

**Modelos de regressão
beta inflacionados**

Raydonal Ospina Martínez

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

Área de Concentração: Estatística

Orientador: Prof^a. Dr^a. Silvia Lopes de Paula Ferrari

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES e da FAPESP.

São Paulo, 7 de abril de 2008

Modelos de regressão beta inflacionados

Este exemplar corresponde à redação
final da tese devidamente corrigida
e defendida por Raydonal Ospina Martínez
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Profa. Dra. Silvia Lopes de Paula Ferrari (Orientadora) - IME/USP.
- Prof. Dr. Edwin Moises Marcos Ortega - ESALQ-USP.
- Prof. Dr. Francisco Cribari-Neto - UFPE.
- Prof. Dr. José de Azevêdo Cysneros - UFPE.
- Prof. Dr. Klaus Leite Pinto Vasconcellos - UFPE.

Resumo

Nos últimos anos têm sido desenvolvidos modelos de regressão beta, que têm uma variedade de aplicações práticas como, por exemplo, a modelagem de taxas, razões ou proporções. No entanto, é comum que dados na forma de proporções apresentem zeros e/ou uns, o que não permite admitir que os dados provêm de uma distribuição contínua.

Nesta tese, são propostas, distribuições de mistura entre uma distribuição beta e uma distribuição de Bernoulli, degenerada em zero e degenerada em um para modelar dados observados nos intervalos $[0, 1]$, $[0, 1)$ e $(0, 1]$, respectivamente. As distribuições propostas são inflacionadas no sentido de que a massa de probabilidade em zero e/ou um excede o que é permitido pela distribuição beta. Propriedades dessas distribuições são estudadas, métodos de estimação por máxima verossimilhança e momentos condicionais são comparados. Aplicações a vários conjuntos de dados reais são examinadas.

Desenvolvemos também modelos de regressão beta inflacionados assumindo que a distribuição da variável resposta é beta inflacionada. Estudamos estimação por máxima verossimilhança. Derivamos expressões em forma fechada para o vetor score, a matriz de informação de Fisher e sua inversa. Discutimos estimação intervalar para diferentes quantidades populacionais (parâmetros de regressão, parâmetro de precisão) e testes de hipóteses assintóticos. Derivamos expressões para o viés de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros, possibilitando a obtenção de estimadores corrigidos que são mais precisos que os não corrigidos em amostras finitas.

Finalmente, desenvolvemos técnicas de diagnóstico para os modelos de regressão beta inflacionados, sendo adotado o método de influência local baseado na curvatura normal conforme. Ilustramos a teoria desenvolvida em um conjuntos de dados reais.

Palavras-chave: Curvatura normal conforme, dados de frações, distribuição beta inflacionada, estimação por máxima verossimilhança, modelo de regressão beta inflacionado, resíduos.

Abstract

The last years have seen new developments in the theory of beta regression models, which are useful for modelling random variables that assume values in the standard unit interval such as proportions, rates and fractions. In many situations, the dependent variable contains zeros and/or ones. In such cases, continuous distributions are not suitable for modeling this kind of data.

In this thesis we propose mixed continuous-discrete distributions to model data observed on the intervals $[0, 1]$, $[0, 1)$ and $(0, 1]$. The proposed distributions are “inflated beta distributions” in the sense that the probability mass at 0 and/or 1 exceeds what is expected for the beta distribution. Properties of the inflated beta distributions are given. Estimation based on maximum likelihood and conditional moments is discussed and compared. Empirical applications using real data set are provided.

Further, we develop inflated beta regression models in which the underlying assumption is that the response follows an inflated beta law. Estimation is performed by maximum likelihood. We provide closed-form expressions for the score function, Fisher’s information matrix and its inverse. Interval estimation for different population quantities (such as regression parameters, precision parameter, mean response) is discussed and tests of hypotheses on the regression parameters can be performed using asymptotic tests. We also derive the second order biases of the maximum likelihood estimators and use them to define bias-adjusted estimators. The numerical results show that bias reduction can be effective in finite samples.

We also develop a set of diagnostic techniques that can be employed to identify departures from the postulated model and influential observations. To that end, we adopt the local influence approach based in the conformal normal curvature. Finally, we consider empirical examples to illustrate the theory developed.

Keywords: Conformal normal curvature, inflated beta distribution, inflated beta regression model, fractional data, maximum likelihood estimation, residuals.

Conteúdo

Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
1.1 Organização da tese	1
1.2 Suporte computacional	2
2 Agregando zero e/ou um a uma distribuição beta	4
2.1 Introdução	4
2.2 A distribuição beta	6
2.3 A distribuição beta inflacionada em zero ou um	7
2.4 A distribuição beta inflacionada em zero e um	15
2.5 Resultados numéricos e discussão	22
2.6 Aplicações	23
2.7 Conclusões	28
3 Modelos de regressão beta inflacionados	30
3.1 Introdução	30
3.2 Modelo de regressão beta inflacionado em zero ou um	32

CONTEÚDO

3.2.1	Definição	32
3.2.2	Processo de estimação	37
3.2.3	Intervalos de confiança e testes de hipóteses	39
3.2.4	Aplicação a dados simulados	42
3.3	Modelo de regressão beta inflacionado em zero e um	42
3.3.1	Definição	43
3.3.2	Processo de estimação	49
3.3.3	Intervalos de confiança e testes de hipóteses	50
3.3.4	Aplicação a dados simulados	52
3.4	Seleção de modelos	54
3.5	Conclusões	55
4	Análise de viés dos estimadores de máxima verossimilhança nos modelos de regressão beta inflacionados	56
4.1	Introdução	56
4.2	Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um	59
4.2.1	Corrigindo os vieses de $\hat{\zeta}$ e $\hat{\eta}$	62
4.2.2	Corrigindo os vieses de $\hat{\mu}$ e $\hat{\alpha}$	63
4.2.3	Viés de $\hat{\sigma}$	64
4.3	Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero e um	66
4.4	Avaliação numérica	70
4.5	Conclusões	74
5	Diagnóstico em modelos de regressão beta inflacionados	77

CONTEÚDO

5.1	Introdução	77
5.2	Resíduos	78
5.2.1	Resíduos para o modelo de regressão beta inflacionado em zero ou um	78
5.2.2	Resíduos para o modelo de regressão beta inflacionado em zero e um	80
5.3	Influência Local	82
5.3.1	Influência local no modelo de regressão beta inflacionado em zero ou um	87
5.3.2	Influência local no modelo de regressão beta inflacionado em zero e um	90
5.4	Conclusões	92
6	Aplicação	93
6.1	Conclusões	101
7	Conclusões	104
7.1	Considerações finais	104
7.2	Trabalhos futuros	105
A	Apêndices	106
A.1	Apêndice 1	106
A.2	Apêndice 2	109
A.3	Apêndice 3	111
A.4	Apêndice 4	117
A.5	Apêndice 5	119
A.6	Apêndice 6	120
A.7	Apêndice 7	123
A.8	Apêndice 8	134
A.9	Apêndice 9	141

CONTEÚDO

A.10 Apêndice 10	143
A.11 Apêndice 11	145
A.12 Apêndice 12	149
Referências Bibliográficas	151

Lista de Figuras

2.1	Densidades beta para diferentes valores de (μ, ϕ)	8
2.2	Densidades BIZ para diferentes escolhas de μ e ϕ ; $\alpha = 0.4$	11
2.3	Distribuições BIZU para diferentes valores de μ e ϕ ; $\alpha = 0.3, \gamma = 0.5$	17
2.4	Histograma de freqüências e distribuição acumulada para a porcentagem de enfermeiros com curso superior em municípios brasileiros.	26
2.5	Histograma de freqüências e distribuição acumulada para a proporção de óbitos em menores de 1 ano por causas mal definidas em municípios brasileiros.	27
2.6	Histograma de freqüências e distribuição acumulada para a proporção de habitantes que moram até a 200 km do litoral em países litorâneos.	28
3.1	Diagrama de dispersão da resposta $y_t, t = 1, \dots, n$, contra a covariável x_t e z_t	43
3.2	Diagrama de dispersão da resposta $y_t, t = 1, \dots, n$, contra as covariáveis x_t, v_t e z_t	53
4.1	Viés relativo de segunda ordem para $\hat{\sigma} = \mathcal{T}(\hat{\phi})$	66
6.1	Histograma de freqüência e gráfico box-plot para a proporção de assinantes de televisão a cabo que adquirem serviços adicionais.	94
6.2	Gráficos de resíduos. Dados de difusão de televisão a cabo.	97
6.3	Gráficos de resíduos do componente discreto e contínuo. Dados de difusão de televisão a cabo.	98
6.4	Gráficos normais de probabilidades com envelopes simulados. Dados de difusão de televisão a cabo.	98

LISTA DE FIGURAS

- 6.5 (a) Autovalores normalizados em módulo, λ_i^* , com valores de q e (b) influência devida à contribuição agregada de todos os autovetores, $m[q]_j$; esquema de ponderação de casos. 99
- 6.6 (a) Contribuição agregada do autovetor 4-influente correspondente ao maior autovalor ($q = 4$), e (b) contribuição agregada dos autovetores 3-influente correspondentes aos dois maiores autovalores ($q = 3$); esquema de ponderação de casos. 100

Lista de Tabelas

2.1	Resultados de simulação para a distribuição BIZ; $\alpha = 0.2, \mu = 0.1, \phi = 2.0, E(y) = 0.08$ e $\text{Var}(y) = 0.0256$	24
2.2	Resultados da simulação para a distribuição BIZU; $\alpha = 0.2, \gamma = 0.3, \mu = 0.1, \phi = 2.0, E(y) = 0.14$ e $\text{Var}(y) = 0.0724$	25
3.1	Estimativas de máxima verossimilhança com erros padrão para os dados simulados; $y_t \sim \text{BIZ}(\alpha_t, \mu_t, \phi)$	42
3.2	Estimativas de máxima verossimilhança com erros padrão para os dados simulados; $y_t \sim \text{BIZU}(\delta_{0t}, \delta_{1t}, \mu_t, \phi)$	54
3.3	Dois estruturas de regressão para o modelo RBIZ.	54
4.1	Viés de $\hat{\sigma} = \mathcal{T}(\hat{\phi})$	65
4.2	Resultados de simulação, $\gamma_0 = -0.5, \gamma_1 = 1.5$	71
4.3	Resultados de simulação, $\beta_0 = 0.5, \beta_1 = 1.8$ e $\phi = 120$	73
4.4	Resultados de simulação, $\rho_0 = -1.3, \rho_1 = 1.5, \gamma_0 = -1.3$ e $\gamma_1 = 1.5$	75
4.5	Resultados de simulação, $\beta_0 = 1.5, \beta_1 = 1.8$ e $\phi = 120$	76
6.1	Medidas resumo das variáveis	94
6.2	Estimativas de máxima verossimilhança com erros padrão.	95
6.3	Estimativas de máxima verossimilhança e suas versões corrigidas com erros padrões	96

LISTA DE TABELAS

6.4	Medidas de influência utilizando a curvatura normal conforme para o esquema de ponderação de casos.	102
6.5	Estimativas, desvio padrão e DRP dos parâmetros do modelo com a amostra completa e tirando as observações influentes	103
6.6	Medida de impacto DQMRP e EAMRP para os valores ajustados (em porcentagem).	103

Capítulo 1

Introdução

A distribuição beta é útil para modelar experimentos aleatórios que produzem resultados no intervalo $(0, 1)$ medidos de forma contínua. Neste aspecto, Paolino (2001), Kieschnick & McCullough (2003), Ferrari & Cribari-Neto (2004) e Smithson & Verkuilen (2006) propuseram modelos paramétricos de regressão para situações em que a distribuição da variável resposta é beta. Nestes modelos, assume-se que a resposta média é relacionada com um preditor linear por meio de uma função de ligação. O preditor linear envolve covariáveis e parâmetros de regressão desconhecidos. Estes modelos também são indexados por um parâmetro de precisão, que em algumas situações pode variar ao longo das observações (Smithson & Verkuilen, 2006). Os modelos propostos têm uma ampla aplicabilidade prática, principalmente, na modelagem de taxas, razões ou proporções. No entanto, é comum que dados na forma de proporções apresentem zeros e/ou uns, o que não permite admitir que os dados provêm de uma distribuição contínua.

A presente proposta de tese de doutorado objetiva desenvolver diferentes aspectos de inferência e diagnóstico em uma classe de modelos de regressão beta que admitem a presença de zeros e/ou uns. No que se refere à modelagem estudamos modelos de regressão beta inflacionados como extensões naturais do modelo de regressão beta proposto por Ferrari & Cribari-Neto (2004), dado que este se assemelha em muitos aspectos aos modelos lineares generalizados.

1.1 Organização da tese

Esta tese se encontra dividida em sete capítulos. No segundo capítulo propomos distribuições de mistura entre uma distribuição beta e uma distribuição de Bernoulli, degenerada em zero e degenerada em um para modelar dados observados nos intervalos $[0, 1]$, $[0, 1)$ e $(0, 1]$ respectivamente. Examinamos algumas propriedades dessas distribuições e comparamos os métodos de estimação por máxima verossimilhança e momentos condicionais. Adicionalmente, apresentamos aplicações a vários conjuntos de dados reais.

O terceiro capítulo consiste na modelagem de regressão beta inflacionada. Inicialmente consideramos um modelo de regressão beta inflacionado em zero ou um, em que a distribuição da variável resposta é uma mistura entre uma distribuição beta e uma distribuição degenerada no ponto zero ou um. Apesar desta abordagem não ser totalmente inédita (Cook, Kieschnick & McCullough, 2006; Hoff, 2007), alguns aspectos inferenciais

1.2. Suporte computacional

não haviam ainda sido explorados. Neste sentido, desenvolvemos inferência baseada em verossimilhança, apresentando expressões para o vetor escore, a matriz de informação de Fisher, algoritmo de estimação, intervalos de confiança, testes de hipótese e uma aplicação a dados simulados. Estudamos também um modelo de regressão beta inflacionado em zero e um, em que a distribuição da variável resposta é uma mistura entre uma distribuição beta e a distribuição de Bernoulli. Para este modelo discutimos a estimação por máxima verossimilhança, e apresentamos expressões para o vetor escore, a matriz de informação de Fisher e algoritmo de estimação entre outros.

No quarto capítulo, derivamos os vieses de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros dos modelos de regressão beta inflacionados, visto que, em geral, quanto menor é o tamanho de amostra, mais viesados são os estimadores de máxima verossimilhança. Assim, a obtenção de expressões que permitam calcular o viés possibilita determinar estimadores corrigidos que são mais precisos que os não corrigidos. Nesse capítulo fornecemos a fórmula geral de Cox & Snell (1968) para os vieses de segunda ordem dos estimadores de máxima verossimilhança dos modelos de regressão beta inflacionados. Adicionalmente, avaliamos através de estudos de simulação as propriedades dos estimadores de máxima verossimilhança junto a seus respectivos estimadores corrigidos.

No quinto capítulo tratamos do problema de diagnóstico para os modelos de regressão beta inflacionados. Neste sentido, para cada modelo investigado no quarto capítulo, derivamos resíduos tornando possível a construção de bandas de confiança (envelopes simulados) para a detecção de possíveis afastamentos das suposições de cada modelo. Em seguida, desenvolvemos a análise de influência local baseada na curvatura normal conforme (Poon & Poon, 1999). Para o modelo de regressão beta inflacionado em zero ou um consideramos dois esquemas de perturbação, a saber: a ponderação de casos e a perturbação individual de covariáveis. Adicionalmente, para o modelo de regressão beta inflacionado em zero e um optamos pelo esquema de perturbação de casos. No sexto capítulo consideramos uma aplicação a dados reais que envolvem a teoria desenvolvida ao longo da tese. Finalmente, no sétimo capítulo apresentamos as conclusões do trabalho.

1.2 Suporte computacional

As avaliações numéricas ao longo desta tese foram realizadas num computador Athlon, AMD 64 Bits através da linguagem de programação matricial Ox na sua versão 4.1 e da linguagem e ambiente de computação estatística R em sua versão 2.5.0 sob os sistemas operacionais Linux e Windows. Ox foi desenvolvida com base na linguagem de programação C e não apresenta os problemas de eficiência computacional inerentes a outras linguagens de alto nível (e.g., S-PLUS, SAS). Seu uso é aconselhável em tarefas computacionalmente intensivas, como métodos de integração de Monte Carlo. Ox é distribuído gratuitamente para uso acadêmico em http://www.doornik.com/download_oxcons.html. Para maiores detalhes, ver Cribari-Neto & Zarkos (2003) e Doornik (2001).

Os gráficos apresentados ao longo desta tese, assim como os scripts desenvolvidos para estimação e diagnóstico dos modelos de regressão beta inflacionados foram produzidos no ambiente computacional R, que se encontra disponível gratuitamente em <http://www.r-project.org/>. Para maiores detalhes sobre R, ver Ihaka & Gentleman (1996) e Cribari-Neto & Zarkos (1999).

1.2. Suporte computacional

Por fim, a tese foi digitada usando o sistema tipográfico \LaTeX desenvolvido por Leslie Lamport em 1985. \LaTeX consta de uma série de macros ou rotinas do sistema \TeX (Knuth, 1986) que facilitam o desenvolvimento e edição de textos científicos.

Capítulo 2

Agregando zero e/ou um a uma distribuição beta

2.1 Introdução

Uma classe importante de problemas envolve dados na forma de taxas e proporções como, por exemplo, taxas de mortalidade, taxas de infecção de doenças, proporção de indivíduos que admitem ter intenção de voto por um candidato em particular, etc. Se a variável resposta é uma proporção medida de forma contínua no intervalo $(0, 1)$ (como, por exemplo, a composição de rochas, a concentração de agentes químicos ou a proporção de renda gasta em serviços de saúde), é comum o uso de uma transformação nos dados para que estes assumam valores na reta real ou no conjunto dos números reais positivos. Entre as transformações mais usuais para este tipo de dados se encontram: a transformação logito, $g(x) = \log\{x/(1-x)\}$, a transformação probito, $g(x) = \Phi^{-1}(x)$, onde $\Phi(\cdot)$ representa a função de distribuição acumulada de uma variável aleatória normal padrão, a transformação log-log complementar, $g(x) = \log\{-\log(1-x)\}$, a transformação log-log, $g(x) = -\log\{-\log(x)\}$, a transformação angular $g(x) = \sin^{-1}(\sqrt{x})$ e a transformação potência (Box & Cox, 1964)

$$g(x) = \begin{cases} (x^\phi - 1)/\phi, & \text{para } \phi \neq 0, \\ \log_e(x), & \text{para } \phi = 0, \end{cases}$$

entre outras. A utilização destas e outras transformações para dados de proporções encontra-se descrita em Atkinson (1985, cap. 7) e sua escolha depende freqüentemente da interpretabilidade do modelo para os dados. Freqüentemente, dados na forma de taxas ou proporções apresentam zeros e/ou uns. Com exceção da transformação angular, as transformações mencionadas acima não estão definidas para $x = 0$ ou $x = 1$, i.e., proporções observadas com o valor zero ou um. Piepho (2003) propõe uma transformação alternativa, que permite modelar dados de proporções com presença de zeros ou uns. Esta transformação baseia-se numa extensão da transformação exponencial dada por Manly (1976). Contudo, as transformações a dados na forma de frações ou proporções modificam a natureza real dos dados e não possibilitam a interpretação direta dos parâmetros envolvidos no modelo.

O foco deste trabalho é a modelagem estatística de dados distribuídos de forma contínua no intervalo $(0,1)$ mas que incluem observações em um ou em ambos extremos. A idéia consiste em assumir que a distribuição

2.1. Introdução

dos dados é uma mistura entre uma distribuição contínua definida no intervalo $(0, 1)$ e a distribuição de Bernoulli, a qual atribui probabilidades aos inteiros 0 e 1. Se os dados são observados no intervalo $(0, 1]$ ou $[0, 1)$ assume-se que a distribuição de probabilidade dos dados é uma mistura entre uma variável distribuída de forma contínua em $(0, 1)$ e uma distribuição degenerada concentrada em 0 ou 1, dependendo do caso. Nesta situação, o modelo proposto faz parte da classe dos modelos inflacionados. A palavra *inflacionado* sugere que a massa de probabilidade de alguns pontos excede o que é permitido pelo modelo proposto. Por exemplo, em dados de contagens, o modelo de Poisson é comumente utilizado e o conjunto de dados estará inflacionado com zeros se a porcentagem de zeros for superior ao que a função de distribuição de Poisson pode associar. Muitos pesquisadores têm analisado modelos inflacionados, em particular modelos para dados discretos com excesso de zeros (Tu, 2002). Nesses modelos, a principal suposição é que a distribuição dos dados é gerada por uma mistura entre uma distribuição padrão de contagem e uma distribuição degenerada concentrada no ponto zero.

Na literatura, alguns destes modelos são conhecidos como *zero-inflated binomial model* (ZIB), *zero-inflated Poisson model* (ZIP) e *zero-inflated negative binomial model* (ZINB). Ridout, Demétrio & Hinde (1998) fornecem uma revisão da literatura sobre o modelo ZIP e de alguns modelos inflacionados com zeros para dados de contagens e, mais recentemente, Hall (2000) e Vieira, Hinde & Demétrio (2000) introduziram modelos de regressão ZIB. Na maioria dos trabalhos sobre modelos ZIP e ZIB se assume uma estrutura de regressão em que os parâmetros da regressão são desconhecidos e se atribui independência entre todas as observações na série de dados. Recentemente, modelos para dados contínuos inflacionados com zeros têm despertado o interesse de pesquisadores dadas as possibilidades de aplicações práticas. A idéia de misturar uma distribuição degenerada no ponto zero com uma distribuição contínua surgiu com os trabalhos de Aitchison (1955, 1969), que introduziu uma mistura de uma distribuição degenerada em zero com uma distribuição lognormal a qual denominou distribuição *delta*. Esta distribuição tem sido usada em estudos econômicos de saúde, por exemplo, na investigação de custos médicos, onde uma grande proporção da população não pode cobrir os custos médicos enquanto relativamente poucos pacientes gravemente doentes dispõem valores altos (Tu & Zhou, 1999). Feuerverger (1979) utilizou uma mistura de uma distribuição degenerada em zero com uma distribuição gama. Este modelo é usado freqüentemente em estudos ambientais como, por exemplo, a predição de chuva, em que os zeros representam os casos de não-precipitação (dias secos) e a distribuição gama é usada para modelar os níveis de precipitação em dias com chuva.

Yoo (2004) considerou um modelo de mistura para modelar despesas com telefonia móvel (*mobile communication expenditure*; MCE). Como alguns indivíduos não gastam em telefonia móvel, o autor considera que a distribuição de probabilidades do MCE é uma mistura entre uma distribuição degenerada em zero e uma outra com suporte nos reais positivos como, por exemplo, gama, exponencial ou Weibull. Heller, Stasinopoulos & Rigby (2006) consideraram uma mistura entre a distribuição normal inversa e uma distribuição degenerada em zero para modelar dados de demanda em seguros. Aqui, a distribuição normal inversa acomoda a assimetria à direita da distribuição da demanda e a ausência de demanda é acomodada pela distribuição degenerada em zero. Cook, Kieschnick & McCullough (2006) propuseram um modelo de regressão para dados de financiamento em que uma suposição é que a variável resposta observada no intervalo $[0, 1)$ segue uma distribuição de mistura entre uma distribuição degenerada no zero e uma distribuição beta. Para este modelo, a média

2.2. A distribuição beta

da distribuição beta é modelada através de um preditor linear usando a ligação logito. Mais recentemente, Hoff (2007) compara diferentes modelos para escores de eficiência DEA (análise envoltória de dados) em função de variáveis exógenas. O autor compara a modelagem Tobit e dois modelos alternativos, entre eles um modelo de regressão beta inflacionado no ponto um. Para este modelo se assume que a distribuição do escore de eficiência observado no intervalo $(0,1]$ segue uma distribuição de mistura entre uma distribuição degenerada no ponto um e uma distribuição beta. Aqui, a probabilidade de que o escore de eficiência é igual a um é modelada através de um preditor linear usando a ligação logito e a média da distribuição beta é modelada através de um outro preditor linear usando a ligação logito.

