos parâmetros da dispersão para os modelos passeio aleatório e inversa Gaussiana, respectivamente.

Assim, utilizando a regra da cadeia é obtido

$$\Phi_1 = \frac{\partial \phi_{PA}}{\partial \phi_{IG}} \cdot \frac{\partial \phi_{IG}}{\partial \tau} = -\frac{1}{2\theta^2} \cdot \delta = -\frac{\delta}{2\theta^2} = \phi$$

$$\Phi_2 = \frac{\partial^2}{\partial \tau^2} \left(-\frac{\delta}{2\theta^2} \right) = \frac{\partial}{\partial \tau} \left(-\frac{\delta}{2\theta^2} \right) = -\frac{\delta}{2\theta^2} = \phi.$$

Apêndice C

Alavancagem Generalizada

Segundo Wei, Hu e Fung (1998), a alavancagem generalizada de $\hat{\boldsymbol{\theta}}$, $\partial\hat{\mathbf{y}}/\partial\mathbf{y}^\top$, é definida como

$$\mathbf{GL}(\hat{\boldsymbol{\theta}}) = \{(\mathbf{D}_\theta)(-\ddot{\mathbf{L}}_{\theta\theta})^{-1}(\ddot{\mathbf{L}}_{\theta\mathbf{y}})\}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

em que $\mathbf{D}_\theta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top$, $\ddot{\mathbf{L}}_{\theta\theta} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ e $\ddot{\mathbf{L}}_{\theta\mathbf{y}} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\mathbf{y}^\top$, com $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ definida em (2.3).

C.1 Cálculos para $\mathbf{GL}^\mu(\hat{\boldsymbol{\theta}})$

$$\begin{aligned} \mathbf{GL}^\mu(\boldsymbol{\theta}) &= (\mathbf{D}_\theta^\mu)(-\ddot{\mathbf{L}}_{\theta\theta})^{-1}(\ddot{\mathbf{L}}_{\theta\mathbf{y}}) \\ &= (\mathbf{M}_1\mathbf{X}, \mathbf{0}) \begin{pmatrix} -\ddot{\mathbf{L}}_{\beta\beta} & -\ddot{\mathbf{L}}_{\beta\gamma} \\ -\ddot{\mathbf{L}}_{\gamma\beta} & -\ddot{\mathbf{L}}_{\gamma\gamma} \end{pmatrix}^{-1} \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \\ \ddot{\mathbf{L}}_{\gamma\mathbf{y}} \end{pmatrix} \\ &= (\mathbf{M}_1\mathbf{X}, \mathbf{0}) \begin{pmatrix} -\ddot{\mathbf{L}}_{\beta\beta}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^\top & -\mathbf{E}^{-1} \end{pmatrix} \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \\ \ddot{\mathbf{L}}_{\gamma\mathbf{y}} \end{pmatrix}, \end{aligned}$$

em que $\mathbf{F} = \ddot{\mathbf{L}}_{\beta\beta}^{-1}\ddot{\mathbf{L}}_{\beta\gamma}$, $\mathbf{F}^\top = \ddot{\mathbf{L}}_{\gamma\beta}\ddot{\mathbf{L}}_{\beta\beta}^{-1}$, $\mathbf{E} = -\ddot{\mathbf{L}}_{\gamma\gamma} + \ddot{\mathbf{L}}_{\gamma\beta}\ddot{\mathbf{L}}_{\beta\beta}^{-1}\ddot{\mathbf{L}}_{\beta\gamma}$, e $\ddot{\mathbf{L}}_{\beta\beta}$, $\ddot{\mathbf{L}}_{\beta\gamma}$, $\ddot{\mathbf{L}}_{\gamma\beta}$ e $\ddot{\mathbf{L}}_{\gamma\gamma}$ são elementos da matriz de informação observada dada em (2.12).