Finalmente, Lesaffre, Rizopoulos & Tsonaka (2007) sugerem uma metodologia para a modelagem de dados observados no intervalo $[0, 1]$. Assume-se a existência de uma variável latente em $(0, 1)$ dado que a variável resposta é observada em $[0, 1]$. Aqui, a t -ésima observação é da forma $y_t = r_t/N_t$, onde $r_t \sim \text{Bin}(U_t, N_t)$ e U_t segue uma distribuição logito-normal. No entanto, se os N_t 's não são conhecidos (como é o caso de muitos exemplos aplicados), este modelo não pode ser utilizado. Os autores também consideram a situação em que a variável resposta é uma versão “grosseira” de uma variável latente com distribuição logito-normal em $(0, 1)$. Em outras palavras, eles assumem que o logito da variável latente transformada tem distribuição normal.

Este capítulo se encontra organizado da seguinte forma. Na Seção 2.2 apresentamos a distribuição beta sob uma parametrização em termos da média da distribuição e um parâmetro de precisão. Na Seção 2.3 definimos a distribuição beta inflacionada em zero ou um e discutimos algumas propriedades. Apresentamos expressões para o vetor escore, a matriz de informação de Fisher e estimadores obtidos por momentos condicionais e de máxima verossimilhança. Adicionalmente, discutimos um procedimento para encontrar os estimadores de máxima verossimilhança dos parâmetros dessa distribuição. Na Seção 2.4 definimos a distribuição beta inflacionada em zero e um. Mostramos algumas de suas propriedades, tais como o fato de pertencer à família exponencial. Para este modelo obtemos o vetor escore e a matriz de informação de Fisher, intervalos de confiança assintóticos e estimadores derivados por momentos condicionais e de máxima verossimilhança. Na Seção 2.5, através de simulação de Monte Carlo, comparamos os estimadores de momentos e de máxima verossimilhança dos parâmetros que indexam as distribuições beta inflacionadas segundo diferentes tamanhos de amostra. Na Seção 2.6 ajustamos as distribuições propostas a vários conjuntos de dados reais. Finalmente, as conclusões deste capítulo são apresentadas na Seção 2.7.

2.2 A distribuição beta

Kieschnick & McCullagh (2003) e Johnson, Kotz & Balakrishnan (1995, p. 235) destacam que a distribuição beta é um modelo apropriado para descrever dados distribuídos de forma contínua no intervalo $(0, 1)$ dada a grande flexibilidade de formas dessa distribuição. Bury (1999) lista um conjunto de aplicações da distribuição beta em engenharia. Janardan & Padmanabhan (1986) modelam variáveis hidrológicas usando a distribuição beta. McNally (1990) utiliza a distribuição beta num estudo referente à capacidade de reprodução em vacas. Graham & Hollands (1990) e Milyutin & Yaromenko (1991) usam a distribuição beta para estudar índices relacionados à transmissão de radiação solar. A potência de sinais de radar é modelada por Maffet

2.3. A distribuição beta inflacionada em zero ou um

& Wackerman (1991) através de uma distribuição beta modificada. Wiley, Herschokoru & Padiou (1989) desenvolvem um modelo beta para estimar a probabilidade de transmissão de HIV durante o contato sexual entre um indivíduo infectado e um indivíduo sadio.

Dizemos que a variável aleatória y tem distribuição beta com parâmetros de forma $p, q > 0$ se sua função de densidade é dada por

$$\pi(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \quad y \in (0, 1),$$

onde $\Gamma(\cdot)$ é a função gama, i.e., $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. Neste trabalho, utilizamos uma parametrização alternativa da distribuição beta. Sejam $\mu = p/(p+q)$ e $\phi = p+q$, i.e., $p = \mu\phi$ e $q = (1-\mu)\phi$. A função de densidade de y pode ser escrita como

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad y \in (0, 1), \quad (2.2.1)$$

$0 < \mu < 1$ e $\phi > 0$. Dizemos que y tem distribuição beta com média μ e precisão ϕ e escrevemos $y \sim \mathcal{B}(\mu, \phi)$. Evidentemente, através da escolha de (μ, ϕ) , obtêm-se diferentes densidades. Assim, na realidade, (2.2.1) forma uma classe de densidades beta.

A Figura 2.1 apresenta diferentes densidades beta correspondentes a alguns valores dos parâmetros (μ, ϕ) . Podemos observar que, sob as diferentes escolhas dos parâmetros, a forma da densidade muda. Em particular, quando $\mu = 1/2$ as densidades se apresentam com forma simétrica; no entanto, quando $\mu \neq 1/2$ as formas são assimétricas. Adicionalmente, notamos a presença de densidades com forma de ‘J’, ‘U’ e de ‘J’ invertido, sendo que o formato ‘U’ ocorre quando $(1-\mu)\phi < 1$ e $\mu\phi < 1$. Já as densidades com forma de ‘J’ são obtidas quando $\{(1-\mu)\phi - 1\}(\mu\phi - 1) < 0$ e as densidades com forma de ‘J’ invertido são obtidas quando $\{(1-\mu)\phi - 1\}(\mu\phi - 1) > 0$.

Para a distribuição beta (2.2.1), deduz-se que o momento de ordem r ao redor de zero é

$$\mu_r = E(y^r) = \frac{\Gamma(\phi)\Gamma(\mu\phi+r)}{\Gamma(\phi+r)\Gamma(\mu\phi)} = \frac{(\mu\phi)_{(r)}}{\phi_{(r)}}, \quad (2.2.2)$$

onde $a_{(r)} = a(a+1)(a+2)\cdots(a+r-1)$. Desta forma, $E(y) = \mu$ e $\text{Var}(y) = V(\mu)/(\phi+1)$ onde $V(\mu) = \mu(1-\mu)$, denota a “função de variância”. O parâmetro μ é a média de y e o parâmetro ϕ é de precisão, no sentido de que, para μ fixo, quanto maior for o valor de ϕ , menor será a variância de y .

2.3 A distribuição beta inflacionada em zero ou um

Em situações práticas, dados na forma de proporções podem incluir zeros ou uns. Empiricamente isto pode estar relacionado a alguma intervenção, truncamento ou censura nos dados. Nesta situação, a distribuição beta não é um modelo adequado para os dados. Se os dados contêm zeros ou uns (mas não ambos) um modelo

2.3. A distribuição beta inflacionada em zero ou um

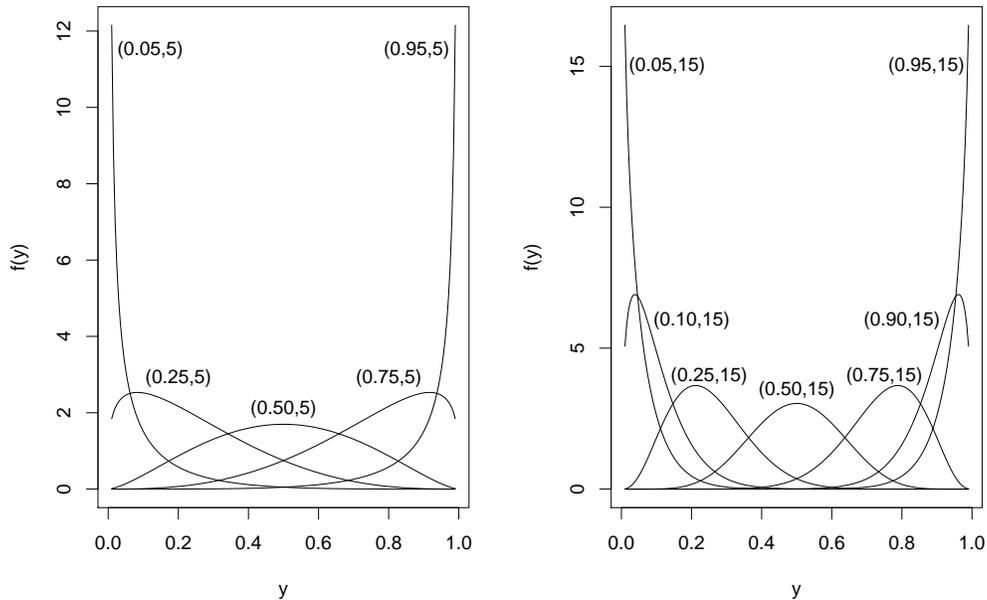


Figura 2.1: Densidades beta para diferentes valores de (μ, ϕ) .

natural consiste em adicionar à distribuição beta um ponto de massa em zero ou um. Desta forma, podemos obter modelos para frações observadas nos intervalos $[0, 1)$ ou $(0, 1]$. Neste contexto, vamos supor que o componente contínuo dos dados é modelado pela distribuição beta (2.2.1) uma vez que esta é bastante flexível para ajustar dados no intervalo $(0, 1)$. Já o componente discreto, i.e., o ponto de massa, será modelado através de uma distribuição degenerada no valor conhecido c , onde c é igual a zero ou um dependendo do caso.

A função de distribuição acumulada da mistura é dada por

$$\text{BI}_c(y; \alpha, \mu, \phi) = \alpha \mathbb{1}_{\{c\}}(y) + (1 - \alpha)F(y; \mu, \phi), \quad (2.3.1)$$

sendo $\mathbb{1}_A(y)$ a função indicadora, com valor 1 se $y \in A$ e 0 se $y \notin A$. A função $F(\cdot; \mu, \phi)$ é a função de distribuição acumulada beta $\mathcal{B}(\mu, \phi)$ e $0 < \alpha < 1$ é o parâmetro de mistura. A função de distribuição BI_c não é absolutamente contínua pois tem um ponto de massa em $y = c$. Note que, com probabilidade α , a variável y é selecionada de uma distribuição degenerada no ponto c e com probabilidade $1 - \alpha$, a variável y é selecionada de uma distribuição beta que é absolutamente contínua com respeito à medida de Lebesgue.

2.3. A distribuição beta inflacionada em zero ou um

A função de densidade de probabilidade de y com respeito à medida gerada pela mistura* é da forma

$$\text{bi}_c(y; \alpha, \mu, \phi) = \begin{cases} \alpha, & \text{se } y = c, \\ (1 - \alpha)f(y; \mu, \phi), & \text{se } y \in (0, 1), \end{cases} \quad (2.3.2)$$

com $0 < \alpha, \mu < 1$ e $\phi > 0$, sendo $f(y; \mu, \phi)$ a função densidade (2.2.1) e escrevemos $y \sim \text{BI}_c(\alpha, \mu, \phi)$. Note que $\alpha = P(y = c)$ representa a probabilidade de se observar zero ($c = 0$) ou um ($c = 1$), conforme o caso. Já μ e ϕ são parâmetros da distribuição beta (2.2.1). Segundo a definição de distribuição inflacionada dada em Tu (2002), a densidade (2.3.2) é uma distribuição beta inflacionada no ponto c , $c = 0$ ou $c = 1$, já que, se $\alpha > 0$, a massa de probabilidade da distribuição beta em $y = c$ é excedida. A densidade (2.3.2) pode ser escrita na forma

$$\text{bi}_c(y; \alpha, \mu, \phi) = \left\{ \alpha^{\mathbb{1}_{\{c\}}(y)} (1 - \alpha)^{1 - \mathbb{1}_{\{c\}}(y)} \right\} \left\{ f(y; \mu, \phi)^{1 - \mathbb{1}_{\{c\}}(y)} \right\}.$$

Esta densidade, como função dos parâmetros, fatora em dois termos, sendo que o primeiro depende somente de α e o segundo depende unicamente de (μ, ϕ) .

Definição 2.3.1. *Seja y uma variável aleatória que segue a distribuição beta inflacionada (2.3.2).*

1. *Se $c = 0$, a distribuição (2.3.2) é chamada de distribuição beta inflacionada no ponto zero (BIZ) e escrevemos $y \sim \text{BIZ}(\alpha, \mu, \phi)$.*
2. *Se $c = 1$, a distribuição (2.3.2) é chamada de distribuição beta inflacionada no ponto um (BIU) e escrevemos $y \sim \text{BIU}(\alpha, \mu, \phi)$.*

Se $y \sim \text{BIZ}(\alpha, \mu, \phi)$, então $\alpha = P(y = 0)$ e se $y \sim \text{BIU}(\alpha, \mu, \phi)$, então, $\alpha = P(y = 1)$. Assim, essas distribuições permitem agregar o valor zero ou o valor um à distribuição beta (2.2.1).

Para obter $E(y^r)$ e $\text{Var}(y)$ usamos as igualdades: $E(y^r) = E[E(y^r | \mathbb{1}_{\{c\}}(y))]$ e $\text{Var}(y) = E[\text{Var}(y | \mathbb{1}_{\{c\}}(y))] + \text{Var}[E(y | \mathbb{1}_{\{c\}}(y))]$, em que

$$\begin{aligned} E(y^r | \mathbb{1}_{\{c\}}(y)) &= \begin{cases} c, & \text{com probabilidade } \alpha, \\ \mu^r, & \text{com probabilidade } 1 - \alpha, \end{cases} \\ \text{Var}(y | \mathbb{1}_{\{c\}}(y)) &= \begin{cases} 0, & \text{com probabilidade } \alpha, \\ \frac{\mu(1-\mu)}{\phi+1}, & \text{com probabilidade } 1 - \alpha. \end{cases} \end{aligned} \quad (2.3.3)$$

*A medida de probabilidade P correspondente a $\text{BI}_c(y; \cdot)$ definida sobre o espaço mensurável $((0, 1) \cup \{c\}, \mathfrak{B})$, sendo \mathfrak{B} a classe de todos os subconjuntos de Borel de $(0, 1) \cup \{c\}$, é tal que $P \ll \lambda + \delta_c$, sendo λ a medida de Lebesgue e δ_c o ponto de massa em c ; i.e., $\delta_c(A) = 1$ se $c \in A$ e $\delta_c(A) = 0$ se $c \notin A$, $A \in \mathfrak{B}$.

2.3. A distribuição beta inflacionada em zero ou um

Assim,

$$\begin{aligned} E(y^r) &= \alpha c + (1 - \alpha)\mu_r, \\ \text{Var}(y) &= (1 - \alpha)\frac{V(\mu)}{\phi + 1} + \alpha(1 - \alpha)(c - \mu)^2, \end{aligned} \quad (2.3.4)$$

onde $V(\mu) = \mu(1 - \mu)$ e μ_r é o r -ésimo momento ao redor de zero da distribuição $\mathcal{B}(\mu, \phi)$ dada em (2.2.1). Note que $E(y^r)$ é a média ponderada entre o r -ésimo momento ao redor de zero da distribuição degenerada em c e o correspondente momento da distribuição $\mathcal{B}(\mu, \phi)$ com pesos α e $1 - \alpha$, respectivamente.

Das equações (2.3.4) temos que a média e a variância da distribuição BIZ são

$$\begin{aligned} E(y) &= (1 - \alpha)\mu, \\ \text{Var}(y) &= (1 - \alpha)\frac{V(\mu)}{\phi + 1} + \alpha(1 - \alpha)\mu^2. \end{aligned}$$

Para a distribuição BIU temos que a média e a variância são

$$\begin{aligned} E(y) &= \alpha + (1 - \alpha)\mu, \\ \text{Var}(y) &= (1 - \alpha)\frac{V(\mu)}{\phi + 1} + \alpha(1 - \alpha)(1 - \mu)^2. \end{aligned}$$

Notamos que para ambas as distribuições, se α e μ estão fixos, quanto maior ϕ menores as variâncias destas distribuições. Então, ϕ pode ser interpretado como um parâmetro de precisão.

A Figura 2.2 apresenta gráficos da distribuição BIZ para diferentes escolhas de μ e ϕ deixando o parâmetro α fixo. Note que para todos μ e ϕ a distribuição BIZ se apresenta com formas assimétricas devido à presença de massa de probabilidade no ponto zero. Percebemos nestes gráficos que a densidade em $(0, 1)$ se apresenta com formas unimodais, ‘J’, ‘U’ e ‘J’ invertida, entre outras. Nestes gráficos, a barra vertical com o ponto acima representa $\alpha = P(y = 0)$. De forma análoga à distribuição BIZ, a distribuição BIU é assimétrica devido à presença de massa de probabilidade no ponto um e, para idênticas escolhas dos parâmetros, as distribuições BIZ e BIU se apresentam com a mesma forma funcional no intervalo $(0,1)$. No entanto, elas se diferenciam no ponto de massa associado, sendo em zero para a distribuição BIZ e em um para a distribuição BIU.

No contexto das famílias exponenciais, Seshadri (1991) estuda o comportamento da função de variância na mistura finita de famílias exponenciais naturais, Lindsay (1995) analisa a geometria da mistura de famílias exponenciais uniparamétricas, Landsman & Makov (2003) propõem uma família de dispersão contaminada para modelos de perda através da mistura de modelos exponenciais de dispersão e Sum & Oommen (1994) examinam o método de momentos para estimar os parâmetros da distribuição de mistura e seus componentes, sendo que a distribuição resultante é obtida através da mistura de densidades da família exponencial com parâmetro de forma conhecido. Nesta linha de idéias, podemos mostrar que a densidade (2.3.2) pertence à família exponencial.

2.3. A distribuição beta inflacionada em zero ou um

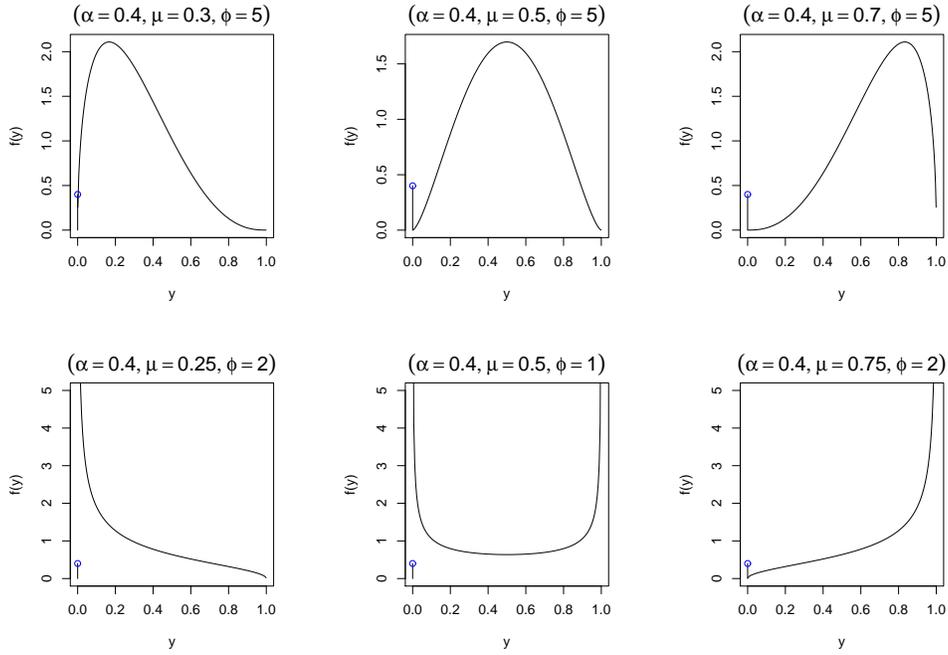


Figura 2.2: Densidades BIZ para diferentes escolhas de μ e ϕ ; $\alpha = 0.4$.

Proposição 2.3.1. *A distribuição beta inflacionada no ponto $c = 0$ ou $c = 1$ dada em (2.3.2) pertence à família exponencial de dimensão 3 de posto completo.*

Prova : Sejam $\eta = (\eta_1, \eta_2, \eta_3)$, com $\eta_1 = [\log(\alpha/(1 - \alpha)) + b(\eta_2, \eta_3)]$, $\eta_2 = \mu\phi$ e $\eta_3 = (1 - \mu)\phi$, sendo $b(\eta_2, \eta_3) = \log(\Gamma(\eta_2)\Gamma(\eta_3)/\Gamma(\eta_2 + \eta_3))$. Considere o vetor de estatísticas $T(y) = (t_1(y), t_2(y), t_3(y))$, onde

$$t_1(y) = \begin{cases} 1, & \text{se } y = c, \\ 0, & \text{se } y \in (0, 1). \end{cases}$$

$$t_2(y) = \begin{cases} \log y, & \text{se } y \in (0, 1), \\ 0, & \text{se } y = c, \end{cases}$$

$$t_3(y) = \begin{cases} \log(1 - y), & \text{se } y \in (0, 1), \\ 0, & \text{se } y = c. \end{cases}$$

2.3. A distribuição beta inflacionada em zero ou um

A distribuição (2.3.2) pode ser escrita como

$$\exp\{\eta^\top T(y) - B^*(\eta)\}h(y), \quad (2.3.5)$$

onde

$$B^*(\eta) = \log\{1 + \exp[\eta_1 - b(\eta_2, \eta_3)]\} + b(\eta_2, \eta_3)$$

é uma função em η a valores reais e

$$h(y) = \begin{cases} 1/\{y(1-y)\}, & \text{se } y \in (0, 1), \\ 0, & \text{se } y = c, \end{cases}$$

é uma função não-negativa definida no conjunto $(0, 1) \cup \{c\}$. A parametrização η define uma transformação bijetora que leva $\mathfrak{X} = (0, 1) \times (0, 1) \times \mathbb{R}^+$ a $\mathfrak{D} = \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}$, um subconjunto aberto de \mathbb{R}^3 . Adicionalmente, é possível mostrar que os t 's e os η 's são linearmente independentes e o espaço paramétrico contém retângulos tridimensionais. Logo, (2.3.5) é a representação canônica da distribuição beta inflacionada no ponto c na família exponencial de dimensão 3 de posto completo. \boxtimes

Consideremos agora y_1, \dots, y_n uma amostra aleatória, onde cada y_t tem densidade (2.3.2). Como consequência da Proposição 2.3.1 o vetor de estatísticas $\sum_{t=1}^n T(y_t) = (T_1, T_2, T_3)$, em que

$$\begin{aligned} T_1 &= \sum_{t=1}^n \mathbb{1}_{\{c\}}(y_t), \\ T_2 &= \sum_{t:y_t \in (0,1)} \log y_t, \\ T_3 &= \sum_{t:y_t \in (0,1)} \log(1 - y_t), \end{aligned} \quad (2.3.6)$$

é uma estatística suficiente completa (Lehmann & Casella, 2002, Corolário 1.6.16 e Teorema 1.6.22). Note ainda que $T_1 \sim \text{Bin}(\alpha, n)$, uma vez que $P(\mathbb{1}_{\{c\}}(y_t) = 1) = P(y_t = c) = \alpha$, para $t = 1, \dots, n$.

A função de verossimilhança para $\theta = (\alpha, \mu, \phi)$ dada a amostra (y_1, \dots, y_n) é

$$L(\theta) = \prod_{t=1}^n \text{bi}_c(y_t; \alpha, \mu, \phi) = L_1(\alpha)L_2(\mu, \phi), \quad (2.3.7)$$

2.3. A distribuição beta inflacionada em zero ou um

em que

$$L_1(\alpha) = \prod_{t=1}^n \alpha^{\mathbb{1}_{\{c\}}(y_t)} (1 - \alpha)^{1 - \mathbb{1}_{\{c\}}(y_t)} = \alpha^{T_1} (1 - \alpha)^{n - T_1},$$

$$L_2(\mu, \phi) = \prod_{t=1}^n f(y_t; \mu, \phi)^{1 - \mathbb{1}_{\{c\}}(y_t)}.$$

Neste caso, a função de verossimilhança $L(\theta)$ fatora em dois termos: um que depende apenas de α e outro que depende apenas de (μ, ϕ) . Assim, os parâmetros são separáveis (Pace & Salvan, 1997, p. 128) e a inferência por máxima verossimilhança sobre (μ, ϕ) pode ser realizada de forma independente do parâmetro α como se esse fosse conhecido, e vice-versa.

O logaritmo da função de verossimilhança assume a forma

$$\ell(\theta) = \log(L(\theta)) = \ell_1(\alpha) + \ell_2(\mu, \phi),$$

onde

$$\ell_1(\alpha) = T_1 \log \alpha + (n - T_1) \log(1 - \alpha),$$

$$\ell_2(\mu, \phi) = (n - T_1) \log \left\{ \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} \right\} + T_2(\mu\phi - 1) + T_3((1 - \mu)\phi - 1).$$

Derivando $\ell(\theta)$ com respeito a cada parâmetro desconhecido, encontramos o vetor escore

$$U(\alpha, \mu, \phi) = (U_\alpha(\alpha), U_\mu(\mu, \phi), U_\phi(\mu, \phi)),$$

em que

$$U_\alpha(\alpha) = \frac{T_1}{\alpha} - \frac{(n - T_1)}{1 - \alpha},$$

$$U_\mu(\mu, \phi) = \phi \{ (n - T_1) [\psi((1 - \mu)\phi) - \psi(\mu\phi)] + T_2 - T_3 \},$$

$$U_\phi(\mu, \phi) = (n - T_1) [\psi(\phi) - \mu\psi(\mu\phi) - (1 - \mu)\psi((1 - \mu)\phi)] + \mu T_2 - (1 - \mu) T_3.$$

Pela separabilidade dada em (2.3.7), a solução do sistema $U_\alpha(\alpha) = 0$ fornece uma expressão em forma fechada para o estimador de máxima verossimilhança de α dada por $\hat{\alpha} = T_1/n$. Este estimador representa a proporção de zeros ($c = 0$) ou de uns ($c = 1$) na amostra. Adicionalmente, $\hat{\alpha}$ é função de uma estatística suficiente completa e é um estimador não-viesado de α , isto é, $E(\hat{\alpha}) = \alpha$. Logo, $\hat{\alpha}$ é um estimador não-viesado de variância uniformemente mínima para α (Lehmann & Casella, 2002, Teorema 2.1.11) com variância dada por $\text{Var}(\hat{\alpha}) = \alpha(1 - \alpha)/n$. Por outro lado, os estimadores de μ e ϕ não podem ser obtidos em forma fechada. Assim, eles devem ser obtidos numericamente pela maximização da função de log-verossimilhança $\ell_2(\mu, \phi)$ usando um algoritmo de otimização não-linear (Newton-Raphson, escore de Fisher, BHHH, etc.); para maiores detalhes, ver Nocedal & Wright (1999).

2.3. A distribuição beta inflacionada em zero ou um

Recentemente, as distribuições BIZ e BIU foram incorporadas no pacote `gam1ss` do ambiente computacional \mathbb{R} (Ospina, 2006), as estimativas por máxima verossimilhança dos parâmetros destas distribuições sendo obtidas usando uma versão ajustada do algoritmo *score* de Fisher (Cole & Green, 1992).

Alternativamente à estimação por máxima verossimilhança, podemos obter estimadores para (μ, ϕ) baseados nos momentos condicionais de y dado que $y \in (0, 1)$. Note que $E(y|y \in (0, 1)) = \mu$ e $\text{Var}(y|y \in (0, 1)) = \mu(1 - \mu)/(\phi + 1)$. Para $T_1 < n$,[†] a solução do sistema de equações $(\bar{y}, s^2)^\top = (\mu, \mu(1 - \mu)/(\phi + 1))^\top$, com $\bar{y} = \sum_{t:y_t \in (0,1)} y_t / (n - T_1)$ e $s^2 = \sum_{t:y_t \in (0,1)} (y_t - \bar{y})^2 / (n - T_1)$, fornece os seguintes estimadores de forma fechada para os parâmetros da distribuição beta inflacionada (2.3.2): $\tilde{\mu} = \bar{y}$ e $\tilde{\phi} = \{\tilde{\mu}(1 - \tilde{\mu})/s^2\} - 1$. Estimadores de forma fechada para $E(y^r)$ e $\text{Var}(y)$ podem ser obtidos ao substituir α, μ e ϕ por $\hat{\alpha}, \tilde{\mu}$ e $\tilde{\phi}$ nas equações dadas em (2.3.4).

A matriz de informação de Fisher para a distribuição beta inflacionada no ponto c é

$$K(\theta) = \begin{pmatrix} \kappa_{\alpha\alpha} & 0 & 0 \\ 0 & \kappa_{\mu\mu} & \kappa_{\mu\phi} \\ 0 & \kappa_{\phi\mu} & \kappa_{\phi\phi} \end{pmatrix}, \quad (2.3.8)$$

em que

$$\begin{aligned} \kappa_{\alpha\alpha} &= 1/\{\alpha(1 - \alpha)\}, \\ \kappa_{\mu\mu} &= (1 - \alpha)\phi^2\{\psi'(\mu\phi) + \psi'((1 - \mu)\phi)\}, \\ \kappa_{\mu\phi} &= \kappa_{\phi\mu} = (1 - \alpha)\phi\{\psi'(\mu\phi)\mu - \psi'((1 - \mu)\phi)(1 - \mu)\}, \\ \kappa_{\phi\phi} &= (1 - \alpha)\{\mu^2\psi'(\mu\phi) + (1 - \mu)^2\psi'((1 - \mu)\phi) - \psi'(\phi)\}. \end{aligned}$$

Observamos que esta matriz não depende do valor de c . Contudo, devemos lembrar que para a distribuição BIZ o parâmetro α representa $P(y = 0)$ enquanto que para a distribuição BIU o parâmetro α representa $P(y = 1)$. Note que na matriz de informação de Fisher temos $\kappa_{\alpha\mu} = \kappa_{\mu\alpha} = \kappa_{\alpha\phi} = \kappa_{\phi\alpha} = 0$, indicando que o parâmetro α é ortogonal ao vetor de parâmetros (μ, ϕ) (Cox & Reid, 1987). Esta ortogonalidade implica que os respectivos componentes do vetor *score* são não correlacionados e $\hat{\alpha}$ e $(\hat{\mu}, \hat{\phi})$ são assintoticamente independentes.

Através do resultado da Proposição 2.3.1 pode ser obtida a normalidade assintótica do estimador de máxima verossimilhança (Bickel & Doksum, 2001, Teorema 5.3.5, p. 322), i.e.,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, K(\theta)^{-1}),$$

em que $\xrightarrow{\mathcal{D}}$ denota convergência em distribuição, $\hat{\theta} = (\hat{\alpha}, \hat{\mu}, \hat{\phi})$ é o estimador de máxima verossimilhança de $\theta = (\alpha, \mu, \phi)$, \mathcal{N}_3 representa a distribuição normal trivariada e $K(\theta)$ é a matriz de informação de Fisher dada em (2.3.8). Além disso, pela normalidade assintótica e dada a consistência do estimador de máxima

[†]Se $T_1 = n$ (todas as observações são iguais a c) não recomendamos o uso das distribuições BIZ e BIU como modelos para dados de proporções.

2.4. A distribuição beta inflacionada em zero e um

verossimilhança $\hat{\theta}$, podemos construir intervalos de confiança de tipo assintótico para os parâmetros desta distribuição. Por exemplo, temos que

$$\hat{\mu} \pm z_{1-\frac{\alpha}{2}}(\hat{\kappa}^{\mu\mu})^{1/2}$$

são os limites do intervalo de confiança assintótico (ICA) para μ de coeficiente de confiança $100(1 - \alpha)\%$, sendo $z_{1-\alpha/2}$ o quantil $(1 - \alpha/2)$ da distribuição $\mathcal{N}(0, 1)$ e $\hat{\kappa}^{\mu\mu}$ é a variância assintótica de $\hat{\mu}$ obtida do inverso da matriz $K(\theta)$ dada em (2.3.8) sendo avaliada no estimador de máxima verossimilhança.

Se estamos interessados em estimar $r(\theta)$, uma função de θ , o método delta (Lehmann & Casella 1998, § 1.9) pode ser usado para obter a distribuição assintótica do estimador de máxima verossimilhança $r(\hat{\theta})$. Se $r(\theta)$ é diferenciável, então $\sqrt{n}(r(\hat{\theta}) - r(\theta)) \xrightarrow{D} \mathcal{N}(0, \lambda(\theta))$, em que $\lambda(\theta) = \dot{r}(\theta)^\top K(\theta)^{-1} \dot{r}(\theta)$ com $\dot{r}(\theta) = \partial r(\theta) / \partial \theta$. Em particular, o estimador de máxima verossimilhança de $E(y) = \alpha c + (1 - \alpha)\mu$ é $\widehat{E}(y) = \hat{\alpha}c + (1 - \hat{\alpha})\hat{\mu}$ e a variância da distribuição limite é $(c - \mu)^2 \kappa^{\alpha\alpha} + (1 - \alpha)^2 \kappa^{\mu\mu}$, em que $\kappa^{\alpha\alpha} = 1/\kappa_{\alpha\alpha}$ e $\kappa^{\mu\mu}$ é o elemento (2, 2) de $K(\theta)^{-1}$. De forma análoga, podemos obter a distribuição assintótica do estimador de máxima verossimilhança de $\text{Var}(y)$.

2.4 A distribuição beta inflacionada em zero e um

Para proporções observadas no intervalo $[0, 1]$ admitimos que as probabilidades de observar os valores zero e um são positivas. Para este tipo de dados, utilizamos uma distribuição derivada da mistura entre uma distribuição beta e uma distribuição de Bernoulli, a qual atribui probabilidades não-negativas aos inteiros 0 e 1. Aqui, a distribuição beta serve para modelar o componente contínuo dos dados e a distribuição de Bernoulli ajusta o componente discreto, i.e., os pontos de massa em zero e um. Especificamente, admitimos que a função de distribuição acumulada dos dados, i.e., de y seja uma combinação convexa de duas funções de distribuição acumuladas na forma

$$\text{BIZU}(y; \alpha, \gamma, \mu, \phi) = \alpha \text{Ber}(y; \gamma) + (1 - \alpha)F(y; \mu, \phi), \quad (2.4.1)$$

em que $\text{Ber}(\cdot; \gamma)$ representa a função de distribuição acumulada de uma variável aleatória de Bernoulli de parâmetro γ e $F(\cdot; \mu, \phi)$ é a função de distribuição acumulada de $\mathcal{B}(\mu, \phi)$. Os parâmetros são $0 < \mu, \gamma, \alpha < 1$ e $\phi > 0$, onde o parâmetro de mistura α permite combinar de forma convexa as duas distribuições.

Segundo a definição de distribuição inflacionada dada em Tu (2002), a distribuição de mistura em (2.4.1) é, na realidade, uma distribuição beta inflacionada por zero e um, uma vez que as massas de probabilidade em zero e um excedem o que é permitido pela função de distribuição beta (2.2.1). Note que a função de distribuição BIZU não é absolutamente contínua pois tem pontos de massa em $y = 0$ e $y = 1$. Percebemos que, com probabilidade α , a variável y é selecionada de uma distribuição de Bernoulli e, com probabilidade $1 - \alpha$, a variável y é selecionada de uma distribuição beta.

Definição 2.4.1. *Seja y uma variável aleatória que assume valores no intervalo fechado $[0, 1]$. Dizemos que y tem distribuição beta inflacionada em zero e um (BIZU) de parâmetros $(\alpha, \gamma, \mu, \phi)$ se sua função de densidade*

2.4. A distribuição beta inflacionada em zero e um

com respeito à medida gerada pela mistura[‡] é da forma

$$\text{bizu}(y; \alpha, \gamma, \mu, \phi) = \begin{cases} \alpha\gamma, & \text{se } y = 1, \\ \alpha(1 - \gamma), & \text{se } y = 0, \\ (1 - \alpha)f(y; \mu, \phi), & \text{se } y \in (0, 1), \end{cases} \quad (2.4.2)$$

com $0 < \alpha, \gamma, \mu < 1$ e $\phi > 0$, sendo $f(y; \mu, \phi)$ a função de densidade beta (2.2.1), e escrevemos $y \sim \text{BIZU}(\alpha, \gamma, \mu, \phi)$.

Note que $P(y = 1) = \alpha\gamma$, $P(y = 0) = \alpha(1 - \gamma)$ e, para $0 < a < b < 1$,

$$P(y \in (a, b)) = (1 - \alpha) \int_a^b f(y; \mu, \phi) dy.$$

Podemos calcular $E(y^r)$ e $\text{Var}(y)$ usando as igualdades: $E(y^r) = E[E(y|\mathbb{1}_{\{0,1\}}(y))]$ e $\text{Var}(y) = E[\text{Var}(y|\mathbb{1}_{\{0,1\}}(y))] + \text{Var}[E(y|\mathbb{1}_{\{0,1\}}(y))]$, em que

$$\begin{aligned} E(y^r | \mathbb{1}_{\{0,1\}}(y)) &= \begin{cases} \gamma, & \text{com probabilidade } \alpha, \\ \mu_r, & \text{com probabilidade } 1 - \alpha, \end{cases} \\ \text{Var}(y | \mathbb{1}_{\{0,1\}}(y)) &= \begin{cases} \gamma(1 - \gamma), & \text{com probabilidade } \alpha, \\ \frac{\mu(1 - \mu)}{\phi + 1}, & \text{com probabilidade } 1 - \alpha. \end{cases} \end{aligned}$$

Desta forma, podemos mostrar que

$$\begin{aligned} E(y^r) &= \alpha\gamma + (1 - \alpha)\mu_r, \\ \text{Var}(y) &= \alpha V_1 + (1 - \alpha)V_2 + \alpha(1 - \alpha)(\gamma - \mu)^2, \end{aligned} \quad (2.4.3)$$

em que $V_1 = \text{Var}(y|\mathbb{1}_{\{0,1\}}(y) = 1) = \gamma(1 - \gamma)$, $V_2 = \text{Var}(y|\mathbb{1}_{\{0,1\}}(y) = 0) = \mu(1 - \mu)/(\phi + 1)$ e μ_r é o r -ésimo momento ao redor de zero da distribuição $\mathcal{B}(\mu, \phi)$ dada em (2.2.1). Note que $E(y^r)$ é a média ponderada entre o r -ésimo momento da distribuição de Bernoulli ao redor de zero e o correspondente momento da distribuição $\mathcal{B}(\mu, \phi)$ com pesos α e $(1 - \alpha)$, respectivamente.

Observamos que $\text{Var}(y) \rightarrow V_2$, quando $\alpha \rightarrow 0$, e $\text{Var}(y) \rightarrow V_1$, quando $\alpha \rightarrow 1$. No caso em que as duas distribuições em (2.4.1) são misturadas na mesma proporção, i.e., $\alpha = 1/2$, temos que $\text{Var}(y) \rightarrow (V_1 + V_2)/2 + (\gamma - \mu)^2/4$. Quando γ, μ se aproximam de $1/2$ e $\phi \rightarrow 0$ temos que $\text{Var}(y) \rightarrow 1/4$. Se μ, γ são iguais a $1/2$ e $\phi = 2$ temos que $\text{Var}(y) = (2\alpha + 1)/12$, que corresponde à variância de uma mistura entre a distribuição uniforme no intervalo $(0, 1)$ e uma distribuição de Bernoulli de parâmetro $1/2$. Neste caso,

[‡]A medida de probabilidade P correspondente a $\text{BIZU}(y; \cdot)$ é definida sobre o espaço mensurável $([0, 1], \mathfrak{B})$ em que \mathfrak{B} é a classe de todos os subconjuntos de Borel de $[0, 1]$, e tal que $P \ll \lambda + m_0 + m_1$, sendo λ a medida de Lebesgue e m_c é um ponto de massa em c , i.e., $m_c(A) = 1$ se $c \in A$ e 0 se $c \notin A$, $A \in \mathfrak{B}$.

2.4. A distribuição beta inflacionada em zero e um

podemos deduzir que se $\alpha \rightarrow 0$, $\text{Var}(y) > 1/12$ para todo $0 < \phi < 2$ e $\text{Var}(y) < 1/12$ quando $\phi > 2$. Para α, μ e ϕ fixos, $P(y = 0) \rightarrow \alpha$ quando $\gamma \rightarrow 0$ e $P(y = 1) \rightarrow \alpha$ quando $\gamma \rightarrow 1$. Se $\gamma = 1/2$ temos que $P(y = 0) = P(y = 1) = \alpha/2$. Adicionalmente, à medida em que $\alpha \rightarrow 1$ a distribuição BIZU tende a concentrar sua massa de probabilidade nos extremos do intervalo $[0, 1]$, i.e., tende à distribuição de Bernoulli de parâmetro γ . Na situação em que $\alpha \rightarrow 0$ a distribuição BIZU tende à distribuição $\mathcal{B}(\mu, \phi)$.

A Figura (2.3) apresenta gráficos da densidade BIZU (2.4.2) para diferentes valores dos parâmetros α, γ, μ e ϕ . Quando $\mu = 0.5$ e $\gamma = 0.5$ a distribuição é simétrica para qualquer escolha de α e $\phi > 1$. Quando $\mu \neq 1/2$ ou $\gamma \neq 1/2$ a distribuição é assimétrica. Note que em $(0, 1)$, se $\mu < 1/2$ e $\phi \leq 2$, a distribuição BIZU tem forma de ‘J’ invertida; para $\mu > 1/2$ e $\phi \leq 2$ a distribuição BIZU tem forma de ‘J’ com massa de probabilidade nos extremos de $[0, 1]$ sendo representada pelas barras com o ponto acima.

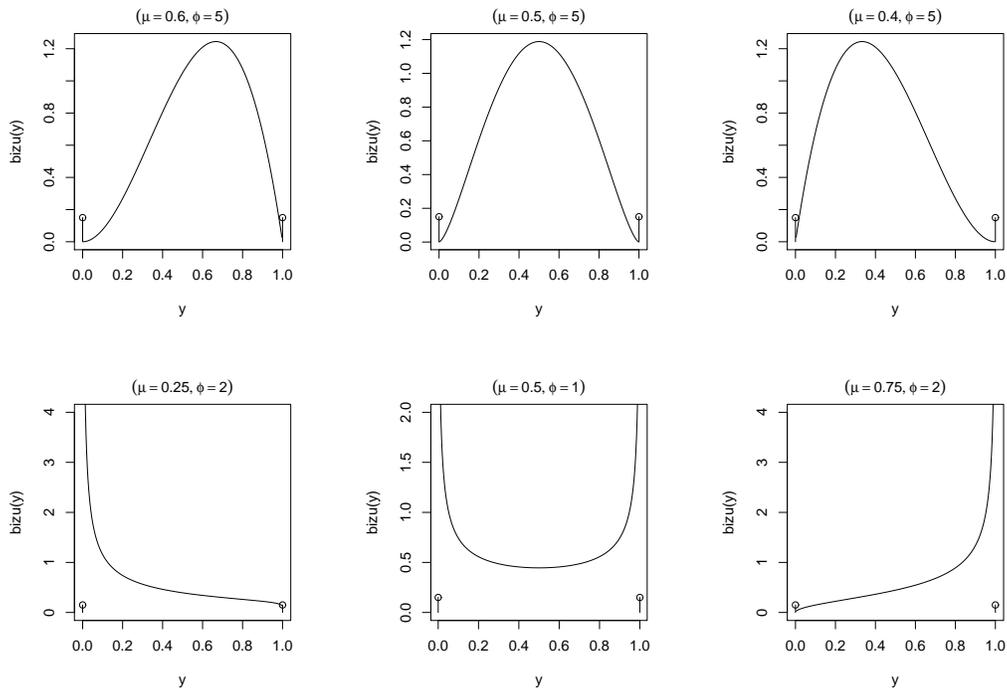


Figura 2.3: Distribuições BIZU para diferentes valores de μ e ϕ ; $\alpha = 0.3, \gamma = 0.5$.

De forma análoga à densidade (2.3.2) é possível mostrar que a densidade (2.4.2) pertence à família exponencial.

Proposição 2.4.1. *A distribuição BIZU dada em (2.4.2) pertence à família exponencial de dimensão 4 de posto completo.*

2.4. A distribuição beta inflacionada em zero e um

Prova : Seja $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)$ com $\eta_1 = [\log(\alpha/(1-\alpha)) - M(\eta_2) + b(\eta_3, \eta_4)]$, $\eta_2 = \log(\gamma/(1-\gamma))$, $\eta_3 = \mu\phi$ e $\eta_4 = (1-\mu)\phi$ em que $M(\eta_2) = \log(1 + e^{\eta_2})$ e $b(\eta_3, \eta_4) = \log(\Gamma(\eta_3)\Gamma(\eta_4)/\Gamma(\eta_3 + \eta_4))$ e seja $T(y) = (t_1(y), t_2(y), t_3(y), t_4(y))$ com

$$\begin{aligned} t_1(y) &= \mathbb{1}_{\{0,1\}}(y), \\ t_2(y) &= y\mathbb{1}_{\{0,1\}}(y) = \mathbb{1}_{\{1\}}(y), \\ t_3(y) &= \begin{cases} \log(y), & \text{se } y \in (0, 1), \\ 0, & \text{se } y \in \{0, 1\}, \end{cases} \\ t_4(y) &= \begin{cases} \log(1-y), & \text{se } y \in (0, 1), \\ 0, & \text{se } y \in \{0, 1\}. \end{cases} \end{aligned}$$

A função de densidade BIZU (2.4.2) pode ser escrita na forma

$$\exp\{\eta^\top T(y) - B^*(\eta)\}h(y), \quad (2.4.4)$$

com

$$B^*(\eta) = \log\{1 + \exp[\eta_1 + M(\eta_2) - b(\eta_3, \eta_4)]\} + b(\eta_3, \eta_4),$$

uma função de η a valores reais e $h(y) = 1/\{y(1-y)\}$ se $y \in (0, 1)$ e 1 se $y \in \{0, 1\}$. A parametrização η define uma transformação bijetora que aplica $\mathfrak{X} = (0, 1) \times (0, 1) \times (0, 1) \times \mathbb{R}^+$ a $\mathfrak{D} = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$, um subconjunto aberto de \mathbb{R}^4 . Adicionalmente, podemos mostrar que os t 's e os η 's são linearmente independentes. Desta forma, o espaço paramétrico contém retângulos de dimensão 4. Assim, (2.4.4) é a representação canônica da função de distribuição BIZU na família exponencial de dimensão 4 de posto completo. \blackbox

Suponhamos que y_1, \dots, y_n é uma amostra aleatória da distribuição BIZU i.e., y_t para $t = 1, \dots, n$, são independentes e identicamente distribuídos seguindo a distribuição BIZU. Como consequência da Proposição 2.4.1 o vetor de estatísticas $\sum_{t=1}^n T(y_t) = (T_1, T_2, T_3, T_4)$, em que

$$\begin{aligned} T_1 &= \sum_{t=1}^n \mathbb{1}_{\{0,1\}}(y_t), \\ T_2 &= \sum_{t=1}^n y_t \mathbb{1}_{\{0,1\}}(y_t) = \sum_{t=1}^n \mathbb{1}_{\{1\}}(y_t), \\ T_3 &= \sum_{t:y_t \in (0,1)} \log(y_t), \\ T_4 &= \sum_{t:y_t \in (0,1)} \log(1-y_t), \end{aligned} \quad (2.4.5)$$

é uma estatística suficiente completa (Lehmann & Casella, 2002, Corolário 1.6.16 e Teorema 1.6.22).

2.4. A distribuição beta inflacionada em zero e um

A função de verossimilhança para $\theta = (\alpha, \gamma, \mu, \phi)$ dada a amostra (y_1, \dots, y_n) é

$$L(\theta) = \prod_{t=1}^n \text{bizu}(y_t; \alpha, \gamma, \mu, \phi) = L_1(\alpha)L_2(\gamma)L_3(\mu, \phi)$$

com

$$\begin{aligned} L_1(\alpha) &= \prod_{t=1}^n \alpha^{\mathbb{1}_{\{0,1\}}(y_t)} (1 - \alpha)^{1 - \mathbb{1}_{\{0,1\}}(y_t)} = \alpha^{T_1} (1 - \alpha)^{(n - T_1)}, \\ L_2(\gamma) &= \prod_{t: y_t \in \{0,1\}} \gamma^{y_t} (1 - \gamma)^{(1 - y_t)} = \gamma^{T_2} (1 - \gamma)^{(T_1 - T_2)}, \\ L_3(\mu, \phi) &= \prod_{t: y_t \in (0,1)} f(y_t; \mu, \phi). \end{aligned}$$

A função de verossimilhança $L(\theta)$ fatora em três termos, a saber, L_1 , L_2 e L_3 ; L_1 depende unicamente de α , L_2 depende somente de γ e L_3 depende unicamente de (μ, ϕ) . Assim, α , γ e (μ, ϕ) são parâmetros separáveis (Pace & Salvani, 1997, p. 128) e a inferência por máxima verossimilhança para α , γ e (μ, ϕ) pode ser realizada de forma independente.