Logo,

$$\begin{aligned} \mathbf{GL}^\mu(\boldsymbol{\theta}) &= (\mathbf{M}_1\mathbf{X}[-\ddot{\mathbf{L}}_{\beta\beta}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top], \mathbf{M}_1\mathbf{X}[-\mathbf{F}\mathbf{E}^{-1}]) \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \\ \ddot{\mathbf{L}}_{\gamma\mathbf{y}} \end{pmatrix} \\ &= \mathbf{M}_1\mathbf{X}\{-\ddot{\mathbf{L}}_{\beta\beta}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top\}\ddot{\mathbf{L}}_{\beta\mathbf{y}} - \mathbf{M}_1\mathbf{X}\mathbf{F}\mathbf{E}^{-1}\ddot{\mathbf{L}}_{\gamma\mathbf{y}}. \end{aligned}$$

Primeiramente, para calcular a inversa de $\mathbf{E} = -\ddot{\mathbf{L}}_{\gamma\gamma} + \ddot{\mathbf{L}}_{\gamma\beta}\ddot{\mathbf{L}}_{\beta\beta}^{-1}\ddot{\mathbf{L}}_{\beta\gamma}$, considere o seguinte resultado de álgebra linear:

$$(\mathbf{A} + \mathbf{BOP})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{O}^{-1} + \mathbf{PA}^{-1}\mathbf{B})^{-1}\mathbf{PA}^{-1}.$$

Então, neste caso, tem-se que $\mathbf{A} = -\ddot{\mathbf{L}}_{\gamma\gamma}^{-1}$, $\mathbf{B} = \ddot{\mathbf{L}}_{\gamma\beta}$, $\mathbf{O} = \ddot{\mathbf{L}}_{\beta\beta}^{-1}$ e $\mathbf{P} = \ddot{\mathbf{L}}_{\beta\gamma}$. Fazendo $\mathbf{C} = -\mathbf{A}^{-1}\mathbf{B}$, $\mathbf{C}^\top = \mathbf{P}\mathbf{A}^{-1}$ e $(\mathbf{R}^\top\mathbf{R})^{-1} = (\mathbf{O}^{-1} + \mathbf{P}\mathbf{A}^{-1}\mathbf{B})^{-1}$, é obtido $(\mathbf{A} + \mathbf{B}\mathbf{O}\mathbf{P})^{-1} = \mathbf{A}^{-1} + \mathbf{C}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{C}^\top$. Assim, $\mathbf{E}^{-1} = -\ddot{\mathbf{L}}_{\gamma\gamma}^{-1} + \mathbf{C}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{C}^\top$. Portanto,

$$\mathbf{F}\mathbf{E}^{-1} = \ddot{\mathbf{L}}_{\beta\beta}^{-1}\ddot{\mathbf{L}}_{\beta\gamma}[-\ddot{\mathbf{L}}_{\gamma\gamma}^{-1} + \mathbf{C}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{C}^\top]$$

e

$$\mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top = \ddot{\mathbf{L}}_{\beta\beta}^{-1}\ddot{\mathbf{L}}_{\beta\gamma}[-\ddot{\mathbf{L}}_{\gamma\gamma}^{-1} + \mathbf{C}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{C}^\top]\ddot{\mathbf{L}}_{\gamma\beta}\ddot{\mathbf{L}}_{\beta\beta}^{-1}.$$

Então,

$$\mathbf{GL}^\mu(\boldsymbol{\theta}) = \mathbf{M}_1\mathbf{X}\{-\ddot{\mathbf{L}}_{\beta\beta}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top\}\ddot{\mathbf{L}}_{\beta\mathbf{y}} - \mathbf{M}_1\mathbf{X}\mathbf{F}\mathbf{E}^{-1}\ddot{\mathbf{L}}_{\gamma\mathbf{y}}. \quad (\text{C.1})$$

Reescrevendo (C.1), tem-se que $GL^\mu(\hat{\boldsymbol{\theta}}) = GL_\beta^\mu(\hat{\boldsymbol{\theta}}) + GL_\gamma^\mu(\hat{\boldsymbol{\theta}})$, em que

$$GL_\beta^\mu(\hat{\boldsymbol{\theta}}) = -\mathbf{M}_1\mathbf{X}(\ddot{\mathbf{L}}_{\beta\beta})^{-1}\ddot{\mathbf{L}}_{\beta\mathbf{y}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

e

$$GL_\gamma^\mu(\hat{\boldsymbol{\theta}}) = \mathbf{M}_1\mathbf{X}\mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top\ddot{\mathbf{L}}_{\beta\mathbf{y}} - \mathbf{M}_1\mathbf{X}\mathbf{F}\mathbf{E}^{-1}\ddot{\mathbf{L}}_{\gamma\mathbf{y}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Ou ainda,

$$GL_\beta^\mu(\hat{\boldsymbol{\theta}}) = -\mathbf{M}_1\mathbf{X}\left\{\mathbf{X}^\top\left[\boldsymbol{\Psi}^{(3,0)}\mathbf{M}_1^2\mathbf{D}(\mathbf{r}) - \boldsymbol{\Psi}^{(2,0)}\mathbf{M}_1^2 + \boldsymbol{\Psi}^{(2,0)}\mathbf{M}_2\mathbf{D}(\mathbf{r})\right]\mathbf{X}\right\}^{-1}\mathbf{X}^\top\boldsymbol{\Psi}^{(2,0)}\mathbf{M}_1|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

e

$$GL_\gamma^\mu(\hat{\boldsymbol{\theta}}) = \mathbf{M}_1\mathbf{X}\mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top\mathbf{X}^\top\boldsymbol{\Psi}^{(2,0)}\mathbf{M}_1 - \mathbf{M}_1\mathbf{X}\mathbf{F}\mathbf{E}^{-1}\left\{\mathbf{S}^\top\left[(\boldsymbol{\Psi}^{(1,1)} + \mathbf{D}(\mathbf{T}'_{\mathbf{y}}))\boldsymbol{\Phi}_1\right]\right\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

sendo $\mathbf{D}(\mathbf{T}'_{\mathbf{y}}) = \text{diag}\{T'(y_1), \dots, T'(y_n)\}$, com $T'(y_\ell) = \frac{\partial T(y_\ell)}{\partial y_\ell}$, $\ell = 1, \dots, n$ e $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$, em que $r_\ell = (y_\ell - \hat{\mu}_\ell)$. Os componentes das matrizes diagonais $\boldsymbol{\Psi}^{(3,0)}$, $\boldsymbol{\Psi}^{(2,0)}$, $\boldsymbol{\Psi}^{(1,1)}$, \mathbf{M}_1 , \mathbf{M}_2 e $\boldsymbol{\Phi}_1$ encontram-se definidos em (2.10).

C.2 Cálculos para $\mathbf{GL}^\phi(\hat{\boldsymbol{\theta}})$

Também é possível observar a influência de y_ℓ em $\hat{\phi}_\ell$, denotada por $\partial\hat{\phi}_\ell/\partial y_\ell$, em que $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^\top$. Assim, pela definição de Wei, Hu e Fung (1998), tem-se que

$$\begin{aligned} \mathbf{GL}^\phi(\boldsymbol{\theta}) &= (\mathbf{D}_\theta^\phi)(-\ddot{\mathbf{L}}_{\theta\theta})^{-1}(\ddot{\mathbf{L}}_{\theta\mathbf{y}}) \\ &= (\mathbf{0}, \boldsymbol{\Phi}_1\mathbf{S}) \begin{pmatrix} -\ddot{\mathbf{L}}_{\beta\beta} & -\ddot{\mathbf{L}}_{\beta\gamma} \\ -\ddot{\mathbf{L}}_{\gamma\beta} & -\ddot{\mathbf{L}}_{\gamma\gamma} \end{pmatrix}^{-1} \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \\ \ddot{\mathbf{L}}_{\gamma\mathbf{y}} \end{pmatrix} \\ &= (\mathbf{0}, \boldsymbol{\Phi}_1\mathbf{S}) \begin{pmatrix} -\ddot{\mathbf{L}}_{\beta\beta}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^\top & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^\top & -\mathbf{E}^{-1} \end{pmatrix} \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \\ \ddot{\mathbf{L}}_{\gamma\mathbf{y}} \end{pmatrix}, \end{aligned}$$

em que $\mathbf{F} = \ddot{\mathbf{L}}_{\beta\beta}^{-1}\ddot{\mathbf{L}}_{\beta\gamma}$, $\mathbf{F}^\top = \ddot{\mathbf{L}}_{\gamma\beta}\ddot{\mathbf{L}}_{\beta\beta}^{-1}$, $\mathbf{E} = -\ddot{\mathbf{L}}_{\gamma\gamma} + \ddot{\mathbf{L}}_{\gamma\beta}\ddot{\mathbf{L}}_{\beta\beta}^{-1}\ddot{\mathbf{L}}_{\beta\gamma}$, e $\ddot{\mathbf{L}}_{\beta\beta}$, $\ddot{\mathbf{L}}_{\beta\gamma}$, $\ddot{\mathbf{L}}_{\gamma\beta}$ e $\ddot{\mathbf{L}}_{\gamma\gamma}$ são elementos da matriz de informação observada dada em (2.12).