O logaritmo da função de verossimilhança assume a forma

$$\ell(\theta) = \log(L(\theta)) = \ell_1(\alpha) + \ell_2(\gamma) + \ell_3(\mu, \phi),$$

sendo

$$\begin{aligned} \ell_1(\alpha) &= T_1 \log \alpha + (n - T_1) \log(1 - \alpha), \\ \ell_2(\gamma) &= T_2 \log \gamma + (T_1 - T_2) \log(1 - \gamma), \\ \ell_3(\mu, \phi) &= (n - T_1) \log \left\{ \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} \right\} + T_3(\mu\phi - 1) + T_4((1 - \mu)\phi - 1). \end{aligned}$$

Pela diferenciação de $\ell_1(\alpha)$ com respeito a α , $\ell_2(\gamma)$ com respeito a γ e $\ell_3(\mu, \phi)$ com respeito a μ e ϕ , nós obtemos o vetor escore $(U_\alpha(\alpha), U_\gamma(\gamma), U_\mu(\mu, \phi), U_\phi(\mu, \phi))$, em que

$$\begin{aligned} U_\alpha(\alpha) &= T_1/\alpha - (n - T_1)/(1 - \alpha), \\ U_\gamma(\gamma) &= T_2/\gamma - (T_1 - T_2)/(1 - \gamma), \\ U_\mu(\mu, \phi) &= \phi\{(n - T_1)[\psi((1 - \mu)\phi) - \psi(\mu\phi)] + T_3 - T_4\}, \\ U_\phi(\mu, \phi) &= (n - T_1)[\psi(\phi) - \mu\psi(\mu\phi) - (1 - \mu)\psi((1 - \mu)\phi)] + \mu T_3 - (1 - \mu)T_4. \end{aligned}$$

Facilmente podemos mostrar que $\hat{\alpha} = T_1/n$ e $\hat{\gamma} = T_2/T_1$ (0/0 definido como 0) são estimadores de máxima verossimilhança de α e γ . Aqui, $\hat{\alpha}$ é proporção de zeros e uns na amostra, $\hat{\gamma}$ é a proporção de uns na sub-amostra de observações que são zeros ou uns. Adicionalmente $\hat{\alpha}$ é função de uma estatística suficiente completa e é um estimador não-viesado de α , isto é, $E(\hat{\alpha}) = \alpha$. Logo, $\hat{\alpha}$ é um estimador não-viesado com

2.4. A distribuição beta inflacionada em zero e um

variância uniformemente mínima para α ; sua variância é dada por $\text{Var}(\hat{\alpha}) = \alpha(1 - \alpha)/n$. Note que o sistema de equações $U_\mu(\mu, \phi) = 0$ e $U_\phi(\mu, \phi) = 0$ é não-linear e neste caso os estimadores de μ e ϕ não podem ser obtidos de forma fechada. Eles devem ser obtidos numericamente pela maximização da função de log-verossimilhança $\ell_2(\mu, \phi)$ usando um algoritmo de otimização não-linear (Newton-Raphson, escore de Fisher, BHHH, etc.). Pela invariância dos estimadores de máxima verossimilhança, temos que os estimadores de máxima verossimilhança das probabilidades de ocorrência de zero e de um são dados por

$$\begin{aligned} P(\widehat{y = 0}) &= \hat{\alpha}(1 - \hat{\gamma}) = \frac{T_1 - T_2}{n}, \\ P(\widehat{y = 1}) &= \hat{\alpha}\hat{\gamma} = \frac{T_2}{n}, \end{aligned}$$

sendo que T_2/n representa a proporção de uns na amostra e $(T_1 - T_2)/n$ é a proporção de zeros na amostra. É fácil mostrar que esses estimadores são não viesados, i.e., $E((T_1 - T_2)/n) = P(y = 0)$ e $E(T_2/n) = P(y = 1)$. Adicionalmente, cada um deles é função de estatísticas suficientes e completas. Logo, $P(\widehat{y = 0})$ e $P(\widehat{y = 1})$ são estimadores não-viesados de variância uniformemente mínima de $P(y = 0)$ e $P(y = 1)$, respectivamente. No caso em que $0 < T_1 < n$,[§] a solução do sistema de equações $(\bar{y}, s^2)^\top = (\mu, \mu(1 - \mu)/(\phi + 1))^\top$, com $\bar{y} = \sum_{t: y_t \in (0,1)} y_t / (n - T_1)$ e $s^2 = \sum_{t: y_t \in (0,1)} (y_t - \bar{y})^2 / (n - T_1)$, fornece os seguintes estimadores de forma fechada para μ e ϕ : $\tilde{\mu} = \bar{y}$, $\tilde{\phi} = \{\tilde{\mu}(1 - \tilde{\mu})/s^2\} - 1$. De forma análoga à distribuição beta inflacionada em zero ou um, $\tilde{\mu}$ e $\tilde{\phi}$ são obtidos através dos momentos condicionais de y dado que $y \in (0, 1)$. Estimadores de forma fechada para $E(y^r)$ e $\text{Var}(y)$ podem ser obtidos substituindo α, γ, μ e ϕ por $\hat{\alpha}, \hat{\gamma}, \tilde{\mu}$ e $\tilde{\phi}$ na equação (2.4.3).

A matriz de informação de Fisher $K(\theta)$ pode ser escrita como

$$K(\theta) = \begin{pmatrix} \kappa_{\alpha\alpha} & 0 & 0 & 0 \\ 0 & \kappa_{\gamma\gamma} & 0 & 0 \\ 0 & 0 & \kappa_{\mu\mu} & \kappa_{\mu\phi} \\ 0 & 0 & \kappa_{\phi\mu} & \kappa_{\phi\phi} \end{pmatrix}, \quad (2.4.6)$$

onde

$$\begin{aligned} \kappa_{\alpha\alpha} &= 1/\{\alpha(1 - \alpha)\}, \\ \kappa_{\gamma\gamma} &= \alpha/\{\gamma(1 - \gamma)\}, \\ \kappa_{\mu\mu} &= (1 - \alpha)\phi^2\{\psi'(\mu\phi) + \psi'((1 - \mu)\phi)\}, \\ \kappa_{\mu\phi} &= \kappa_{\phi\mu} = (1 - \alpha)\phi\{\psi'(\mu\phi)\mu - \psi'((1 - \mu)\phi)(1 - \mu)\} \\ \kappa_{\phi\phi} &= (1 - \alpha)\{\mu^2\psi'(\mu\phi) + (1 - \mu)^2\psi'((1 - \mu)\phi) - \psi'(\phi)\}. \end{aligned}$$

Assim, α, γ e (μ, ϕ) são parâmetros ortogonais. Esta ortogonalidade implica que os respectivos componentes

[§]Se $T_1 = 0$ (todas as observações estão em $(0, 1)$) ou $T_1 = n$ (todas as observações são 0 ou 1) a distribuição BIZU não é recomendada.

2.4. A distribuição beta inflacionada em zero e um

do vetor escore são não-correlacionados e $\hat{\alpha}$, $\hat{\gamma}$, e $(\hat{\mu}, \hat{\phi})$ são assintoticamente independentes. Adicionalmente, da Proposição 2.3.1 temos que (Bickel & Doksum, 2001, Teorema 5.3.5, p. 322)

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}_4(0, K(\theta)^{-1}),$$

em que $\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\mu}, \hat{\phi})$ é o estimador de máxima verossimilhança de $\theta = (\alpha, \gamma, \mu, \phi)$, \mathcal{N}_4 representa a distribuição normal multivariada de dimensão 4 e $K(\theta)$ é a matriz (2.4.6). Pela normalidade assintótica do estimador de máxima verossimilhança $\hat{\theta}$, podemos construir intervalos de confiança de tipo assintótico para os parâmetros da distribuição BIZU. Por exemplo, temos que

$$\hat{\mu} \pm z_{1-\frac{\alpha}{2}}(\hat{\kappa}^{\mu\mu})^{1/2}$$

são os limites do intervalo de confiança assintótico (ICA) para μ com coeficiente de confiança $100(1 - \alpha)\%$, sendo $z_{1-\alpha/2}$ o quantil $(1 - \alpha/2)$ da distribuição $\mathcal{N}(0, 1)$ e $\hat{\kappa}^{\mu\mu}$ é a variância assintótica de $\hat{\mu}$ obtida do inverso da matriz de informação $K(\theta)$ dada em (2.4.6) e sendo avaliada no estimador de máxima verossimilhança. O método delta (ver Seção 2.3) é útil para obter a distribuição assintótica do estimador de máxima verossimilhança de alguma função $r(\theta)$ diferenciável em θ . Por exemplo, se $r(\theta) = E(y) = \alpha\gamma + (1 - \alpha)\mu$, a variância assintótica da distribuição limite de $\widehat{E}(y) = r(\hat{\theta})$ é $\alpha^2\kappa^{\gamma\gamma} + (1 - \alpha)^2\kappa^{\mu\mu} + (\gamma - \mu)^2\kappa^{\alpha\alpha}$. Aqui, $\kappa^{\alpha\alpha} = 1/\kappa_{\alpha\alpha}$, $\kappa^{\gamma\gamma} = 1/\kappa_{\gamma\gamma}$ e $\kappa^{\mu\mu}$ é o (3, 3)-elemento do inverso da matriz $K(\theta)$ dada em (2.4.6).

Para a distribuição BIZU podem ser consideradas outras parametrizações. Por exemplo, podemos formular uma parametrização na qual o parâmetro da distribuição de Bernoulli satisfaz a relação $\gamma = \delta_1/\alpha$ e o parâmetro de mistura $\alpha = \delta_0 + \delta_1$. Desta forma, a densidade BIZU pode ser escrita como

$$\text{bizu}(y; \delta_0, \delta_1, \mu, \phi) = \begin{cases} \delta_0, & \text{se } y = 0, \\ \delta_1, & \text{se } y = 1, \\ (1 - \delta_0 - \delta_1)f(y; \mu, \phi), & \text{se } y \in (0, 1), \end{cases} \quad (2.4.7)$$

sendo $f(y; \mu, \phi)$ a densidade da distribuição beta (2.2.1). Nesta parametrização a interpretação dos parâmetros é mais intuitiva, uma vez que $\delta_0 = P(y = 0)$, $\delta_1 = P(y = 1)$ e μ, ϕ são os parâmetros da distribuição beta (2.2.1). Claramente, sob esta parametrização, a distribuição BIZU também tem representação canônica na família exponencial de dimensão 4 de posto completo. Contudo, esta parametrização induz a restrição no espaço paramétrico dada por $0 < \delta_0 + \delta_1 < 1$. A estimação dos parâmetros pode ser alcançada por máxima verossimilhança penalizada.

A função de log-verossimilhança da distribuição BIZU pode ser maximizada iterativamente usando os algoritmos RS ou CG (Rigby & Stasinopoulos, 2005). Estes algoritmos usam o procedimento de retroajuste (backfitting) para melhorar o desempenho do procedimento de escore de Fisher. O algoritmo CG (Cole & Green, 1992) usa a função de log-verossimilhança dos dados e as primeiras derivadas (opcionalmente podem utilizar o valor esperado das derivadas de segunda ordem) com respeito aos parâmetros da distribuição.

2.5. Resultados numéricos e discussão

Finalmente, se consideramos o vetor de parâmetros conjunto $\theta = (\delta_0, \delta_1, \mu, \phi)$ podemos mostrar que a matriz de informação de Fisher da distribuição BIZU assume a forma

$$K(\theta) = \begin{pmatrix} \kappa_{\delta_0\delta_0} & \kappa_{\delta_0\delta_1} & 0 & 0 \\ \kappa_{\delta_1\delta_0} & \kappa_{\delta_1\delta_1} & 0 & 0 \\ 0 & 0 & \kappa_{\mu\mu} & \kappa_{\mu\phi} \\ 0 & 0 & \kappa_{\phi\mu} & \kappa_{\phi\phi} \end{pmatrix},$$

em que

$$\begin{aligned} \kappa_{\delta_0\delta_0} &= (1 - \delta_1)/\delta_0(1 - \delta_0 - \delta_1), \\ \kappa_{\delta_0\delta_1} &= \kappa_{\delta_1\delta_0} = 1/(1 - \delta_0 - \delta_1), \\ \kappa_{\delta_1\delta_1} &= (1 - \delta_0)/\delta_1(1 - \delta_0 - \delta_1), \\ \kappa_{\mu\mu} &= (1 - \alpha)\phi^2\{\psi'(\mu\phi) + \psi'((1 - \mu)\phi)\}, \\ \kappa_{\mu\phi} &= \kappa_{\phi\mu} = (1 - \alpha)\phi\{\psi'(\mu\phi)\mu - \psi'((1 - \mu)\phi)(1 - \mu)\}, \\ \kappa_{\phi\phi} &= (1 - \alpha)\{\mu^2\psi'(\mu\phi) + (1 - \mu)^2\psi'((1 - \mu)\phi) - \psi'(\phi)\}. \end{aligned}$$

Note que $\kappa_{\delta_0\delta_1} \neq 0$, o que implica que os parâmetros δ_0 e δ_1 não são globalmente ortogonais e desta forma seus respectivos componentes no vetor escore são correlacionados em contraste à parametrização da distribuição BIZU dada em (2.4.2). Adicionalmente, percebemos que (δ_0, δ_1) é ortogonal a (μ, ϕ) e assintoticamente temos que $(\hat{\delta}_0, \hat{\delta}_1)$ é independente de $(\hat{\mu}, \hat{\phi})$.

2.5 Resultados numéricos e discussão

Através de simulações de Monte Carlo, avaliamos os desempenhos dos estimadores de máxima verossimilhança (MV) e de momentos condicionais (MC) para os parâmetros das distribuições $BIZ(\alpha, \mu, \phi)$ e $BIZU(\alpha, \gamma, \mu, \phi)$ em amostras de tamanho finito e sob diferentes critérios. Nosso estudo se concentra na estimação de μ , ϕ , $E(y)$ e $\text{Var}(y)$. Para a distribuição BIZU também consideramos a estimação por MV de γ . Como $\hat{\alpha}$ e $\tilde{\alpha}$ coincidem e são estimadores não-viesados de variância uniformemente mínima para α nas distribuições BIZ e BIZU, não mostramos resultados de simulação referentes à estimação deste parâmetro. Os parâmetros da distribuição BIZ foram fixados em $\alpha = 0.2$, $\mu = 0.1$ e $\phi = 2$. No caso da distribuição BIZU, tomamos $\alpha = 0.2$, $\gamma = 0.3$, $\mu = 0.1$ e $\phi = 2$. Para o experimento de Monte Carlo foram consideradas $R = 5000$ réplicas. Os tamanhos de amostra são $n = 10, 20, 50, 100, 500$ e 1000 . As estimativas dos parâmetros obtidas pelos momentos condicionais foram utilizadas como estimativas iniciais para o processo de maximização da função de log-verossimilhança de cada uma das distribuições. A função de log-verossimilhança de cada uma das distribuições é maximizada através do método BFGS com derivadas analíticas (Press et al., 1992). Para a análise de resultados da estimação pontual, foram calculados para cada distribuição e para cada tamanho de amostra: a média $\hat{E}(\hat{\theta}) = R^{-1} \sum_{i=1}^R \hat{\theta}_i$, a estimativa do viés $\hat{B}(\hat{\theta}) = \hat{E}(\hat{\theta}) - \theta$, a estimativa do viés relativo $\hat{B}_R(\hat{\theta}) = \theta^{-1}(\hat{E}(\hat{\theta}) - \theta)$ e a estimativa da raiz quadrada do erro quadrático médio $\sqrt{\text{EQM}} = \sqrt{R^{-1} \sum_{i=1}^R (\hat{\theta}_i - \theta)^2}$ das 5000 estimativas, em que θ é o valor verdadeiro do parâmetro e $\hat{\theta}$ seu respectivo estimador.

2.6. Aplicações

Todo o procedimento de cálculo foi programado na linguagem de programação Ox (Cribari–Neto & Zarkos, 2003; Doornik, 2001).

Na Tabela 2.1 apresentamos os resultados de simulação para a distribuição BIZ. Observamos que, em módulo, as estimativas do viés e do viés relativo estimado de μ por MV e MC estão próximas de zero, sendo que o viés diminui com o aumento do tamanho de amostra. As estimativas de \sqrt{EQM} dos estimadores obtidos por MV e MC para μ são similares. Para todos os tamanhos de amostra, os estimadores de ϕ por MV e MC são consideravelmente viesados, sendo que o viés do estimador obtido por MC é mais pronunciado. Note que as estimativas MC da média e \sqrt{EQM} de ϕ podem ser muito maiores que as de MV. Por exemplo, para $n = 10$ e $\phi = 2$, o viés e \sqrt{EQM} são respectivamente, 3.4 e 10.5 para o estimador MV e 6.5 e 21.2 para o estimador de MC. Os estimadores de MV e MC para ϕ se apresentam com viés positivo e assim a variância da variável resposta é subestimada.

Na Tabela 2.2 se encontram os resultados obtidos da simulação para a distribuição BIZU. Podemos observar que, em módulo, os estimadores de MV e MC de γ e μ são ligeiramente viesados, sendo que este viés diminui com o aumento do tamanho de amostra. Novamente, os estimadores de MV e MC de ϕ são bastante viesados e o estimador de MV se apresenta mais eficiente que o estimador de MC. Por exemplo, para $n = 20$ e $\phi = 2$, o viés e \sqrt{EQM} são, respectivamente, 1.0 e 2.4 para o estimador de MV e 1.7 e 3.9 para o estimador de MC.

Em resumo, nas distribuições BIZ e BIZU o estimador de MV de ϕ é mais eficiente que o estimador de MC. No entanto, ambos os estimadores se apresentam bastante viesados para tamanhos de amostra pequenos (por exemplo, $n = 10$). Por outro lado, os estimadores de MV e de MC para os outros parâmetros têm desempenhos iguais ou similares e podem ser aproximadamente não-viesados se os tamanhos de amostra não são muito pequenos. Os resultados de simulação para a distribuição BIU não são apresentados pois estes podem ser obtidos da distribuição BIZ, lembrando que a única diferença entre as distribuições BIZ e BIU é a alocação do ponto de massa, sendo em zero para a distribuição BIZ e em um para a distribuição BIU, i.e., $BIU(\alpha, \mu, \phi) = 1 - BIZ(\alpha, 1 - \mu, \phi)$. Finalmente, os estimadores de MV e MC de $E(y)$ e $Var(y)$ têm desempenhos semelhantes em termos das medidas consideradas, sendo ambos quase não viesados se a amostra for suficientemente grande.

2.6 Aplicações

O objetivo desta seção é ilustrar o uso das distribuições beta inflacionadas BIZ e BIZU para dados restritos aos intervalos $[0, 1)$, e $[0, 1]$. Para nossa análise, apresentamos três conjuntos de dados com seus respectivos modelos ajustados. Para efeito de comparação, também avaliamos um modelo Tobit para cada conjunto de dados. Em cada caso apresentamos o histograma de freqüências e o gráfico da distribuição empírica sobrepondo as respectivas distribuições acumuladas ajustadas, beta inflacionadas e do modelo Tobit. Os parâmetros de cada distribuição beta inflacionada e do modelo Tobit são estimados pelo método de máxima verossimilhança. Aqui, usamos as implementações na linguagem de programação R (Ihaka & Gentleman, 1996) das distribuições BIZ e BIU do pacote `gamlss` desenvolvidas por Ospina (2006), da distribuição BIZU por Stasinopoulos, Rigby & Akantziliotou (2006) e do modelo Tobit do pacote `VGAM` (Yee & Wild, 1996).

2.6. Aplicações

Tabela 2.1: Resultados de simulação para a distribuição BIZ; $\alpha = 0.2$, $\mu = 0.1$, $\phi = 2.0$, $E(y) = 0.08$ e $\text{Var}(y) = 0.0256$.

Par.	n	Média		Viés		$\sqrt{\text{EQM}}$	
		MC	MV	MC	MV	MC	MV
μ	10	0.1009	0.0945	0.0009	-0.0055	0.0637	0.0591
	20	0.1005	0.0971	0.0005	-0.0029	0.0432	0.0406
	50	0.1004	0.0989	0.0004	-0.0011	0.0276	0.0262
	100	0.0999	0.0992	-0.0001	-0.0008	0.0194	0.0186
	500	0.0998	0.0996	-0.0002	-0.0004	0.0086	0.0083
	1000	0.0997	0.0997	-0.0003	-0.0003	0.0061	0.0058
ϕ	10	8.5145	5.3980	6.5145	3.3980	21.1880	10.4990
	20	3.6960	2.9660	1.6960	0.9660	4.6428	2.8064
	50	2.4009	2.2750	0.4009	0.2750	1.1773	0.8370
	100	2.1840	2.1371	0.1840	0.1371	0.6793	0.5036
	500	2.0340	2.0292	0.0340	0.0292	0.2487	0.1959
	1000	2.0150	2.0133	0.0150	0.0133	0.1700	0.1342
$E(y)$	10	0.0785	0.0739	-0.0015	-0.0061	0.0503	0.0471
	20	0.0802	0.0775	0.0002	-0.0025	0.0354	0.0334
	50	0.0804	0.0792	0.0004	-0.0008	0.0228	0.0218
	100	0.0799	0.0794	-0.0001	-0.0006	0.0161	0.0155
	500	0.0798	0.0797	-0.0002	-0.0003	0.0072	0.0069
	1000	0.0800	0.0799	0.0000	-0.0001	0.0050	0.0048
$\text{Var}(y)$	10	0.0229	0.0228	-0.0027	-0.0028	0.0229	0.0203
	20	0.0244	0.0242	-0.0012	-0.0014	0.0167	0.0153
	50	0.0253	0.0252	-0.0003	-0.0004	0.0109	0.0104
	100	0.0254	0.0253	-0.0002	-0.0003	0.0078	0.0075
	500	0.0255	0.0255	-0.0001	-0.0001	0.0035	0.0034
	1000	0.0256	0.0255	-0.0000	-0.0001	0.0024	0.0024

O primeiro conjunto de dados, com 645 observações, foi obtido através dos indicadores brasileiros de atendimento qualificado para serviços prioritários no ano 2000. A variável de interesse (y) é a porcentagem de enfermeiros com curso superior nos municípios brasileiros. Os dados foram extraídos do banco de dados das Nações Unidas através do *Atlas de Desenvolvimento Humano no Brasil* disponível no sítio <http://www.pnud.org.br/atlas/>.

2.6. Aplicações

Tabela 2.2: Resultados da simulação para a distribuição BIZU; $\alpha = 0.2, \gamma = 0.3, \mu = 0.1, \phi = 2.0, E(y) = 0.14$ e $\text{Var}(y) = 0.0724$.

Par.	n	Média		Viés		$\sqrt{\text{EQM}}$	
		MC	MV	MC	MV	MC	MV
γ	10		0.4462		0.1462		0.1348
	20		0.3882		0.0882		0.1616
	50		0.3102		0.0102		0.1374
	100		0.3015		0.0015		0.1047
	500		0.2995		-0.0005		0.0466
	1000		0.3011		0.0011		0.0323
μ	10	0.1017	0.0955	0.0017	-0.0045	0.0661	0.0607
	20	0.1003	0.0969	0.0003	-0.0031	0.0445	0.0418
	50	0.1006	0.0992	0.0006	-0.0008	0.0278	0.0265
	100	0.0999	0.0993	-0.0001	-0.0007	0.0195	0.0186
	500	0.1002	0.1000	0.0002	0.0000	0.0086	0.0083
	1000	0.1001	0.0999	0.0001	-0.0001	0.0062	0.0060
ϕ	10	10.9191	6.6869	8.9191	4.6869	33.3939	19.5225
	20	3.7055	2.9909	1.7055	0.9909	3.8645	2.3889
	50	2.4123	2.2821	0.4123	0.2821	1.1997	0.8229
	100	2.1824	2.1380	0.1824	0.1380	0.6407	0.4891
	500	2.0321	2.0210	0.0321	0.0210	0.2424	0.1894
	1000	2.0187	2.0115	0.0187	0.0115	0.1726	0.1363
$E(y)$	10	0.2002	0.1956	0.0602	0.0556	0.0679	0.0655
	20	0.1627	0.1601	0.0227	0.0201	0.0524	0.0512
	50	0.1423	0.1411	0.0023	0.0011	0.0363	0.0357
	100	0.1401	0.1396	0.0001	-0.0004	0.0266	0.0263
	500	0.1400	0.1398	0.0000	-0.0002	0.0121	0.0119
	1000	0.1402	0.1401	0.0002	0.0001	0.0085	0.0084
$\text{Var}(y)$	10	0.1129	0.1139	0.0405	0.0415	0.0313	0.0306
	20	0.0871	0.0873	0.0147	0.0149	0.0297	0.0293
	50	0.0727	0.0727	0.0003	0.0003	0.0232	0.0229
	100	0.0717	0.0717	-0.0007	-0.0006	0.0177	0.0176
	500	0.0721	0.0721	-0.0003	-0.0003	0.0080	0.0080
	1000	0.0724	0.0724	0.0000	0.0000	0.0057	0.0056

2.6. Aplicações

Na Figura 2.4, o histograma de freqüências dos dados em $(0, 1)$ apresenta forma de ‘J’, uma característica que é modelada facilmente pela distribuição BIZ. Neste gráfico a barra com o ponto acima representa a quantidade de zeros na amostra. Aqui, consideramos também um modelo Tobit censurado à esquerda (TCE), ao assumir que $y_t = y_t^*$ se $y_t^* > 0$, e $y_t = 0$ se $y_t^* \leq 0$, onde os $y_t^* \sim \mathcal{N}(\mu, \sigma^2)$ são variáveis aleatórias independentes. As estimativas de MV dos parâmetros (erros padrão entre parênteses) da distribuição BIZ são $\hat{\alpha} = 0.0155$ (0.0049), $\hat{\mu} = 0.1263$ (0.0042) e $\hat{\phi} = 4.691$ (0.220), e para o modelo Tobit, $\hat{\mu} = 0.1177$ (0.0060) e $\hat{\sigma} = 0.1433$ (0.0040). O gráfico da distribuição acumulada empírica junto das distribuições acumuladas estimadas (ver Figura 2.4) indica que a distribuição BIZ é a que melhor se ajusta aos dados.

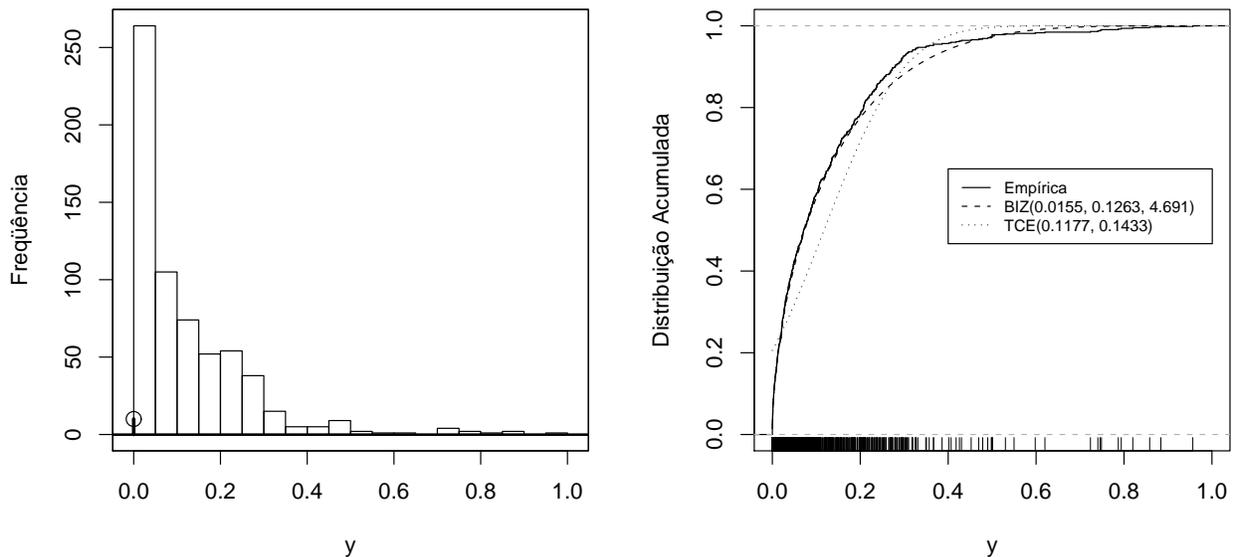


Figura 2.4: Histograma de freqüências e distribuição acumulada para a porcentagem de enfermeiros com curso superior em municípios brasileiros.

Em seguida, consideramos um conjunto de dados formado por 5561 observações da proporção de óbitos em menores de 1 ano por causas mal definidas nos municípios brasileiros do ano 2000. Os dados foram obtidos do Banco de Dados do Sistema Único de Saúde DATASUS, disponível de forma gratuita no sítio <http://www.datasus.gov.br>. Este conjunto de dados contém 3364 zeros e 172 uns. A Figura 2.5 apresenta o histograma de freqüências da proporção de óbitos em menores de 1 ano. Neste gráfico as barras com os pontos acima representam as quantidades de zeros e uns na amostra. Para este conjunto de dados ajustamos uma distribuição BIZU sob a parametrização $(\delta_0, \delta_1, \mu, \phi)$.

2.6. Aplicações

Adicionalmente, ajustamos um modelo Tobit duplamente censurado (TDC), i.e., assumimos que os $y_t = y_t^*$ se $0 < y_t^* < 1$, $y_t = 0$ se $y_t^* \leq 0$ e $y_t = 1$ se $y_t^* \geq 1$, em que $y_t^* \sim \mathcal{N}(\mu, \sigma^2)$ são variáveis aleatórias independentes. As estimativas de MV dos parâmetros são $\hat{\delta}_0 = 0.6055$ (0.0066), $\hat{\delta}_1 = 0.0313$ (0.0023), $\hat{\mu} = 0.2974$ (0.0043) e $\hat{\phi} = 0.4562$ (0.0050) para a distribuição BIZU e $\hat{\mu} = -0.1555$ (0.0088) e $\hat{\sigma} = 0.5420$ (0.0085) para o modelo Tobit. O gráfico das curvas da distribuição acumulada empírica e das distribuições acumuladas ajustadas é apresentado na Figura 2.5. Uma simples inspeção visual mostra claramente que apenas a distribuição BIZU é um modelo teórico adequado para o conjunto de dados

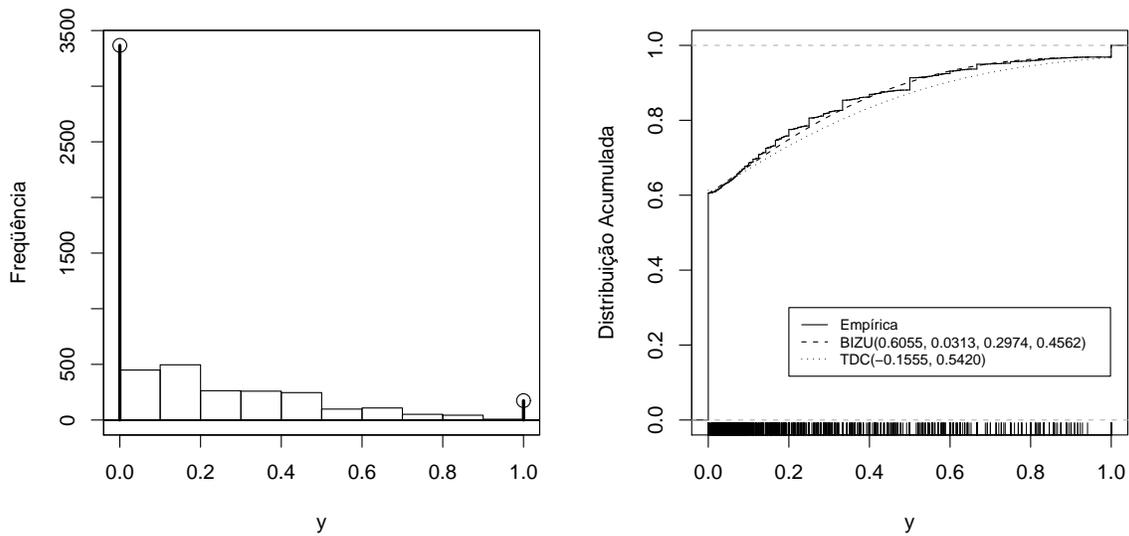


Figura 2.5: Histograma de frequências e distribuição acumulada para a proporção de óbitos em menores de 1 ano por causas mal definidas em municípios brasileiros.

Finalmente, o terceiro conjunto de dados corresponde a 223 observações da proporção de habitantes que moram até a 200 km do litoral em países litorâneos no ano 2002. Os dados são fornecidos pelo *Center for International Earth Science Information Network* e estão disponíveis na internet no sítio <http://sedac.ciesin.columbia.edu/plue/nagd/place>. Na Figura 2.6, percebe-se que o histograma de frequências dos dados apresenta forma de ‘U’; esta propriedade é facilmente modelada por uma distribuição BIZU. Para estes dados ajustamos uma distribuição BIZU sob a parametrização $(\delta_0, \delta_1, \mu, \phi)$ e um modelo Tobit duplamente censurado. As estimativas de MV dos parâmetros são $\hat{\delta}_0 = 0.1141$ (0.0215), $\hat{\delta}_1 = 0.4064$ (0.0332), $\hat{\mu} = 0.6189$ (0.0279) e $\hat{\phi} = 0.6615$ (0.0204). Para o modelo Tobit as estimativas são $\hat{\mu} = 0.8766$ (0.0518) e $\hat{\sigma} = 0.6975$ (0.0368). A Figura 2.6 mostra as curvas da distribuição acumulada em-

2.7. Conclusões

pírica e das distribuições acumuladas ajustadas. Claramente, o modelo de Tobit não se ajusta bem aos dados. Por outro lado, o gráfico das curvas da distribuição acumulada empírica e da distribuição acumulada estimada da distribuição BIZU mostra que o modelo teórico BIZU se ajusta muito bem aos dados, uma vez que não há grandes afastamentos entre as duas curvas. Desta forma, podemos concluir que a distribuição BIZU é um modelo adequado para os dados.

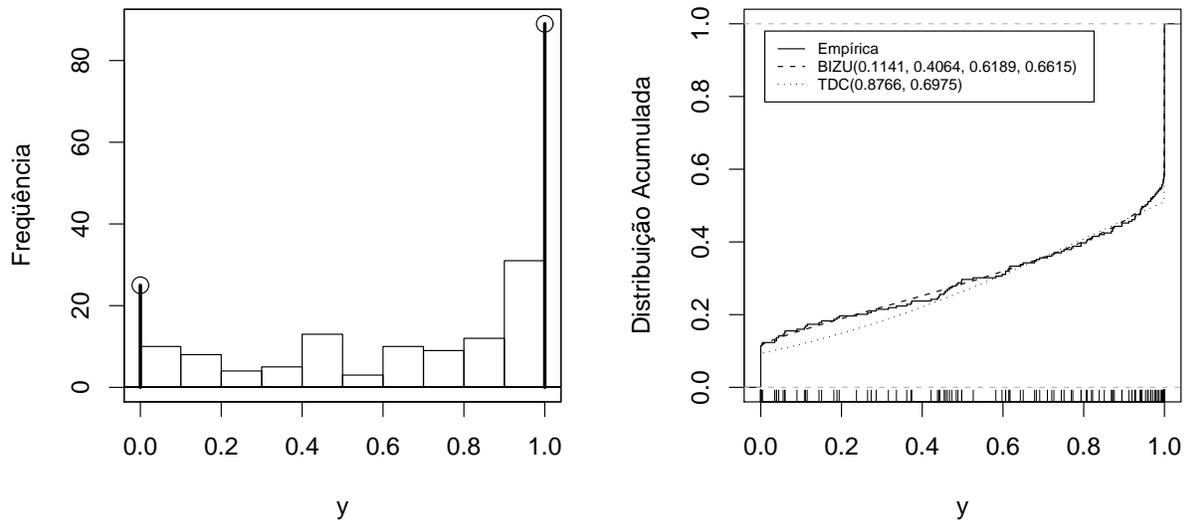


Figura 2.6: Histograma de freqüências e distribuição acumulada para a proporção de habitantes que moram até a 200 km do litoral em países litorâneos.

2.7 Conclusões

A distribuição beta é usada para modelar dados que são medidos de forma contínua no intervalo aberto $(0,1)$. Porém, conjuntos de dados que contêm zeros e/ou uns não podem ser modelados usando a distribuição beta. Neste capítulo, propusemos distribuições de mistura entre uma distribuição absolutamente contínua definida em $(0,1)$ e distribuições discretas para modelar dados que são observados nos intervalos $[0, 1)$, $(0, 1]$ ou $[0, 1]$. As distribuições propostas são distribuições beta inflacionadas no sentido que a massa de probabilidade em zero e/ou um excede o que é permitido pela distribuição beta. Propriedades das distribuições beta inflacionadas são analisadas.

Também discutimos para estes modelos a estimação pelos métodos de máxima verossimilhança e momentos condicionais e comparamos tais estimadores através de simulação de Monte Carlo.

2.7. Conclusões

Três aplicações empíricas com conjuntos de dados reais mostram que as distribuições beta inflacionadas são modelos bastante flexíveis para modelar dados de frações observados no intervalo unitário fechado ou semi-aberto. Nós usamos conjuntos de dados com formas bem diferentes de histograma de frequências (com formas unimodal, ‘ U ’ e ‘ J ’ invertida no intervalo $(0,1)$) e mostramos que as distribuições beta inflacionadas são modelos teóricos adequados para estes conjuntos de dados.

Sugerimos aos usuários interessados em modelar dados na forma de frações, proporções ou taxas medidas de forma contínua, que utilizem a distribuição beta inflacionada sempre que zeros e/ou uns apareçam no conjunto de dados.

Capítulo 3

Modelos de regressão beta inflacionados

3.1 Introdução

Na modelagem estatística uma estratégia tipicamente adotada para situações em que uma variável dependente (resposta) é medida de forma contínua no intervalo $(0, 1)$ é o uso de transformações de tal forma que a nova variável transformada assuma valores na reta. Isto permite modelar a média da nova variável através de um preditor linear baseado em um conjunto de variáveis explicativas conhecidas e parâmetros desconhecidos (Atkinson, 1985, Cap. 7). No entanto, esta metodologia possui algumas limitações, uma das quais é a perda de interpretabilidade dos parâmetros da regressão em termos da variável resposta original. Numa outra linha, Paolino (2001), Kieschnick & McCullough (2003), Ferrari & Cribari-Neto (2004) e Smithson & Verkulien (2006) introduziram modelos de regressão para variáveis aleatórias que são regidas por uma distribuição de probabilidades beta, em que a média da variável resposta é relacionada a um preditor linear (definido por regressores e parâmetros de regressão desconhecidos) por meio de uma função de ligação.

Há diversos artigos referentes à análise de dados que são observados de forma contínua no intervalo $(0,1)$. Dentre eles podemos citar Cox (1996) e Papke & Wooldridge (1996) em modelagem de quasi-verossimilhança, Song & Tan (2000) e Song, Qiu & Tan (2004) em equações de estimação generalizadas e Paolino (2001), Kieschnick & McCullough (2003), Ferrari & Cribari-Neto (2004), Vasconcellos & Cribari-Neto (2005) e Smithson & Verkuilen (2006) em modelagem de regressão paramétrica com distribuição beta para a variável resposta (modelos de regressão beta). Nestes trabalhos, assume-se que a média condicional da variável resposta (proporção, taxa ou fração), que é medida de forma contínua no intervalo $(0,1)$, pode ser modelada através de uma função não-linear envolvendo variáveis explicativas, i.e., a média da resposta é relacionada a um preditor linear (definido por regressores e parâmetros de regressão desconhecidos) por meio de uma função de ligação. Aqui, a média condicional é limitada ao intervalo $(0,1)$ e se admite que a variância condicional é uma função (em geral quadrática) da média condicional. Desta forma, a variância condicional muda à medida em que a média condicional varia no intervalo $(0,1)$.

Dentre as diferentes especificações dos modelos de regressão beta, a proposta de Ferrari & Cribari-Neto (2004) é a que se apresenta mais estruturada, uma vez que a parametrização que eles utilizam per-

3.1. Introdução

mite modelar de forma direta a média usando um preditor linear e uma função de ligação geral, sendo que esta especificação é a mais similar à dos modelos lineares generalizados (McCullagh & Nelder, 1989). Ferrari & Cribari-Neto (2004) apresentam resultados de inferência estatística sob o enfoque da teoria de verossimilhança, algumas técnicas de diagnóstico e aplicações e fornecem um programa desenvolvido na linguagem de programação matricial Ox (Doornik, 2001; Cribari-Neto & Zarkos, 2003) para o ajuste do modelo e a construção de gráficos para fins de diagnóstico. Recentemente, sua proposta foi implementada no pacote estatístico R sob o nome de `betareg` disponível de forma gratuita em <http://cran.r-project.org> e no pacote estatístico SAS (SAS institute, 2005) sob os procedimentos `proc nlmixed` e `proc glimmix` (ver <http://www.ats.ucla.edu/STAT/sas/>). Adicionalmente, Espinheira, Ferrari & Cribari-Neto (2007, 2008) apresentaram novos resíduos e medidas de influência local para a análise de diagnóstico no modelo de regressão beta.

Modelos que usam a distribuição beta ou outra distribuição contínua no intervalo $(0, 1)$ são inadequados em situações em que os valores observados da variável resposta apresentam zeros e/ou uns. Assim, parece mais apropriado admitir que a distribuição da variável resposta atribui probabilidade positiva a esses valores, o que não é contemplado por nenhuma distribuição contínua em $(0, 1)$.

Recentemente, Cook, Kieschnick & McCullough (2006) propuseram um modelo de regressão para dados de financiamento, em que uma suposição é que a distribuição da variável resposta observada no intervalo $[0, 1)$ segue uma mistura entre uma distribuição degenerada no zero e uma distribuição beta. Para este modelo, a média da distribuição beta é modelada através de um preditor linear usando a ligação logito. Mais adiante, Hoff (2007) compara diferentes aproximações para modelar escores de eficiência DEA (análise envoltória de dados) em função de variáveis exógenas. Nesse trabalho, o autor compara a modelagem Tobit com dois modelos alternativos, entre eles, um modelo de regressão beta inflacionado no ponto um. Para este modelo se assume que a distribuição do escore de eficiência observado no intervalo $(0, 1]$ segue uma distribuição de mistura entre uma distribuição degenerada no ponto um e uma distribuição beta. Aqui, a probabilidade de que o escore de eficiência seja igual a um é modelada através de um preditor linear usando a ligação logito e a média da distribuição beta é modelada através de um outro preditor linear usando a mesma ligação.

Mais recentemente, Lesaffre, Rizoupoulus & Tsonaka (2007) exploram o uso da transformação logística para modelar a distribuição de escores limitados em $(0, 1)$. Para o caso em que os escores são definidos no intervalo $[0, 1]$, os autores assumem que uma variável latente em $(0, 1)$ conduz a um escore observado em $[0, 1]$, como é o caso da proporção de dias em que um paciente toma corretamente uma droga ou do índice de Barthel*. Nessa modelagem, se assume que o componente contínuo da distribuição de mistura segue uma distribuição logística-normal em que a média da distribuição é modelada através de covariadas. Adicionalmente, as probabilidades do escore ser igual a zero e um (componente discreto) podem ser modeladas usando variáveis explicativas através de regressores lineares usando a ligação logito.

Neste capítulo abordamos aspectos inferenciais para modelos de regressão beta que apresentam valores zeros e/ou uns na variável resposta. Neste sentido estenderemos o modelo de regressão beta inflacionado em

*O índice de Barthel é uma média da capacidade de uma pessoa para realizar dez atividades básicas da vida cotidiana, obtendo-se uma estimativa quantitativa do grau de dependência do indivíduo.

3.2. Modelo de regressão beta inflacionado em zero ou um

zero (Cook, Kieschnick & McCullough, 2006) e o modelo de regressão beta inflacionado em um (Hoff, 2007) ao sugerir uma parametrização que permita modelar de forma direta a média da distribuição beta e a massa de probabilidade em zero ou um usando preditores lineares e funções de ligação adequadas. Apresentamos resultados de inferência estatística sob o enfoque da teoria de verossimilhança, entre eles a estimação, a construção de intervalos de confiança de tipo assintótico e testes de hipóteses. Adicionalmente, estudaremos um modelo de regressão beta inflacionado em zero e um como uma extensão natural do modelo de regressão beta inflacionado em zero ou um. Aqui, analisamos aspectos de inferência estatística, tais como a estimação e testes de hipóteses. Também discutimos brevemente a seleção de modelos.

Este capítulo se encontra organizado da seguinte forma. Na Seção 3.2 definimos o modelo de regressão beta inflacionado em zero ou um. Abordamos aspectos de estimação por máxima verossimilhança, onde fornecemos expressões para o vetor escore, a matriz de informação de Fisher e um algoritmo de estimação, entre outros. Na Seção 3.3 definimos o modelo de regressão beta inflacionado em zero e um. Discutimos aspectos de estimação por máxima verossimilhança e apresentamos expressões para o vetor escore, a matriz de informação de Fisher e um algoritmo de estimação, entre outros. Na Seção 3.4 discutimos brevemente a seleção de modelos. Finalmente, as conclusões deste capítulo são apresentadas na Seção 3.5.

3.2 Modelo de regressão beta inflacionado em zero ou um

Há situações práticas em que dados de proporções ou frações apresentam zeros ou uns (mas não ambos) e é esse o caso que trataremos nesta seção. Consideremos, por exemplo, a situação em que o interesse recai na proporção da renda familiar gasta em telefonia celular. Há famílias que não possuem telefone celular e o gasto com esse item é nulo. Por outro lado, dificilmente alguma família gasta a totalidade da renda mensal nesse item. Neste caso, um modelo estatístico que permita adicionar a uma distribuição contínua para a variável resposta no intervalo $(0, 1)$ um ponto de massa em um dos extremos, i.e., massa de probabilidade positiva no zero ou um, parece ser mais adequado.