Logo,

$$\begin{aligned} \mathbf{GL}^\phi(\boldsymbol{\theta}) &= (\boldsymbol{\Phi}_1\mathbf{S}[-\mathbf{E}^{-1}\mathbf{F}^\top], \boldsymbol{\Phi}_1\mathbf{S}\mathbf{E}^{-1}) \begin{pmatrix} \ddot{\mathbf{L}}_{\beta\mathbf{y}} \\ \ddot{\mathbf{L}}_{\gamma\mathbf{y}} \end{pmatrix} \\ &= \boldsymbol{\Phi}_1\mathbf{S}[-\mathbf{E}^{-1}\mathbf{F}^\top]\ddot{\mathbf{L}}_{\beta\mathbf{y}} + \boldsymbol{\Phi}_1\mathbf{S}\mathbf{E}^{-1}\ddot{\mathbf{L}}_{\gamma\mathbf{y}}. \end{aligned}$$

Dos cálculos anteriores, tem-se que

$$\mathbf{E}^{-1} = -\ddot{\mathbf{L}}_{\gamma\gamma}^{-1} + \mathbf{C}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{C}^\top$$

e

$$\mathbf{E}^{-1}\mathbf{F}^\top = [-\ddot{\mathbf{L}}_{\gamma\gamma}^{-1} + \mathbf{C}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{C}^\top]\ddot{\mathbf{L}}_{\gamma\beta}\ddot{\mathbf{L}}_{\beta\beta}^{-1}.$$

Então,

$$\begin{aligned} \mathbf{GL}^\phi(\boldsymbol{\theta}) &= \boldsymbol{\Phi}_1\mathbf{S}[-\mathbf{E}^{-1}\mathbf{F}^\top]\ddot{\mathbf{L}}_{\beta\mathbf{y}} + \boldsymbol{\Phi}_1\mathbf{S}\mathbf{E}^{-1}\ddot{\mathbf{L}}_{\gamma\mathbf{y}} \\ &= -\boldsymbol{\Phi}_1\mathbf{S}\mathbf{E}^{-1}[\mathbf{F}^\top\ddot{\mathbf{L}}_{\beta\mathbf{y}} - \ddot{\mathbf{L}}_{\gamma\mathbf{y}}], \end{aligned}$$

que pode ser reescrita da seguinte forma $\mathbf{GL}^\phi(\hat{\boldsymbol{\theta}}) = \mathbf{GL}_\beta^\phi(\hat{\boldsymbol{\theta}}) + \mathbf{GL}_\gamma^\phi(\hat{\boldsymbol{\theta}})$, sendo

$$\mathbf{GL}_\beta^\phi(\hat{\boldsymbol{\theta}}) = -\boldsymbol{\Phi}_1\mathbf{S}\mathbf{E}^{-1}\mathbf{F}^\top\ddot{\mathbf{L}}_{\beta\mathbf{y}} + \boldsymbol{\Phi}_1\mathbf{S}[\mathbf{V}\mathbf{W}^{-1}\mathbf{V}^\top]\ddot{\mathbf{L}}_{\gamma\mathbf{y}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

e

$$\mathbf{GL}_\gamma^\phi(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Phi}_1 \mathbf{S} - \ddot{\mathbf{L}}_{\gamma\gamma}^{-1} \ddot{\mathbf{L}}_{\gamma y} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

em que $\mathbf{VW}^{-1}\mathbf{V}^\top = \mathbf{C}(\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{C}^\top$.

Ou ainda,

$$\begin{aligned} GL_\beta^\phi(\hat{\boldsymbol{\theta}}) &= -\boldsymbol{\Phi}_1 \mathbf{S} \mathbf{E}^{-1} \mathbf{F}^\top \mathbf{X}^\top \boldsymbol{\Psi}^{(2,0)} \mathbf{M}_1 \\ &\quad + \boldsymbol{\Phi}_1 \mathbf{S} \mathbf{V} \mathbf{W}^{-1} \mathbf{V}^\top \left\{ \mathbf{S}^\top \left[\boldsymbol{\Psi}^{(1,1)} + \mathbf{D}(\mathbf{T}'_y) \right] \boldsymbol{\Phi}_1 \right\} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \end{aligned}$$

e

$$\begin{aligned} GL_\gamma^\phi(\hat{\boldsymbol{\theta}}) &= \boldsymbol{\Phi}_1 \mathbf{S} \left(-\ddot{\mathbf{L}}_{\gamma\gamma} \right)^{-1} \ddot{\mathbf{L}}_{\gamma y} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= \boldsymbol{\Phi}_1 \mathbf{S} \left(-\ddot{\mathbf{L}}_{\gamma\gamma} \right)^{-1} \left[\mathbf{S}^\top \left(\boldsymbol{\Psi}^{(1,1)} + \mathbf{D}(\mathbf{T}'_y) \right) \boldsymbol{\Phi}_1 \right] |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \end{aligned}$$

com $\mathbf{E} = -\ddot{\mathbf{L}}_{\gamma\gamma} + \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1} \ddot{\mathbf{L}}_{\beta\gamma}$, $\mathbf{F}^\top = \ddot{\mathbf{L}}_{\gamma\beta} \ddot{\mathbf{L}}_{\beta\beta}^{-1}$, $\mathbf{V} = (-\ddot{\mathbf{L}}_{\gamma\gamma})^{-1} \ddot{\mathbf{L}}_{\gamma\beta}$, $\mathbf{V}^\top = \ddot{\mathbf{L}}_{\beta\gamma} (-\ddot{\mathbf{L}}_{\gamma\gamma})^{-1}$, $\mathbf{W} = \ddot{\mathbf{L}}_{\beta\beta} - \ddot{\mathbf{L}}_{\beta\gamma} \ddot{\mathbf{L}}_{\gamma\gamma}^{-1} \ddot{\mathbf{L}}_{\gamma\beta}$, e $\ddot{\mathbf{L}}_{\beta\beta}$, $\ddot{\mathbf{L}}_{\beta\gamma}$, $\ddot{\mathbf{L}}_{\gamma\beta}$ e $\ddot{\mathbf{L}}_{\gamma\gamma}$ são elementos da matriz de informação observada dada em (2.12). Os componentes das matrizes diagonais $\boldsymbol{\Psi}^{(2,0)}$, $\boldsymbol{\Psi}^{(1,1)}$, \mathbf{M}_1 e $\boldsymbol{\Phi}_1$ estão definidos em (2.10) e $\mathbf{D}(\mathbf{T}'_y) = \text{diag}\{T'(y_1), \dots, T'(y_n)\}$, com $T'(y_\ell) = \frac{\partial T(y_\ell)}{\partial y_\ell}$, $\ell = 1, \dots, n$.

Apêndice D

Influência Global

Suponha que o logaritmo da função de verossimilhança para o parâmetro $\boldsymbol{\theta}$ seja agora expresso na forma

$$\ell_{\delta}(\boldsymbol{\theta}; \mathbf{y}) = \sum_{\ell=1}^n \delta_{\ell} \ell(\boldsymbol{\theta}; y_{\ell}), \quad (\text{D.1})$$

em que $\ell(\boldsymbol{\theta}; y_{\ell})$ denota o logaritmo da função de verossimilhança correspondente à ℓ -ésima observação e δ_{ℓ} é um tipo de perturbação, definida tal que $0 \leq \delta_{\ell} \leq 1$. Quando $\delta_{\ell} = 1, \forall \ell$, significa que não há perturbação no modelo e quando $\delta_{\ell} = 0$ significa que a ℓ -ésima observação foi excluída.

A estimativa de $\boldsymbol{\beta}$ fica, supondo a estrutura (D.1), dada por

$$\hat{\boldsymbol{\beta}}_{\delta} = (\mathbf{X}^{\top} \boldsymbol{\Delta} \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^{\top} \boldsymbol{\Delta} \hat{\mathbf{W}}_{\beta\beta} \hat{\boldsymbol{\xi}}_1,$$

em que $\boldsymbol{\Delta}$ é uma matriz de uns com δ na ℓ -ésima posição. Ou seja,

$$\boldsymbol{\Delta} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \delta & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Assim,

$$\begin{aligned}\mathbf{X}^\top \Delta \hat{\mathbf{W}}_{\beta\beta} \mathbf{X} &= \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \left[I - \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 - \delta & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \right] \mathbf{X} \\ &= \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X} - (1 - \delta) \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \mathbf{x}_\ell^\top.\end{aligned}$$

Analogamente, tem-se que $\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\boldsymbol{\xi}}_1 = \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\boldsymbol{\xi}}_1 - (1 - \delta) \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \hat{\boldsymbol{\xi}}_{1\ell}$.