3.2.1 Definição

Sejam y_1, \dots, y_n variáveis aleatórias independentes cada uma com função de densidade beta inflacionada no ponto c ($c = 0$ ou $c = 1$) dada por (2.3.2), i.e., $y_t \sim \text{BI}_c(\alpha_t, \mu_t, \phi)$. O modelo de regressão beta inflacionado no ponto c ($c = 0$ ou $c = 1$) denotado por RBI_c é definido pelos componentes sistemáticos

$$\begin{aligned} h(\alpha_t) &= \sum_{i=1}^M z_{ti} \gamma_i = \zeta_t, \\ g(\mu_t) &= \sum_{i=1}^m x_{ti} \beta_i = \eta_t, \end{aligned} \tag{3.2.1}$$

em que $\gamma = (\gamma_1, \dots, \gamma_M)^\top$ e $\beta = (\beta_1, \dots, \beta_m)^\top$ são vetores de parâmetros de regressão desconhecidos tais que $\gamma \in \mathbb{R}^M$ e $\beta \in \mathbb{R}^m$, z_{t1}, \dots, z_{tM} e x_{t1}, \dots, x_{tm} são observações de variáveis exógenas conhecidas;

3.2. Modelo de regressão beta inflacionado em zero ou um

$m + M < n$. Note que os z 's e os x 's podem coincidir total ou parcialmente. Assumimos que as funções de ligação $h : (0, 1) \rightarrow \mathbb{R}$ e $g : (0, 1) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. Para este modelo podemos escolher diferentes funções de ligação h e g . Por exemplo, para μ podemos usar a especificação logito, $g(\mu) = \log\{\mu/(1 - \mu)\}$, a função probito, $g(\mu) = \Phi^{-1}(\mu)$, onde $\Phi(\cdot)$ representa a função de distribuição acumulada de uma variável aleatória normal padrão, a função de ligação log-log complementar, $g(\mu) = \log\{-\log(1 - \mu)\}$ e a função de ligação log-log, $g(\mu) = \log\{-\log(\mu)\}$, entre outras. Estas mesmas especificações podem ser usadas para α . Uma discussão detalhada destas funções de ligação pode ser encontrada em McCullagh & Nelder (1989, §4.3.1) e a utilização de outras transformações encontra-se descrita em Atkinson (1985, Cap. 7). Note que c é fixo para todas as observações e $\alpha_t = P(y_t = c)$. Já μ_t é a média condicional de y_t dado que $y_t \in (0, 1)$ e ϕ é um parâmetro de precisão que é constante para todas as observações. Seguindo Cook, Kieschnick & McCullough (2006), denotaremos por RBIZ o modelo de regressão beta inflacionado no ponto zero ($c = 0$) e por RBIU o modelo de regressão beta inflacionado no ponto um ($c = 1$).

O modelo de regressão beta inflacionado no ponto c ($c = 0$ ou $c = 1$) pertence a uma família mais ampla de modelos de regressão, conhecida como modelos aditivos generalizados para locação, escala e forma, em inglês "generalized additive models for location, scale and shape", GAMLSS (Rigby & Stasinopoulos, 2005). Nesta família os parâmetros das distribuições consideradas podem ser modelados através de preditores lineares ou através de funções não paramétricas (tipicamente suavizadores lineares).

Se consideramos o vetor de parâmetros $\theta = (\gamma^\top, \beta^\top, \phi)^\top$, a função de verossimilhança para o modelo RBI_c é da forma

$$L(\theta) = \prod_{t=1}^n \text{bi}_c(y_t; \alpha_t, \mu_t, \phi) = L_1(\gamma)L_2(\beta, \phi), \quad (3.2.2)$$

em que

$$L_1(\gamma) = \prod_{t=1}^n \alpha_t^{\mathbb{1}_{\{c\}}(y_t)} (1 - \alpha_t)^{1 - \mathbb{1}_{\{c\}}(y_t)},$$

$$L_2(\beta, \phi) = \prod_{t: y_t \in (0, 1)} f(y_t; \mu_t, \phi),$$

sendo $\mathbb{1}_A(y)$ a função indicadora, com valor 1 se $y \in A$ e 0 se $y \notin A$, e os parâmetros μ_t e α_t definidos como funções de γ e β , respectivamente, através de (3.2.1), i.e., $\alpha_t = h^{-1}(\zeta_t)$ e $\mu_t = g^{-1}(\eta_t)$. A função de densidade $\text{bi}_c(\cdot; \cdot, \cdot, \cdot)$ está definida em (2.3.2). Note que a função de verossimilhança $L(\theta)$ pode ser fatorada em dois termos, um que depende apenas do vetor de parâmetros γ e outro que depende somente do vetor de parâmetros β e de ϕ . Assim, os vetores de parâmetros γ e $(\beta^\top, \phi)^\top$ são separáveis (Pace & Salvani, 1997, p. 128) e a inferência por máxima verossimilhança sobre $(\beta^\top, \phi)^\top$ pode ser realizada de forma independente do vetor de parâmetros γ como se esse fosse conhecido e vice-versa. Note ainda que o componente discreto $L_1(\gamma)$ envolve apenas os parâmetros utilizados para modelar a probabilidade de ocorrência de zero ou de um. Por outro lado, o componente contínuo $L_2(\beta, \phi)$ envolve apenas os parâmetros usados para modelar a distribuição condicional da variável resposta dado que esta pertence ao intervalo $(0, 1)$.

3.2. Modelo de regressão beta inflacionado em zero ou um

A formulação deste modelo de regressão tem uma particular vantagem, que reside na interpretação dos parâmetros. Por exemplo, em estudos de consumo, como é o caso da despesa em telefonia celular, o interesse pode estar em analisar a proporção de renda familiar gasta nesse item. Neste caso, a distribuição da variável resposta deve considerar o fato de que algumas famílias não apresentam despesa com esse item (presença de zeros), i.e., a distribuição de probabilidades da variável resposta tem uma bimodalidade e, desta forma, um modelo de regressão beta inflacionado em zero permitiria avaliar os efeitos da heterogeneidade entre consumidores e não-consumidores.

O logaritmo da função de verossimilhança para $\theta = (\gamma^\top, \beta^\top, \phi)^\top$ é da forma

$$\ell(\theta) = \ell_1(\gamma) + \ell_2(\beta, \phi), \quad (3.2.3)$$

em que

$$\begin{aligned} \ell_1(\gamma) &= \sum_{t=1}^n \ell_t(\alpha_t), \\ \ell_2(\beta, \phi) &= \sum_{t:y_t \in (0,1)} \ell_t(\mu_t, \phi), \end{aligned} \quad (3.2.4)$$

sendo

$$\begin{aligned} \ell_t(\alpha_t) &= \mathbb{1}_{\{c\}}(y_t) \log \alpha_t + (1 - \mathbb{1}_{\{c\}}(y_t)) \log(1 - \alpha_t) \\ \ell_t(\mu_t, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t) \phi) + (\mu_t \phi - 1) \log y_t \\ &\quad + \{(1 - \mu_t) \phi - 1\} \log(1 - y_t), \end{aligned} \quad (3.2.5)$$

com $c = 0$ ou $c = 1$ dependendo do caso. Para $t = 1, \dots, n$, a variável aleatória $\mathbb{1}_{\{c\}}(y_t)$ é de tipo Bernoulli, em que $P(\mathbb{1}_{\{c\}}(y_t) = 1) = \alpha_t$ é associada a covariadas através de um preditor linear ζ_t e uma função de ligação h da forma apresentada em (3.2.1). Assim, $\ell_1(\gamma)$ é a função de log-verossimilhança de um modelo linear generalizado com resposta binária. Para maiores detalhes, ver McCullagh & Nelder (1989, §4.4.1). Adicionalmente, note que $\ell_2(\beta, \phi)$ é a função de log-verossimilhança de um modelo de regressão beta em que a variável resposta é restrita às observações em $(0,1)$ (Ferrari & Cribari-Neto, 2004).

A função escore é obtida pela diferenciação da função de log-verossimilhança com respeito a cada um dos parâmetros desconhecidos. Assim, pela separabilidade dos vetores de parâmetros γ e $(\beta^\top, \phi)^\top$, podemos obter de forma independente o escore para γ e o escore para $(\beta^\top, \phi)^\top$. Assim, para $R = 1, \dots, M$, temos

$$U_R = \frac{\partial \ell_1(\gamma)}{\partial \gamma_R} = \sum_{t=1}^n \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \frac{\partial \zeta_t}{\partial \gamma_R} = \sum_{t=1}^n \frac{\mathbb{1}_{\{c\}}(y_t) - \alpha_t}{\alpha_t(1 - \alpha_t)} \frac{d\alpha_t}{d\zeta_t} z_{tR} \quad (3.2.6)$$

e, para $r = 1, \dots, m$, temos

$$U_r = \frac{\partial \ell_2(\beta, \phi)}{\partial \beta_r} = \sum_{t:y_t \in (0,1)} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r}$$

3.2. Modelo de regressão beta inflacionado em zero ou um

$$U_r = \sum_{t:y_t \in (0,1)} \phi \left[\log\left(\frac{y_t}{1-y_t}\right) - \{\psi(\mu_t\phi) - \psi((1-\mu_t)\phi)\} \right] \frac{d\mu_t}{d\eta_t} x_{tr},$$

em que

$$\begin{aligned} \frac{d\alpha_t}{d\zeta_t} &= \frac{dh^{-1}(\zeta_t)}{d\zeta_t} = \frac{1}{h'(\alpha_t)}, \\ \frac{d\mu_t}{d\eta_t} &= \frac{dg^{-1}(\eta_t)}{d\eta_t} = \frac{1}{g'(\mu_t)}. \end{aligned}$$

Definindo

$$y_t^* = \begin{cases} \log\left(\frac{y_t}{1-y_t}\right), & \text{se } y_t \in (0, 1), \\ 0, & \text{caso contrário,} \end{cases} \quad (3.2.7)$$

e

$$\mu_t^* = E(y_t^* | \mathbb{1}_{\{c\}}(y_t) = 0) = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi), \quad (3.2.8)$$

obtemos

$$U_r = \phi \sum_{t=1}^n (1 - \mathbb{1}_{\{c\}}(y_t)) (y_t^* - \mu_t^*) \frac{d\mu_t}{d\eta_t} x_{tr}. \quad (3.2.9)$$

Se definimos os vetores $y^* = (y_1^*, \dots, y_n^*)^\top$, $y^c = (\mathbb{1}_{\{c\}}(y_1), \dots, \mathbb{1}_{\{c\}}(y_n))^\top$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)^\top$ e $\alpha^* = (\alpha_1, \dots, \alpha_n)^\top$ de dimensão $(n \times 1)$ e as matrizes diagonais $H = \text{diag}\{1 - \mathbb{1}_{\{c\}}(y_1), \dots, 1 - \mathbb{1}_{\{c\}}(y_n)\}$, $G = \text{diag}\{d\alpha_1/d\zeta_1, \dots, d\alpha_n/d\zeta_n\}$, $P = \text{diag}\{1/[\alpha_1(1-\alpha_1)], \dots, 1/[\alpha_n(1-\alpha_n)]\}$, $T = \text{diag}\{d\mu_1/d\eta_1, \dots, d\mu_n/d\eta_n\}$, de dimensão $(n \times n)$, podemos escrever matricialmente os vetores escore para γ e β , respectivamente, como

$$\begin{aligned} U_\gamma(\gamma) &= Z^\top P G (y^c - \alpha^*), \\ U_\beta(\beta, \phi) &= \phi X^\top T H (y^* - \mu^*), \end{aligned} \quad (3.2.10)$$

em que Z é uma matriz de valores fixos conhecidos $(n \times M)$ cuja t -ésima linha é $z_t^\top = (z_{t1}, \dots, z_{tM})$ e X é uma matriz de valores fixos conhecidos $(n \times m)$ com t -ésima linha $x_t^\top = (x_{t1}, \dots, x_{tm})$. Note que $U_\gamma(\gamma)$ é obtido de forma independente de $U_\beta(\beta, \phi)$ dada a fatoração da verossimilhança em (3.2.2). Finalmente, o escore para o parâmetro de precisão é obtido independentemente de γ através da derivada da função de log-verossimilhança com respeito ao parâmetro ϕ . Assim,

$$\begin{aligned} U_\phi &= \frac{\partial \ell_2(\beta, \phi)}{\partial \phi} = \sum_{t:y_t \in (0,1)} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \phi} \\ &= \sum_{t:y_t \in (0,1)} \left\{ \mu_t \left(\log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi) - \psi((1-\mu_t)\phi)] \right) + \log(1-y_t) + \psi(\phi) - \psi((1-\mu_t)\phi) \right\}. \end{aligned}$$

3.2. Modelo de regressão beta inflacionado em zero ou um

Definimos

$$s(y_t) = \begin{cases} \log(1 - y_t), & \text{se } y_t \in (0, 1), \\ 0, & \text{se } y = c. \end{cases} \quad (3.2.11)$$

Logo,

$$U_\phi = \sum_{t=1}^n (1 - \mathbb{1}_{\{c\}}(y_t)) \{ \mu_t (y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi) \}. \quad (3.2.12)$$

Se definimos $D^* = \text{diag}\{d_1^*, \dots, d_n^*\}$ sendo $d_t^* = \mu_t (y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi)$ temos que

$$U_\phi = \text{tr}(HD^*), \quad (3.2.13)$$

em que $\text{tr}(\cdot)$ é o traço de uma matriz quadrada.

Nas equações (A.1.6) - (A.1.10) do Apêndice A.1 são fornecidos os valores esperados das derivadas de segunda ordem da função de log-verossimilhança $\ell(\theta)$. Definimos as matrizes diagonais $Q = GPG = \text{diag}\{q_1, \dots, q_n\}$, $\Delta = \text{diag}\{\delta_1, \dots, \delta_n\}$, $W = \text{diag}\{w_1, \dots, w_n\}$, $D = \text{diag}\{d_1, \dots, d_n\}$ e o vetor $\mathbf{c} = (c_1, \dots, c_n)^\top$, sendo que, para $t = 1, \dots, n$, $p_t = 1/[\alpha_t(1 - \alpha_t)]$, $q_t = p_t(d\alpha_t/d\zeta_t)^2$, $\delta_t = 1 - \alpha_t$, $w_t = \psi'(\mu_t\phi) + \psi'(1 - \mu_t)\phi$, $d_t = (1 - \mu_t)^2\psi'((1 - \mu_t)\phi) + \mu_t^2\psi'(\mu_t\phi) - \psi'(\phi)$ e $c_t = \phi[\mu_t\psi'(\mu_t\phi) - (1 - \mu_t)\psi'((1 - \mu_t)\phi)]$ são funções dos parâmetros γ , β e ϕ através de α_t , μ_t e ϕ . Desta forma, a matriz de informação de Fisher $K(\theta)$ para o modelo RBI_c tem a forma

$$K(\theta) = \begin{pmatrix} K_{\gamma\gamma} & 0 \\ 0 & K_{\vartheta}(\vartheta) \end{pmatrix}, \quad (3.2.14)$$

sendo a matriz $K_{\gamma\gamma}(\gamma) = K_{\gamma\gamma}$ a matriz de informação de Fisher de γ e

$$K_{\vartheta}(\vartheta) = K_{\vartheta}(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix} \quad (3.2.15)$$

a matriz de informação de Fisher de $\vartheta = (\beta^\top, \phi)^\top$. Aqui, $K_{\gamma\gamma} = Z^\top QZ$, $K_{\beta\beta} = \phi^2 X^\top \Delta T W T X$, $K_{\beta\phi} = K_{\phi\beta}^\top = X^\top \Delta T \mathbf{c}$, $K_{\phi\phi} = \text{tr}(\Delta D)$. Observamos que a matriz em (3.2.14) não depende do ponto de inflação c . Note ainda que $K_{\gamma\gamma}$ não depende de (β, ϕ) e que $K_{\beta\beta}$, $K_{\beta\phi}$, $K_{\phi\beta}$ e $K_{\phi\phi}$ não dependem de γ . Adicionalmente, o vetor de parâmetros γ é ortogonal ao vetor de parâmetros $\vartheta = (\beta^\top, \phi)^\top$ o que implica que os respectivos componentes do vetor escore são não correlacionados. Além disso, temos deste fato que, assintoticamente, o estimador de máxima verossimilhança de γ é independente dos estimadores de máxima verossimilhança de β e ϕ . Sejam $W_{\beta\beta} = \phi^2 \Delta T W T$, $W_{\beta\phi} = \Delta T \mathbf{c}$, $W_{\phi\beta} = W_{\beta\phi}^\top$ e $W_{\phi\phi} = \text{tr}(\Delta D)$. Definimos a matriz \widetilde{W} de dimensão $(n + 1) \times (n + 1)$ por

$$\widetilde{W} = \begin{pmatrix} W_{\beta\beta} & W_{\beta\phi} \\ W_{\phi\beta} & W_{\phi\phi} \end{pmatrix}, \quad (3.2.16)$$

3.2. Modelo de regressão beta inflacionado em zero ou um

e \tilde{X} , a matriz aumentada de dimensão $(n + 1) \times (m + 1)$, da forma

$$\tilde{X} = \begin{pmatrix} X & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.2.17)$$

Então, a matriz de informação de Fisher de $\vartheta = (\beta^\top, \phi)^\top$ pode ser escrita como

$$K_\vartheta(\vartheta) = \tilde{X}^\top \tilde{W} \tilde{X}. \quad (3.2.18)$$

Usando a expressão padrão para a inversa de matrizes particionadas (ver, por exemplo, Rao, 1973, p. 33), deduz-se que a inversa da matriz de informação de Fisher é

$$K(\theta)^{-1} = \begin{pmatrix} K_\gamma(\gamma)^{-1} & 0 & 0 \\ 0 & K^{\beta\beta} & K^{\beta\phi} \\ 0 & K^{\phi\beta} & K^{\phi\phi} \end{pmatrix} \quad (3.2.19)$$

com

$$K^{\gamma\gamma} = (Z^\top QZ)^{-1},$$

$$K^{\beta\beta} = (X^\top W_{\beta\beta} X)^{-1} \left\{ I_m + \frac{X^\top T \mathbf{c} \mathbf{c}^\top T^\top X (X^\top W_{\beta\beta} X)^{-1}}{\text{tr}(D) - \mathbf{c}^\top T^\top X (X^\top W_{\beta\beta} X)^{-1} X^\top T \mathbf{c}} \right\},$$

$$K^{\beta\phi} = (K^{\phi\beta})^\top = -[\text{tr}(D) - \mathbf{c}^\top T^\top X (X^\top W_{\beta\beta} X)^{-1} X^\top T \mathbf{c}]^{-1} (X^\top W_{\beta\beta} X)^{-1} X^\top T \mathbf{c},$$

$$K^{\phi\phi} = [\text{tr}(D) - \mathbf{c}^\top T^\top X (X^\top W_{\beta\beta} X)^{-1} X^\top T \mathbf{c}]^{-1},$$

sendo I_m a matriz identidade de dimensão $m \times m$. Pela separabilidade dos parâmetros γ e $(\beta^\top, \phi)^\top$, o estimador de máxima verossimilhança de γ é obtido independentemente de $(\beta^\top, \phi)^\top$ como a solução do sistema não linear $U_\gamma(\gamma) = 0$. Já o estimador de máxima verossimilhança de $(\beta^\top, \phi)^\top$ é obtido como solução do sistema não linear $(U_\beta(\beta, \phi)^\top, U_\phi(\beta, \phi)^\top)^\top = 0$. Observamos que, em ambos os casos, tais estimadores não possuem forma fechada. Assim, eles devem ser obtidos numericamente pela maximização da função de log-verossimilhança usando um algoritmo de otimização não-linear, tal como o algoritmo de Newton (Newton-Raphson, escore de Fisher, BHHH, etc.) ou um algoritmo quasi-Newton (BFGS); ver Press et al. (1992, Capítulos 9 e 10).

3.2.2 Processo de estimação

O objetivo desta seção é apresentar um procedimento para ajuste do modelo RBI_c. Para estimar os parâmetros usamos o método escore de Fisher, bastando para tal ter conhecimento da função escore e da matriz de informação de Fisher. Inicialmente, para calcular a estimativa de máxima verossimilhança de γ , o

3.2. Modelo de regressão beta inflacionado em zero ou um

algoritmo escore de Fisher é expresso por

$$\begin{aligned}\gamma^{(m+1)} &= \gamma^{(m)} + (Z^\top Q^{(m)} Z)^{-1} Z^\top P^{(m)} G^{(m)} (y^c - \alpha^{*(m)}) \\ &= (Z^\top Q^{(m)} Z)^{-1} Z^\top Q^{(m)} \tau_1^{(m)},\end{aligned}\tag{3.2.20}$$

onde $\tau_1^{(m)} = Z\gamma^{(m)} + (Q^{(m)})^{-1}P^{(m)}G^{(m)}(y^c - \alpha^{*(m)})$.

O algoritmo escore de Fisher para calcular a estimativa de máxima verossimilhança de $\vartheta = (\beta^\top, \phi)^\top$ é expresso por

$$\vartheta^{(m+1)} = (\tilde{X}^\top \tilde{W}^{(m)} \tilde{X})^{-1} \tilde{X}^\top \tilde{W}^{(m)} \tau^*(m),\tag{3.2.21}$$

sendo

$$\tau^*(m) = \tilde{X} \vartheta^{(m)} + (\tilde{W}^{(m)})^{-1} \tilde{T}^{(m)} z^*(m)$$

um vetor de dimensão $(n+1) \times 1$, em que \tilde{T} é a matriz aumentada de dimensão $(n+1) \times (n+1)$ da forma

$$\tilde{T} = \begin{pmatrix} \phi^\top H & 0 \\ 0 & \text{tr}(HD^*) \end{pmatrix}$$

e $z^* = ((y^* - \mu^*)^\top, 1)^\top$ é um vetor de dimensão $(n+1) \times 1$.

Note que para cada iteração das expressões (A.9.1) e (A.9.2) o método escore de Fisher corresponde a uma regressão ponderada de uma variável dependente modificada sobre sua respectiva matriz modelo. Para encontrar o estimador de γ a variável modificada é τ_1 com matriz modelo Z e para encontrar o estimador de ϑ a variável modificada é τ^* com matriz modelo \tilde{X} . Os processos anteriores são repetidos até que a distância (por exemplo, euclidiana) entre $\gamma^{(m+1)}$ e $\gamma^{(m)}$ e a distância entre $\vartheta^{(m+1)}$ e $\vartheta^{(m)}$ sejam menores que uma tolerância especificada.

Em geral, os algoritmos de maximização precisam da especificação de valores iniciais. Assumiremos como estimativa inicial para γ a obtida pela regressão linear auxiliar no intercepto, i.e., consideramos como vetor inicial $\gamma_1 = h^{-1}(\sum_{t=1}^n \mathbb{1}_{\{c\}}(y_t)/n)$, $\gamma_2 = 0, \dots, \gamma_M = 0$. Note que $\sum_{t=1}^n \mathbb{1}_{\{c\}}(y_t)/n$ representa a proporção de zeros ou de uns na amostra, dependendo do caso. Seguindo Ferrari & Cribari-Neto (2004), para a estimação de β podemos tomar como vetor inicial a estimativa do vetor β obtida a partir de uma regressão linear da variável resposta transformada $g(y_1^\dagger), \dots, g(y_{m_0}^\dagger)$ em X_0 , i.e., $(X_0^\top X_0)^{-1} X_0^\top z^\dagger$, onde o vetor z^\dagger é da forma $z^\dagger = g(y^\dagger)^\top = (g(y_1^\dagger), \dots, g(y_{m_0}^\dagger))^\top$ sendo $y^\dagger = (y_1^\dagger, \dots, y_{m_0}^\dagger)^\top$ o sub-vetor da variável resposta formado pelas observações que estão no intervalo $(0, 1)$, X_0 é a sub-matriz de covariadas associadas a y^\dagger e m_0 representa o número de observações da variável resposta que se encontram no intervalo $(0, 1)$.

3.2. Modelo de regressão beta inflacionado em zero ou um

Adicionalmente, dado que $\text{Var}(y_t|y_t \in (0, 1)) = \mu_t(1-\mu_t)/(1+\phi)$, temos que $\phi = \mu_t(1-\mu_t)/\text{Var}(y_t|y_t \in (0, 1)) - 1$ e o valor inicial sugerido para ϕ é

$$\frac{1}{m_0} \sum_{t=1}^{m_0} \frac{\check{\mu}_t(1-\check{\mu}_t)}{\check{\sigma}_t^2} - 1, \quad (3.2.22)$$

em que $\check{\mu}_t = g^{-1}(x_{0t}^\top (X_0^\top X_0)^{-1} X_0^\top z^\dagger)$ e $\check{\sigma}_t^2 = \check{e}^\top \check{e} / [(m_0 - m) \{g'(\check{\mu}_t)\}^2]$. Aqui, o vetor de resíduos de mínimos quadrados ordinários,

$$\check{e} = z^\dagger - X_0 (X_0^\top X_0)^{-1} X_0^\top z^\dagger,$$

é obtido da regressão linear auxiliar de z^\dagger sobre X_0 . Note que a expressão (3.2.22) está baseada numa aproximação linear de $g(\cdot)$. Assim,

$$\text{Var}(g(y_t^\dagger)) \approx \text{Var}\{g(\mu_t) + (y_t^\dagger - \mu_t)g'(\mu_t)\} = \text{Var}(y_t^\dagger)\{g'(\mu_t)\}^2,$$

ou seja, $\text{Var}(y_t^\dagger) \approx \text{Var}\{g(y_t^\dagger)\}\{g'(\mu_t)\}^{-2}$.