Para calcular $(\mathbf{X}^\top \Delta \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1}$, considere o seguinte resultado de álgebra linear:

$$(\mathbf{A} + \mathbf{U}\mathbf{V}^\top)^{-1} = \mathbf{A}^{-1} - \frac{(\mathbf{A}^{-1}\mathbf{U})(\mathbf{V}^\top \mathbf{A}^{-1})}{1 + \mathbf{V}^\top \mathbf{A}^{-1} \mathbf{U}},$$

com $\mathbf{A} = \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X}$, $\mathbf{U} = -(1 - \delta) \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}$ e $\mathbf{V}^\top = \mathbf{x}_\ell^\top$.

Logo,

$$\begin{aligned}(\mathbf{X}^\top \Delta \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} &= (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \\ &\quad - \frac{[(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} (-(1 - \delta) \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell})] [\mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})]}{1 + \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} (-(1 - \delta) \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell})} \\ &= (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \\ &\quad + \frac{[(1 - \delta) (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1}]}{1 - (1 - \delta) \hat{\omega}_{\beta\beta_\ell}^{1/2} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}} \\ &= (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \\ &\quad + \frac{[(1 - \delta) (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1}]}{1 - (1 - \delta) \hat{h}_{\ell\ell}},\end{aligned}$$

em que $\hat{h}_{\ell\ell} = \hat{\omega}_{\beta\beta_\ell}^{1/2} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}$.

Portanto,

$$\begin{aligned}
\hat{\beta}_\delta &= (\mathbf{X}^\top \Delta \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \Delta \hat{\mathbf{W}}_{\beta\beta} \hat{\xi}_1 \\
&= \left\{ (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} + \frac{[(1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1}]}{1 - (1-\delta)\hat{h}_{\ell\ell}} \right\} \\
&\quad [\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\xi}_1 - (1-\delta)\mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \hat{\xi}_{1\ell}] \\
&= (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\xi}_1 - (1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \hat{\xi}_{1\ell} \\
&\quad + \frac{(1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\xi}_1}{1 - (1-\delta)\hat{h}_{\ell\ell}} \\
&\quad - \frac{(1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} (1-\delta)\mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell} \hat{\xi}_{1\ell}}{1 - (1-\delta)\hat{h}_{\ell\ell}} \\
&= \hat{\beta} - \frac{(1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}}{1 - (1-\delta)\hat{h}_{\ell\ell}} [(1 - (1-\delta)\hat{h}_{\ell\ell}) \hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell} \\
&\quad - \hat{\omega}_{\beta\beta_\ell}^{1/2} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\xi}_1 \\
&\quad + \hat{\omega}_{\beta\beta_\ell}^{1/2} \mathbf{x}_\ell^\top \hat{\omega}_{\beta\beta_\ell}^{1/2} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} (1-\delta)\mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell}] \\
&= \hat{\beta} - \frac{(1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}}{1 - (1-\delta)\hat{h}_{\ell\ell}} [\hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell} - (1-\delta)\hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell} \hat{h}_{\ell\ell} \\
&\quad - \hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\eta}_\ell + (1-\delta)\hat{\omega}_{\beta\beta_\ell}^{1/2} \mathbf{x}_\ell^\top \hat{\omega}_{\beta\beta_\ell}^{1/2} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell}] \\
&= \hat{\beta} - \frac{(1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}}{1 - (1-\delta)\hat{h}_{\ell\ell}} [\hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell} - (1-\delta)\hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell} \hat{h}_{\ell\ell} \\
&\quad - \hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\eta}_\ell + (1-\delta)\hat{\omega}_{\beta\beta_\ell}^{1/2} \hat{\xi}_{1\ell} \hat{h}_{\ell\ell}] \\
&= \hat{\beta} - \frac{(1-\delta)(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}}{1 - (1-\delta)\hat{h}_{\ell\ell}} [\hat{\omega}_{\beta\beta_\ell}^{1/2} (\hat{\xi}_{1\ell} - \hat{\eta}_\ell)] \\
&= \hat{\beta} - \frac{(1-\delta)\hat{r}_\ell^\beta (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}}{1 - (1-\delta)\hat{h}_{\ell\ell}},
\end{aligned}$$

em que $\hat{r}_\ell^\beta = \hat{\omega}_{\beta\beta_\ell}^{1/2} (\hat{\xi}_{1\ell} - \hat{\eta}_\ell)$, $\hat{\eta}_\ell = \mathbf{x}_\ell^\top \hat{\beta} = \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \hat{\xi}_1$ e $\hat{h}_{\ell\ell} = \hat{\omega}_{\beta\beta_\ell}^{1/2} \mathbf{x}_\ell^\top (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2}$.

O caso $\delta = 0$ significa que o ℓ -ésimo ponto foi excluído, ou seja, $\hat{\beta}_{(\ell)}$, que fica dado por

$$\hat{\beta}_{(\ell)} = \hat{\beta} - \frac{\hat{r}_\ell^\beta}{(1 - \hat{h}_{\ell\ell})} (\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{x}_\ell \hat{\omega}_{\beta\beta_\ell}^{1/2},$$

em que $\hat{h}_{\ell\ell}$ corresponde ao ℓ -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}} = \hat{\mathbf{W}}_{\beta\beta}^{1/2} \mathbf{X}(\mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{W}}_{\beta\beta}^{1/2}$.

Analogamente, tem-se que a estimativa de γ fica, supondo a estrutura (D.1), dada por

$$\hat{\gamma}_\delta = (\mathbf{S}^\top \Delta \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{S}^\top \Delta \hat{\mathbf{W}}_{\gamma\gamma} \hat{\boldsymbol{\xi}}_2,$$

com

$$(\mathbf{S}^\top \Delta \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} = (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} + \frac{[(1-\delta)(\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\boldsymbol{\omega}}_{\gamma\ell} \mathbf{s}_\ell^\top (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1}]}{1 - (1-\delta)\hat{p}_{\ell\ell}}$$

e

$$\mathbf{S}^\top \Delta \hat{\mathbf{W}}_{\gamma\gamma} \hat{\boldsymbol{\xi}}_2 = \mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \hat{\boldsymbol{\xi}}_2 - (1-\delta) \mathbf{s}_\ell \hat{\boldsymbol{\omega}}_{\gamma\ell} \hat{\boldsymbol{\xi}}_{2\ell},$$

em que $\hat{p}_{\ell\ell} = \hat{\boldsymbol{\omega}}_{\gamma\ell}^{1/2} \mathbf{s}_\ell^\top (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\boldsymbol{\omega}}_{\gamma\ell}^{1/2}$.

Assim,

$$\hat{\gamma}_\delta = \hat{\gamma} - \frac{(1-\delta)\hat{r}_\ell^\gamma (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\boldsymbol{\omega}}_{\gamma\ell}^{1/2}}{1 - (1-\delta)\hat{p}_{\ell\ell}},$$

com $\hat{r}_\ell^\gamma = \hat{\boldsymbol{\omega}}_{\gamma\ell}^{1/2} (\hat{\boldsymbol{\xi}}_{2\ell} - \hat{\boldsymbol{\tau}}_\ell)$, $\hat{\boldsymbol{\tau}}_\ell = \mathbf{s}_\ell^\top \hat{\gamma} = \mathbf{s}_\ell^\top (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \hat{\boldsymbol{\xi}}_2$ e $\hat{p}_{\ell\ell} = \hat{\boldsymbol{\omega}}_{\gamma\ell}^{1/2} \mathbf{s}_\ell^\top (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\boldsymbol{\omega}}_{\gamma\ell}^{1/2}$.

Para $\delta = 0$, o que significa que o ℓ -ésimo ponto foi excluído,

$$\hat{\gamma}_{(\ell)} = \hat{\gamma} - \frac{\hat{r}_\ell^\gamma}{(1 - \hat{p}_{\ell\ell})} (\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{s}_\ell \hat{\boldsymbol{\omega}}_{\gamma\ell}^{1/2},$$

em que $\hat{p}_{\ell\ell}$ corresponde ao ℓ -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{P}} = \hat{\mathbf{W}}_{\gamma\gamma}^{1/2} \mathbf{S}(\mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma} \mathbf{S})^{-1} \mathbf{S}^\top \hat{\mathbf{W}}_{\gamma\gamma}^{1/2}$.