Ainda, os modelos de regressão RBIZ e RBIU podem ser incorporados no pacote `gamlss` no R (Stasinopoulos, Rigby & Akantziliotou, 2006). Para a estimação dos parâmetros é utilizada máxima verossimilhança. A função de log-verossimilhança do modelo é maximizada iterativamente usando os algoritmos RS ou CG (Rigby & Stasinopoulos, 2005). Estes algoritmos usam o procedimento de retroajuste (backfitting) para melhorar o desempenho do método de escore de Fisher. Os algoritmos RS e CG usam a função de log-verossimilhança dos dados e as primeiras derivadas (opcionalmente podem utilizar o valor esperado das derivadas de segunda ordem) com respeito aos parâmetros da distribuição, neste caso, os parâmetros da distribuição (2.3.2) na forma $\nu = \alpha$, $\mu = \mu$ e $\sigma = \phi$. O algoritmo CG utilizado no pacote `gamlss` é uma generalização do algoritmo de Cole & Green (1992), em que podem ser utilizados os valores esperados das derivadas cruzadas de segunda ordem. A implementação no R dos modelos de regressão RBIZ e RBIU foi desenvolvida recentemente por Ospina (2006) através dos uso das distribuições beta inflacionadas em zero e/ou.

3.2.3 Intervalos de confiança e testes de hipóteses

Sob as condições de regularidade usuais (Cox & Hinkley, 1974, Cap. 9), temos que $\hat{\theta}$ e $K(\hat{\theta})$ são estimadores consistentes de θ e $K(\theta)$, respectivamente, em que $K(\hat{\theta})$ é a matriz de informação de Fisher (3.2.14) avaliada em $\hat{\theta}$. Assumindo que $J(\theta) = \lim_{n \rightarrow \infty} K(\theta)/n$ existe e é não-singular, temos que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_{M+m+1}(0, J(\theta)^{-1}),$$

em que $\hat{\theta} = (\hat{\gamma}^\top, \hat{\beta}^\top, \hat{\phi})^\top$ é o estimador de máxima verossimilhança de $\theta = (\gamma^\top, \beta^\top, \phi)^\top$ e \mathcal{N}_{M+m+1} é a distribuição normal $(M + m + 1)$ -variada. Aqui, $\xrightarrow{\mathcal{D}}$ denota convergência em distribuição. Assim, pela normalidade assintótica do estimador de máxima verossimilhança $\hat{\theta}$, podemos construir intervalos de

3.2. Modelo de regressão beta inflacionado em zero ou um

confiança de tipo assintótico para os parâmetros deste modelo de regressão. Portanto,

$$\left(\widehat{\gamma}_R - z_{1-\frac{\varsigma}{2}} (\widehat{K}_{RR}^{\gamma\gamma})^{1/2}, \widehat{\gamma}_R + z_{1-\frac{\varsigma}{2}} (\widehat{K}_{RR}^{\gamma\gamma})^{1/2} \right),$$

para $R = 1, \dots, M$,

$$\left(\widehat{\beta}_r - z_{1-\frac{\varsigma}{2}} (\widehat{K}_{rr}^{\beta\beta})^{1/2}, \widehat{\beta}_r + z_{1-\frac{\varsigma}{2}} (\widehat{K}_{rr}^{\beta\beta})^{1/2} \right),$$

para $r = 1, \dots, m$, e

$$\left(\widehat{\phi} - z_{1-\frac{\varsigma}{2}} (\widehat{K}^{\phi\phi})^{1/2}, \widehat{\phi} + z_{1-\frac{\varsigma}{2}} (\widehat{K}^{\phi\phi})^{1/2} \right)$$

são intervalos de confiança assintóticos (ICA) para γ_R, β_r e ϕ com coeficiente de confiança $100(1 - \varsigma)\%$, respectivamente. As variâncias assintóticas de $\widehat{\gamma}_R, \widehat{\beta}_r$ e $\widehat{\phi}$ são $\widehat{K}_{RR}^{\gamma\gamma}, \widehat{K}_{rr}^{\beta\beta}$ e $\widehat{K}^{\phi\phi}$, respectivamente, sendo $\widehat{K}_{RR}^{\gamma\gamma}$ o (R, R) -ésimo elemento da matriz $K^{\gamma\gamma}$ avaliado em $\widehat{\gamma}$, $\widehat{K}_{rr}^{\beta\beta}$ o (r, r) -ésimo elemento da matriz $K^{\beta\beta}$ avaliado em $(\widehat{\beta}^\top, \widehat{\phi}^\top)^\top$ e $\widehat{K}^{\phi\phi}$ o elemento $K^{\phi\phi}$ avaliado em $(\widehat{\beta}^\top, \widehat{\phi}^\top)^\top$. Ainda, para $0 < \varsigma < 1/2$, $z_{1-\frac{\varsigma}{2}}$ representa o quantil $1 - \varsigma/2$ da distribuição normal padrão $\mathcal{N}(0, 1)$.

Finalmente, através do método delta multivariado[†] pode ser obtido para a resposta média $\mu_t^\diamond = E(y_t)$, $t = 1, \dots, n$, do modelo RBI_c o seguinte intervalo de confiança assintótico com coeficiente de confiança $100(1 - \varsigma)\%$:

$$\left(\widehat{\mu}_t^\diamond - z_{1-\frac{\varsigma}{2}} \text{e.p.}(\widehat{\mu}_t^\diamond), \widehat{\mu}_t^\diamond + z_{1-\frac{\varsigma}{2}} \text{e.p.}(\widehat{\mu}_t^\diamond) \right),$$

em que

$$\widehat{\mu}_t^\diamond = c\widehat{\alpha}_t + (1 - \widehat{\alpha}_t)\widehat{\mu}_t = ch^{-1}(\widehat{\zeta}_t) + (1 - h^{-1}(\widehat{\zeta}_t))g^{-1}(\widehat{\eta}_t)$$

e

$$\text{e.p.}(\widehat{\mu}_t^\diamond) = \sqrt{\{(1 - \widehat{\mu}_t)/h(\widehat{\zeta}_t)\}^2 z_t^\top \widehat{K}^{\gamma\gamma} z_t + \{(1 - \widehat{\alpha}_t)/g(\widehat{\eta}_t)\}^2 x_t^\top \widehat{K}^{\beta\beta} x_t},$$

sendo $\widehat{K}^{\gamma\gamma}$ e $\widehat{K}^{\beta\beta}$ os blocos $K^{\gamma\gamma}$ e $K^{\beta\beta}$ da inversa da matriz de informação de Fisher (3.2.19) avaliados nos respectivos estimadores de máxima verossimilhança.

Ajustado o modelo RBI_c , pode ser de interesse do pesquisador realizar testes de hipóteses sobre os parâmetros do modelo, com o objetivo de verificar a significância das variáveis explicativas. Suponha que estamos interessados em testar um subconjunto dos vetores de parâmetros γ e β . Particionando os vetores de parâmetros $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$ e $\beta = (\beta_1^\top, \beta_2^\top)^\top$ em que $\gamma_1 = (\gamma_1, \dots, \gamma_{M_1})^\top$, $\gamma_2 = (\gamma_{M_1+1}, \dots, \gamma_M)^\top$, $\beta_1 = (\beta_1, \dots, \beta_{m_1})^\top$ e $\beta_2 = (\beta_{m_1+1}, \dots, \beta_m)^\top$ podemos estar interessados em testar a hipótese $\mathcal{H}_0^1 : \gamma_1 = \gamma_1^{(0)}; \beta_1 = \beta_1^{(0)}$ contra a hipótese $\mathcal{H}_1^1 : \text{violação de pelo menos uma igualdade, em que } \gamma_1^{(0)} \text{ e } \beta_1^{(0)} \text{ são vetores de parâmetros especificados de dimensões } M_1 \text{ e } m_1, \text{ respectivamente. Assumimos que } 0 \leq M_1 \leq M \text{ e}$

[†]O método delta multivariado afirma que se T_n é uma seqüência de vetores p -dimensionais tais que $\sqrt{n}[T_n - \theta] \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, \Sigma)$ e considerando $g(T_n)$ uma função a valores reais, i.e., $g : \mathbb{R}^p \rightarrow \mathbb{R}$ tal que $\dot{g}(\theta) = \partial g(x)/\partial x|_{x=\theta}$ não é identicamente nula e contínua numa vizinhança de θ , então $\sqrt{n}[g(T_n) - g(\theta)] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$ com $\tau^2 = [\dot{g}(\theta)]^\top \Sigma [\dot{g}(\theta)]$. Para maiores detalhes, ver Lehmann & Casella (2002, § 1.9).

3.2. Modelo de regressão beta inflacionado em zero ou um

$0 \leq m_1 \leq m$ (o caso trivial $M_1 = m_1 = 0$ é excluído). A estatística de teste da razão de log-verossimilhanças é

$$\Lambda = 2\{\ell(\widehat{\gamma}, \widehat{\beta}, \widehat{\phi}) - \ell(\widetilde{\gamma}, \widetilde{\beta}, \widetilde{\phi})\},$$

onde $\ell(\gamma, \beta, \phi)$ é a função de log-verossimilhança (3.2.3) e $(\widetilde{\gamma}^\top, \widetilde{\beta}^\top, \widetilde{\phi})^\top$ é o estimador de máxima verossimilhança restrito de $(\gamma^\top, \beta^\top, \phi)^\top$ obtido ao impor-se a hipótese nula. Quando o modelo restrito não se situa na fronteira do espaço paramétrico e, em condições usuais de regularidade, tem-se que, sob \mathcal{H}_0 , $\Lambda \xrightarrow{D} \chi_{M_1+m_1}^2$. Dessa forma, o teste pode ser realizado usando valores críticos aproximados de uma distribuição χ^2 com $M_1 + m_1$ graus de liberdade.

Alternativamente ao teste da razão de verossimilhanças pode ser utilizado o teste escore. Note que a hipótese nula $\mathcal{H}_0^1: \gamma_1 = \gamma_1^{(0)}; \beta_1 = \beta_1^{(0)}$ induz partições sobre as matrizes de regressores. Neste caso, sejam $Z = [Z_1, Z_2]$ e $X = [X_1, X_2]$ as respectivas matrizes particionadas de covariadas do modelo RBI_c, em que Z_1, Z_2, X_1 e X_2 são matrizes de posto completo de dimensões $n \times M_1, n \times (M - M_1), n \times m_1$, e $n \times (m - m_1)$ respectivamente. No caso em que $M_1 = M$ definimos $Z_1 = Z$. Analogamente, se $m_1 = m$ definimos $X_1 = X$. Seja $U_{1\gamma}$ o vetor de dimensão M_1 que contém os primeiros M_1 elementos do vetor escore $U_\gamma(\gamma)$ e seja $K_{11}^{\gamma\gamma}$ a matriz de dimensão $M_1 \times M_1$ formada pelas primeiras M_1 linhas e as primeiras M_1 colunas da matriz $K^{\gamma\gamma}$ definida em (3.2.19). De forma análoga, definimos $U_{1\beta}$ o vetor de dimensão m_1 que contém os primeiros m_1 elementos do vetor escore $U_\beta(\beta, \phi)$ e seja $K_{11}^{\beta\beta}$ a matriz de dimensão $m_1 \times m_1$ formada pelas primeiras m_1 linhas e as primeiras m_1 colunas da matriz $K^{\beta\beta}$ definida em (3.2.19). Desta forma, podemos mostrar que a partição induzida pela hipótese \mathcal{H}_0^1 conduz a $U_{1\gamma} = Z_1^\top PG(y^c - \alpha^*)$ e $U_{1\beta} = \phi X_1^\top TH(y^* - \mu^*)$. Daí, a estatística escore ξ pode ser escrita como a soma de duas forma quadráticas, a saber:

$$\xi = \widetilde{U}_{1\gamma}^\top \widetilde{K}_{11}^{\gamma\gamma} \widetilde{U}_{1\gamma} + \widetilde{U}_{1\beta}^\top \widetilde{K}_{11}^{\beta\beta} \widetilde{U}_{1\beta}.$$

Aqui, o ‘til’ indica que as quantidades são avaliadas no estimador de máxima verossimilhança restrito (impondo-se a hipótese nula). Sob a hipótese nula e em condições usuais de regularidade, a estatística escore tem assintoticamente distribuição $\chi_{M_1+m_1}^2$.

Finalmente, podemos considerar o teste de Wald para testar a hipótese \mathcal{H}_0^1 . Neste caso, a estatística de Wald é

$$\varpi = (\widehat{\gamma}_1 - \gamma_1^{(0)})^\top (\widehat{K}_{11}^{\gamma\gamma})^{-1} (\widehat{\gamma}_1 - \gamma_1^{(0)}) + (\widehat{\beta}_1 - \beta_1^{(0)})^\top (\widehat{K}_{11}^{\beta\beta})^{-1} (\widehat{\beta}_1 - \beta_1^{(0)}).$$

Aqui, o ‘chapéu’ denota que as quantidades são avaliadas no estimador irrestrito. Novamente, sob a hipótese nula e em condições gerais de regularidade, temos que $\varpi \xrightarrow{D} \chi_{M_1+m_1}^2$. Desta forma, o teste pode ser avaliado usando os valores críticos aproximados de uma distribuição $\chi_{M_1+m_1}^2$. Em particular, para testar a significância do r -ésimo parâmetro β_r , $r = 1, \dots, m$, podemos utilizar a raiz quadrada sinalizada da estatística de Wald, isto é $\widehat{\beta}_r / \text{s.e.}(\widehat{\beta}_r)$ em que $\text{s.e.}(\widehat{\beta}_r)$ é o erro padrão assintótico do estimador de máxima verossimilhança de $\widehat{\beta}_r$, obtido do elemento (r, r) da matriz $K^{\beta\beta}$ avaliada no estimador de máxima verossimilhança. Neste caso a distribuição limite da estatística de teste sob a hipótese nula é normal padrão. Analogamente, pode ser testada a significância do parâmetro γ_R , $R = 1, \dots, M$.

3.3. Modelo de regressão beta inflacionado em zero e um

3.2.4 Aplicação a dados simulados

Através de dados simulados, ilustramos o uso do pacote `gamlss` do R para estimar por máxima verossimilhança os parâmetros do modelo de regressão beta inflacionado RBIZ utilizando a implementação proposta por Ospina (2006). Consideramos um modelo de regressão beta inflacionado no zero (RBIZ) com estrutura

$$\begin{aligned}g(\mu_t) &= \beta_0 + \beta_1 x_t, \\h(\alpha_t) &= \gamma_0 + \gamma_1 z_t,\end{aligned}\tag{3.2.23}$$

$t = 1, \dots, n$, onde g e h são funções de ligação logito. Desta forma, a variável resposta é $y_t \sim \text{BIZ}(\alpha_t, \mu_t, \phi)$ com $t = 1, \dots, n$. Os valores das covariáveis x_t e z_t são realizações independentes de uma variável aleatória uniforme $\mathcal{U}(0, 1)$. O tamanho de amostra é $n = 500$ e os valores verdadeiros dos parâmetros são $\gamma_0 = -1.0$, $\gamma_1 = 0.5$, $\beta_0 = -1.5$, $\beta_1 = 1.5$ e $\phi = 50$. Para ajustar este modelo de regressão no pacote `gamlss` do R utilizamos

```
fit = gamlss(y~(-1+x), nu.formula=~(-1+z), family=BEZI)
```

em que `y~(-1+x)` corresponde à estrutura de regressão do componente contínuo, `nu.formula=~(-1+z)` é a estrutura de regressão do componente discreto e `family=BEZI` indica que a família a ser modelada é a distribuição beta inflacionada em zero. Maiores detalhes sobre o uso do pacote `gamlss` podem ser encontrados em Stasinopoulos, Rigby & Akantziliotou (2006).

Neste exemplo, 31.2% das observações da variável resposta são iguais a zero. A Figura 3.1 apresenta os diagramas de dispersão da variável resposta y_t , $t = 1, \dots, n$, contra as covariáveis x_t e z_t , respectivamente. Na Tabela 3.1 encontram-se as estimativas de máxima verossimilhança com seus respectivos erros padrão. Segundo o exemplo, percebemos que todas as covariáveis são significativas. Para mais detalhes do resumo de estatísticas do ajuste do modelo simulado ver o Apêndice A.2.

Tabela 3.1: Estimativas de máxima verossimilhança com erros padrão para os dados simulados; $y_t \sim \text{BIZ}(\alpha_t, \mu_t, \phi)$.

Parâmetro	γ_0	γ_1	β_0	β_1	ϕ
Estimativa	-0.96280	0.37523	-1.53124	1.53170	56.87893
Erro-Padrão	0.01656	0.02678	0.00460	0.00696	4.30175

3.3 Modelo de regressão beta inflacionado em zero e um

Nesta seção, consideramos um modelo estatístico que é uma generalização natural do modelo de regressão beta inflacionado em zero ou um. Para este modelo admitimos que a variável resposta (y) assume valores no intervalo $[0, 1]$ com probabilidade positiva de ocorrer os valores zero e um. Como exemplo, consideramos os dados referentes a um estudo em que 104 estudantes do primeiro ano de psicologia da Universidade Nacional

3.3. Modelo de regressão beta inflacionado em zero e um

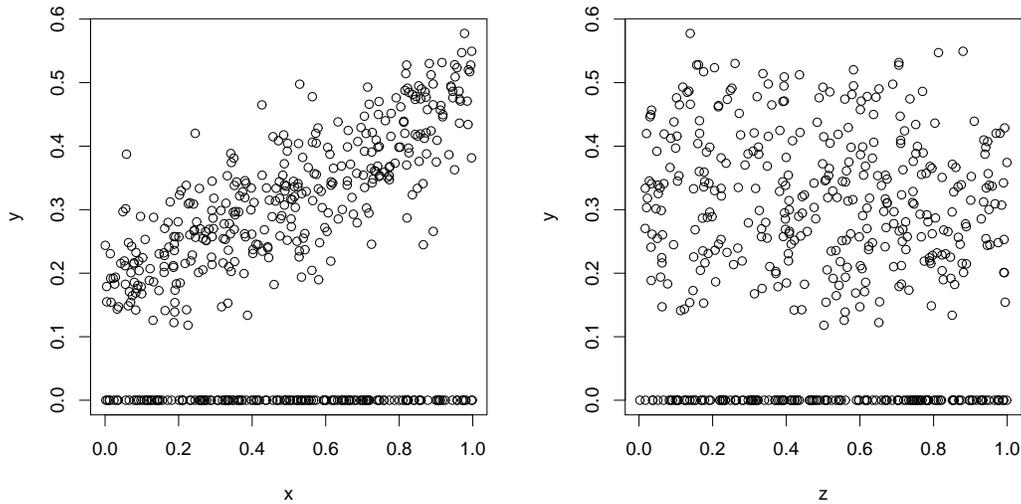


Figura 3.1: Diagrama de dispersão da resposta y_t , $t = 1, \dots, n$, contra a covariável x_t e z_t .

da Austrália simularam vereditos para uma suposta infração (Smithson & Verkuilen, 2006). Neste caso, a variável de interesse (y) é uma taxa percentual $[0, 100\%]$ que mede o grau de confiança do jurado em seu próprio veredito. Aqui, $y = 0$ representa desconfiança total no veredito do jurado e $y = 100\%$ representa confiança total no veredito do jurado. Neste estudo, a resposta é associada a covariadas. Uma possível covariada é o tipo de veredito disponível, que no caso pode ser de dois tipos. O primeiro é convencional (culpa ou absolvição) e o outro é condicional (culpa, absolvição ou falta de provas). A outra possível covariável é a evidência de testemunhos contraditórios. Assim, os dados refletem que a distribuição de probabilidades do grau de confiança no veredito do jurado tem pontos de massa nos extremos $y = 0$ e $y = 1$. Uma distribuição derivada da mistura entre uma distribuição contínua no intervalo $(0, 1)$ (por exemplo a distribuição beta) e uma distribuição de Bernoulli, que atribui probabilidades positivas aos valores 0 e 1, pode caracterizar a heterogeneidade da probabilidade do grau de confiança do jurado em seu veredito.

3.3.1 Definição

Sejam y_1, \dots, y_n variáveis aleatórias independentes, cada uma com função de densidade dada por (2.4.7), i.e., $y_t \sim \text{BIZU}(\delta_{0t}, \delta_{1t}, \mu_t, \phi)$. O modelo de regressão beta inflacionado em zero e um (RBIZU) é definido

3.3. Modelo de regressão beta inflacionado em zero e um

por (2.4.7) e pelos componentes sistemáticos

$$g(\mu_t) = \sum_{i=1}^k x_{ti}\beta_i = \eta_t, \quad (3.3.1)$$

$$H(\delta_{0t}, \delta_{1t}) = (h_0(\delta_{0t}, \delta_{1t}), h_1(\delta_{0t}, \delta_{1t})) = (\zeta_{0t}, \zeta_{1t}),$$

em que $\mu_t = E(y_t | y_t \in (0, 1))$, $\delta_{0t} = P(y_t = 0)$, $\delta_{1t} = P(y_t = 1)$ e $1 - \delta_{0t} - \delta_{1t} = P(y_t \in (0, 1))$. As funções $\eta_t = x_t^\top \beta$, $\zeta_{0t} = v_t^\top \rho$ e $\zeta_{1t} = z_t^\top \gamma$ são preditores lineares; $\beta = (\beta_1, \dots, \beta_k)^\top$, $\rho = (\rho_1, \dots, \rho_{k_0})^\top$ e $\gamma = (\gamma_1, \dots, \gamma_{k_1})^\top$ são vetores de parâmetros de regressão desconhecidos a serem estimados tais que $\beta \in \mathbb{R}^k$, $\rho \in \mathbb{R}^{k_0}$ e $\gamma \in \mathbb{R}^{k_1}$. Aqui, $x_t = (x_{t1}, \dots, x_{tk})^\top$, $v_t = (v_{t1}, \dots, v_{tk_0})^\top$ e $z_t = (z_{t1}, \dots, z_{tk_1})^\top$ representam os valores observados de k , k_0 e k_1 variáveis exógenas conhecidas respectivamente.

Admitimos que a função de ligação $g : (0, 1) \rightarrow \mathbb{R}$ é estritamente monótona e duas vezes diferenciável. Já a função H é uma transformação bijetora do conjunto $\mathcal{C} = \{(\delta_{0t}, \delta_{1t}) : 0 < \delta_{0t} < 1, 0 < \delta_{1t} < 1 - \delta_{0t}\}$ a \mathbb{R}^2 , duplamente diferenciável. Sob as condições impostas para H , garante-se que as derivadas parciais de $\delta_{0t} = h_0^*(\zeta_{0t}, \zeta_{1t})$ e de $\delta_{1t} = h_1^*(\zeta_{0t}, \zeta_{1t})$ são contínuas em \mathbb{R}^2 e δ_{0t} , δ_{1t} podem ser escritos em termos de ζ_{0t} e ζ_{1t} de forma única (Rudin, 1976. Teorema 9.28, p. 224). Note que a função H pode ser escolhida de forma geral para satisfazer as condições exigidas acima. Por exemplo, podemos considerar H tal que

$$H(\delta_{0t}, \delta_{1t}) = (h_0(\delta_{0t}, \delta_{1t}), h_1(\delta_{0t}, \delta_{1t})) = \left(h\left(\frac{\delta_{0t}}{1 - \delta_{0t} - \delta_{1t}}\right), h\left(\frac{\delta_{1t}}{1 - \delta_{0t} - \delta_{1t}}\right) \right), \quad (3.3.2)$$

sendo a função $h : \mathbb{R}^+ \rightarrow \mathbb{R}$ estritamente monótona e duas vezes diferenciável. Note que h_0 e h_1 são funções de \mathbb{R}^2 em \mathbb{R} .

Suponhamos que no modelo RBIZU tomamos a função de ligação $g(\mu_t) = \log(\mu_t / (1 - \mu_t))$ e em (3.3.2) escolhemos h como sendo a função logaritmo. Logo, $h_0(\delta_{0t}, \delta_{1t}) = \log(\delta_{0t} / (1 - \delta_{0t} - \delta_{1t})) = \zeta_{0t}$ e $h_1(\delta_{0t}, \delta_{1t}) = \log(\delta_{1t} / (1 - \delta_{0t} - \delta_{1t})) = \zeta_{1t}$. Desta forma,

$$\frac{P(y_t = 0)}{P(y_t \in (0, 1))} = \frac{\delta_{0t}}{1 - \delta_{0t} - \delta_{1t}} = \exp(\zeta_{0t}),$$

$$\frac{P(y_t = 1)}{P(y_t \in (0, 1))} = \frac{\delta_{1t}}{1 - \delta_{0t} - \delta_{1t}} = \exp(\zeta_{1t}).$$

Assim,

$$\delta_{0t} = P(y_t = 0) = \frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}},$$

$$\delta_{1t} = P(y_t = 1) = \frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}},$$

$$1 - \delta_{0t} - \delta_{1t} = P(y_t \in (0, 1)) = \frac{1}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}.$$

3.3. Modelo de regressão beta inflacionado em zero e um

Sob a especificação anterior, o modelo RBIZU será chamado de modelo de regressão logístico beta inflacionado em zero e um (RLBIZU).

Agora, consideremos o vetor de parâmetros $\theta = (\rho^\top, \gamma^\top, \beta^\top, \phi)^\top$. A função de verossimilhança para o modelo de regressão beta inflacionado em zero e um é da forma

$$L(\theta) = \prod_{t=1}^n \text{bizu}(y_t; \delta_{0t}, \delta_{1t}, \mu_t, \phi) = L_1(\rho, \gamma) L_2(\beta, \phi) \quad (3.3.3)$$

sendo $\text{bizu}(\cdot; \cdot, \cdot, \cdot, \cdot)$ a distribuição beta inflacionada em zero e um definida em (2.4.7) e

$$L_1(\rho, \gamma) = \prod_{t=1}^n \delta_{0t}^{\mathbb{1}_{\{0\}}(y_t)} \delta_{1t}^{\mathbb{1}_{\{1\}}(y_t)} (1 - \delta_{0t} - \delta_{1t})^{1 - \mathbb{1}_{\{0\}}(y_t) - \mathbb{1}_{\{1\}}(y_t)},$$

$$L_2(\beta, \phi) = \prod_{t: y_t \in (0,1)} f(y_t; \mu_t, \phi),$$

em que μ_t , δ_{0t} e δ_{1t} são definidos através de (3.3.1) como funções dos parâmetros β , ρ e γ , $\mathbb{1}_A(y)$ é a função indicadora, com valor 1 se $y \in A$ e 0 se $y \notin A$, e $f(y_t; \cdot, \cdot)$ é a densidade beta definida em (2.2.1).

Note que a função de verossimilhança $L(\theta)$ pode ser fatorada em dois termos, um que depende apenas do vetor de parâmetros $(\rho^\top, \gamma^\top)^\top$ e outro que depende somente do vetor de parâmetros β e de ϕ . Assim, os vetores de parâmetros $(\rho^\top, \gamma^\top)^\top$ e $(\beta^\top, \phi)^\top$ são separáveis (Pace & Salvani, 1997, p. 128) e a inferência por máxima verossimilhança sobre $(\beta^\top, \phi)^\top$ pode ser realizada de forma independente do vetor de parâmetros $(\rho^\top, \gamma^\top)^\top$ como se este fosse conhecido e vice-versa. Note ainda que $L_1(\rho, \gamma)$ envolve apenas os parâmetros utilizados para modelar as probabilidades de ocorrência de zero e de um (componente discreto). Por outro lado, $L_2(\beta, \phi)$ envolve apenas os parâmetros usados para modelar a distribuição condicional da variável resposta dado que esta pertence ao intervalo $(0, 1)$, i.e., do componente contínuo.

A função de log-verossimilhança do modelo RBIZU baseada numa amostra de n observações independentes é

$$\ell(\theta) = \sum_{t=1}^n \log(\text{bizu}(y_t; \delta_{0t}, \delta_{1t}, \mu_t, \phi)) = \ell_1(\rho, \gamma) + \ell_2(\beta, \phi), \quad (3.3.4)$$

em que

$$\ell_1(\rho, \gamma) = \sum_{t=1}^n \ell_t(\delta_{0t}, \delta_{1t}),$$

$$\ell_2(\beta, \phi) = \sum_{t: y_t \in (0,1)} \ell_t(\mu_t, \phi), \quad (3.3.5)$$

3.3. Modelo de regressão beta inflacionado em zero e um

onde

$$\begin{aligned}\ell_t(\delta_{0t}, \delta_{1t}) &= \mathbb{1}_{\{0\}}(y_t) \log \delta_{0t} + \mathbb{1}_{\{1\}}(y_t) \log \delta_{1t} \\ &\quad + (1 - \mathbb{1}_{\{0\}}(y_t) - \mathbb{1}_{\{1\}}(y_t)) \log(1 - \delta_{0t} - \delta_{1t}), \\ \ell_t(\mu_t, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t \\ &\quad + \{(1 - \mu_t)\phi - 1\} \log(1 - y_t).\end{aligned}\tag{3.3.6}$$

Se definimos o vetor de parâmetros $\pi = (\pi_{0t}, \pi_{1t}, \pi_{2t})$, em que $\pi_{0t} = \delta_{0t}$, $\pi_{1t} = \delta_{1t}$, $\pi_{2t} = 1 - \delta_{0t} - \delta_{1t}$, e a variável auxiliar $\mathfrak{S} = (\mathfrak{S}_{0t}, \mathfrak{S}_{1t}, \mathfrak{S}_{2t})$, sendo $\mathfrak{S}_{0t} = \mathbb{1}_{\{0\}}(y_t)$, $\mathfrak{S}_{1t} = \mathbb{1}_{\{1\}}(y_t)$ e $\mathfrak{S}_{2t} = 1 - \mathbb{1}_{\{0\}}(y_t) - \mathbb{1}_{\{1\}}(y_t)$, temos que $\pi_{0t} + \pi_{1t} + \pi_{2t} = 1$ e $\mathfrak{S}_{0t} + \mathfrak{S}_{1t} + \mathfrak{S}_{2t} = 1$. Daí, a função $\ell_1(\rho, \gamma)$ em (3.3.5) pode ser escrita como

$$\ell_1(\rho, \gamma) = \ell(\pi) = \sum_{t=1}^n \mathfrak{S}_{0t} \log \pi_{0t} + \mathfrak{S}_{1t} \log \pi_{1t} + \mathfrak{S}_{2t} \log \pi_{2t},$$

a qual corresponde à função de log-verossimilhança de um modelo trinomial na variável auxiliar \mathfrak{S} . A função $\ell_2(\beta, \phi)$ em (3.3.5) é idêntica à obtida em (3.2.5) uma vez que o componente contínuo usado para modelar a distribuição condicional da variável resposta dado que esta pertence ao intervalo $(0, 1)$ é a mesma, i.e., o componente contínuo é modelado através da distribuição beta (2.2.1). Adicionalmente, $\ell_2(\beta, \phi)$ corresponde à função de log-verossimilhança de um modelo de regressão beta (Ferrari & Cribari-Neto, 2004).

A função escore é obtida pela diferenciação da função de log-verossimilhança com respeito a cada um dos parâmetros desconhecidos (ver equações (A.3.3) - (A.3.9) do Apêndice A.3). Definimos os vetores de dimensão $n \times 1$: $y_{\{0\}} = (\mathbb{1}_{\{0\}}(y_1), \dots, \mathbb{1}_{\{0\}}(y_n))^T$, $y_{\{1\}} = (\mathbb{1}_{\{1\}}(y_1), \dots, \mathbb{1}_{\{1\}}(y_n))^T$, $y_{(0,1)} = (\mathbb{1}_{(0,1)}(y_1), \dots, \mathbb{1}_{(0,1)}(y_n))^T$, $y^* = (y_1^*, \dots, y_n^*)^T$ e $\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$, sendo y^* e μ^* definidos em (3.2.7) e (3.2.8) respectivamente. Adicionalmente, definimos matrizes diagonais de dimensão $n \times n$, sendo $\Delta_0 = \text{diag}\{1/\delta_{01}, \dots, 1/\delta_{0n}\}$, $\Delta_1 = \text{diag}\{1/\delta_{11}, \dots, 1/\delta_{1n}\}$, $\Delta_{(0,1)} = \text{diag}\{1/(1 - \delta_{01} - \delta_{11}), \dots, 1/(1 - \delta_{0n} - \delta_{1n})\}$, $T = \text{diag}\{d\mu_1/d\eta_1, \dots, d\mu_n/d\eta_n\}$, $T_0 = \text{diag}\{\partial\delta_{01}/\partial\zeta_{01}, \dots, \partial\delta_{0n}/\partial\zeta_{0n}\}$, $T_1 = \text{diag}\{\partial\delta_{11}/\partial\zeta_{11}, \dots, \partial\delta_{1n}/\partial\zeta_{1n}\}$, $T_{01} = \text{diag}\{\partial\delta_{01}/\partial\zeta_{11}, \dots, \partial\delta_{0n}/\partial\zeta_{1n}\}$ e $T_{10} = \text{diag}\{\partial\delta_{11}/\partial\zeta_{01}, \dots, \partial\delta_{1n}/\partial\zeta_{0n}\}$. Temos que o vetor escore para $\Upsilon = (\rho^T, \gamma^T)^T$ é

$$U(\Upsilon) = (U_\rho(\rho, \gamma)^T, U_\gamma(\rho, \gamma)^T)^T,$$

em que

$$\begin{aligned}U_\rho(\rho, \gamma) &= V^T T_0 (\Delta_0 y_{\{0\}} - \Delta_{(0,1)} y_{(0,1)}) + V^T T_{10} (\Delta_1 y_{\{0\}} - \Delta_{(0,1)} y_{(0,1)}), \\ U_\gamma(\rho, \gamma) &= Z^T T_{01} (\Delta_0 y_{\{0\}} - \Delta_{(0,1)} y_{(0,1)}) + Z^T T_1 (\Delta_1 y_{\{1\}} - \Delta_{(0,1)} y_{(0,1)}),\end{aligned}\tag{3.3.7}$$

V é uma matriz $n \times k_0$ de covariadas cuja t -ésima linha é v_t e Z é uma matriz $n \times k_1$ de covariadas cuja t -ésima linha é z_t . Se definimos, respectivamente, as matrizes aumentadas \tilde{T} e \tilde{Z} de dimensões $(2n \times 2n)$ e $(2n \times (k_0 + k_1))$ da forma

$$\tilde{T} = \begin{pmatrix} T_0 & T_{10} \\ T_{01} & T_1 \end{pmatrix}, \quad \tilde{Z} = \begin{pmatrix} V & 0 \\ 0 & Z \end{pmatrix},\tag{3.3.8}$$

3.3. Modelo de regressão beta inflacionado em zero e um

e o vetor auxiliar $y_{\Delta}^{\top} = ((\Delta_0 y_{\{0\}} - \Delta_{(0,1)} y_{(0,1)})^{\top}, (\Delta_1 y_{\{1\}} - \Delta_{(0,1)} y_{(0,1)})^{\top})^{\top}$, temos que o escore para Υ pode ser escrito como

$$U(\Upsilon) = \tilde{Z}^{\top} \tilde{T} y_{\Delta}.$$

Particularmente, se consideramos o modelo RLBIZU temos, através de algumas manipulações algébricas, que os vetores escore dados em (3.3.7) se reduzem a

$$\begin{aligned} U_{\rho}(\rho, \gamma) &= V^{\top} (y_{\{0\}} - \delta_0), \\ U_{\gamma}(\rho, \gamma) &= Z^{\top} (y_{\{1\}} - \delta_1), \end{aligned} \quad (3.3.9)$$

em que $\delta_0 = (\delta_{01}, \dots, \delta_{0n})^{\top}$ e $\delta_1 = (\delta_{11}, \dots, \delta_{1n})^{\top}$.

O vetor escore para β pode ser escrito matricialmente como

$$U_{\beta}(\beta, \phi) = \phi X^{\top} \text{m} \text{diag}(y_{(0,1)}) T (y^* - \mu^*), \quad (3.3.10)$$

em que $\text{m} \text{diag}(\cdot)$ é o operador que transforma um vetor numa matriz diagonal.

Finalmente, da equação (A.3.9) do Apêndice A.3 temos que o escore do parâmetro de precisão é dado por

$$U_{\phi}(\beta, \phi) = \sum_{t=1}^n \mathbb{1}_{(0,1)}(y_t) \{ \mu_t (y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi) \}, \quad (3.3.11)$$

em que $s(y_t)$ foi definido em (3.2.11). Note que as expressões (3.3.10) e (3.3.11) são as mesmas expressões obtidas em (3.2.10) e (3.2.13) respectivamente, uma vez que o componente contínuo da variável resposta é modelado pela distribuição beta definida em (2.2.1).

A seguir obteremos a matriz de informação de Fisher para $\theta = (\rho^{\top}, \gamma^{\top}, \beta^{\top}, \phi)^{\top}$. Para isso usaremos as equações (A.3.13) - (A.3.19) fornecidas no Apêndice A.3. Sejam $W = \text{diag}\{w_1, \dots, w_n\}$ e $D = \text{diag}\{d_1, \dots, d_n\}$, em que $w_t = \psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi)$ e $d_t = (1 - \mu_t)^2 \psi'((1 - \mu_t)\phi) + \mu_t^2 \psi'(\mu_t \phi) - \psi'(\phi)$ e seja $\mathbf{c} = (\mathbf{c}_1, \dots, \mathbf{c}_n)^{\top}$, com $\mathbf{c}_t = \phi [\mu_t w_t + \psi'((1 - \mu_t)\phi)]$. Pode-se mostrar (ver Apêndice A.3) que a matriz de informação de Fisher $K(\theta)$ do modelo RBIZU é

$$K(\theta) = \begin{pmatrix} K_{\Upsilon}(\Upsilon) & 0 \\ 0 & K_{\vartheta}(\vartheta) \end{pmatrix}, \quad (3.3.12)$$

em que $\Upsilon = (\rho^{\top}, \gamma^{\top})^{\top}$, $\vartheta = (\beta^{\top}, \phi)^{\top}$ e os blocos não-nulos são

$$K_{\Upsilon}(\Upsilon) = \begin{pmatrix} K_{\rho\rho} & K_{\rho\gamma} \\ K_{\gamma\rho} & K_{\gamma\gamma} \end{pmatrix}$$

3.3. Modelo de regressão beta inflacionado em zero e um

e

$$K_{\vartheta}(\vartheta) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix},$$

onde os componentes de cada bloco estão dados por

$$\begin{aligned} K_{\rho\rho} &= V^{\top} \{T_0^2 \Delta_0 - \Delta_{(0,1)}(T_0 + T_{10})^2 + \Delta_1 T_{10}^2\} V, \\ K_{\gamma\gamma} &= Z^{\top} \{T_1^2 \Delta_1 - \Delta_{(0,1)}(T_1 + T_{01})^2 + \Delta_0 T_{01}^2\} Z, \\ K_{\rho\gamma} &= K_{\gamma\rho}^{\top} = Z^{\top} \{T_0 \Delta_0 T_{01} - (T_0 + T_{10}) \Delta_{(0,1)}(T_1 + T_{01}) + T_1 \Delta_1 T_{10}\} V, \\ K_{\beta\beta} &= \phi^2 X^{\top} \{\Delta_{(0,1)}^{-1} T W T\} X, \\ K_{\beta\phi} &= K_{\phi\beta}^{\top} = X^{\top} \Delta_{(0,1)}^{-1} T \mathbf{c}, \\ K_{\phi\phi} &= \text{tr}(\Delta_{(0,1)}^{-1} D), \end{aligned}$$

sendo $\text{tr}(\cdot)$ o traço de uma matriz. Note que as expressões para W , D são idênticas às obtidas em (3.2.14) uma vez que provêm do componente contínuo do modelo, i.e., da distribuição beta.

Notamos que o vetor de parâmetros $(\rho^{\top}, \gamma^{\top})^{\top}$ é ortogonal ao vetor de parâmetros $(\beta^{\top}, \phi)^{\top}$, o que facilita a estimação dos parâmetros do modelo e o cálculo da matriz de covariância assintótica de $\hat{\theta}$ dada por $K(\theta)^{-1}$, uma vez que os respectivos componentes do vetor escore são não-correlacionados.

Se definimos \tilde{Q} como sendo a matriz de dimensão $2n \times 2n$ dada por

$$\tilde{Q} = \begin{pmatrix} Q_0 & Q_{01} \\ Q_{01}^{\top} & Q_1 \end{pmatrix}, \quad (3.3.13)$$

com elementos $Q_0 = T_0^2 \Delta_0 - \Delta_{(0,1)}(T_0 + T_{10})^2 + \Delta_1 T_{10}^2$, $Q_{01} = Q_{01}^{\top} = T_0 \Delta_0 T_{01} - (T_0 + T_{10}) \Delta_{(0,1)}(T_1 + T_{01}) + T_1 \Delta_1 T_{10}$ e $Q_1 = T_1^2 \Delta_1 - \Delta_{(0,1)}(T_1 + T_{01})^2 + \Delta_0 T_{01}^2$, temos que

$$K_{\Upsilon}(\Upsilon) = \tilde{Z}^{\top} \tilde{Q} \tilde{Z}.$$

De forma análoga ao modelo RBI_c apresentado na seção anterior, a matriz $K_{\vartheta}(\vartheta)$ pode ser escrita como

$$K_{\vartheta}(\vartheta) = \tilde{X}^{\top} \tilde{W} \tilde{X},$$

em que \tilde{X} e \tilde{W} foram definidas na seção anterior por (3.2.2) e (3.2.16), respectivamente.

Novamente, se consideramos o modelo RLBIZU temos uma simplificação nos componentes da matriz de informação de Fisher correspondentes a (γ, ρ) a saber:

3.3. Modelo de regressão beta inflacionado em zero e um

$$\begin{aligned} K_{\rho\rho} &= V^\top \text{diag}\{\delta_{01}(1 - \delta_{01}), \dots, \delta_{0n}(1 - \delta_{0n})\}V, \\ K_{\gamma\gamma} &= Z^\top \text{diag}\{\delta_{11}(1 - \delta_{11}), \dots, \delta_{1n}(1 - \delta_{1n})\}Z, \\ K_{\gamma\rho} &= K_{\rho\gamma}^\top = Z^\top \text{diag}\{-\delta_{01}\delta_{11}, \dots, -\delta_{0n}\delta_{1n}\}V. \end{aligned}$$

Usando uma expressão padrão para a inversa de matrizes particionadas, deduz-se que a inversa da matriz de informação de Fisher do modelo RBIZU é

$$K(\theta)^{-1} = \begin{pmatrix} K_{\Upsilon}(\Upsilon)^{-1} & 0 \\ 0 & K_{\vartheta}(\vartheta)^{-1} \end{pmatrix} = \begin{pmatrix} K^{\rho\rho} & K^{\rho\gamma} & 0 & 0 \\ K^{\gamma\rho} & K^{\gamma\gamma} & 0 & 0 \\ 0 & 0 & K^{\beta\beta} & K^{\beta\phi} \\ 0 & 0 & K^{\phi\beta} & K^{\phi\phi} \end{pmatrix}, \quad (3.3.14)$$

em que

$$\begin{aligned} K^{\rho\rho} &= K_{\rho\rho}^{-1}(I_{k_0} + K_{\rho\gamma}(K_{\gamma\gamma} - K_{\rho\gamma}K_{\rho\rho}^{-1}K_{\rho\gamma})^{-1}K_{\rho\gamma}K_{\rho\rho}^{-1}), \\ K^{\rho\gamma} &= K^{\gamma\rho\top} = -K_{\rho\rho}^{-1}K_{\rho\gamma}(K_{\gamma\gamma} - K_{\rho\gamma}K_{\rho\rho}^{-1}K_{\rho\gamma})^{-1}, \\ K^{\gamma\gamma} &= (K_{\gamma\gamma} - K_{\rho\gamma}K_{\rho\rho}^{-1}K_{\rho\gamma})^{-1}, \end{aligned}$$

e I_{k_0} sendo a matriz identidade de dimensão $k_0 \times k_0$. Já os componentes $K^{\beta\beta}$, $K^{\beta\phi}$ e $K^{\phi\phi}$ são os mesmos que foram obtidos em (3.2.19).

3.3.2 Processo de estimação

Para estimar os parâmetros do modelo de regressão beta inflacionado em zero e um pelo método de máxima verossimilhança, devemos solucionar as equações $U_\rho(\rho, \gamma) = 0$, $U_\gamma(\rho, \gamma) = 0$, $U_\beta(\beta, \phi) = 0$ e $U_\phi(\beta, \phi) = 0$. No entanto, percebemos que tais soluções não podem ser expressas em forma fechada. Desta forma, os estimadores de máxima verossimilhança devem ser obtidos numericamente pela maximização da função de log-verossimilhança usando um algoritmo de otimização não-linear, tal como o algoritmo de Newton (Newton-Raphson, escore de Fisher, BHHH, etc.) ou um algoritmo quasi-Newton (BFGS). Aqui, decidimos desenvolver o método escore de Fisher. Note que para encontrar a estimativa do vetor de parâmetros $\Upsilon = (\rho^\top, \gamma^\top)^\top$ o algoritmo escore de Fisher é expresso por

$$\begin{aligned} \Upsilon^{(m+1)} &= \Upsilon^{(m)} + (\tilde{Z}^\top \tilde{Q}^{(m)} \tilde{Z})^{-1} \tilde{Z}^\top \tilde{T}^{(m)} y_\Delta^{(m)} \\ &= (\tilde{Z}^\top \tilde{Q}^{(m)} \tilde{Z})^{-1} \tilde{Z}^\top \tilde{Q}^{(m)} \tilde{y}^{(m)}, \end{aligned}$$

em que a matriz \tilde{Q} foi definida em (3.3.13) e as matrizes \tilde{Z} , \tilde{T} são as definidas em (3.3.8). Desta forma, o vetor

$$\tilde{y}^{(m)} = \tilde{Z}\Upsilon^{(m)} + (\tilde{Q}^{(m)})^{-1}\tilde{T}^{(m)}y_\Delta^{(m)}$$

3.3. Modelo de regressão beta inflacionado em zero e um

é uma variável dependente modificada. O processo de estimação escore de Fisher para $\vartheta = (\beta^\top, \phi)^\top$ é idêntico ao processo de estimação (A.9.2). Cada iteração do método escore de Fisher corresponde a uma regressão ponderada de uma variável dependente modificada localmente sobre sua respectiva matriz modelo e os processos iterativos são repetidos até que a distância (por exemplo, euclidiana) entre $\Upsilon^{(m+1)}$ e $\Upsilon^{(m)}$ e a distância entre $\vartheta^{(m+1)}$ e $\vartheta^{(m)}$ com $m = 0, 1, \dots$ sejam menores que uma tolerância especificada.

O algoritmo de estimação descrito acima precisa da especificação de um valor inicial. Os valores iniciais dos parâmetros que modelam o componente contínuo do modelo RBIZU podem ser os sugeridos por Ferrari & Cribari–Neto (2004). Assim, o vetor inicial para o vetor β é obtido a partir de uma regressão linear da variável resposta transformada $g(y_1), \dots, g(y_{n_0})$ sobre X_0 , i.e., $\beta^{(0)} = (X_0^\top X_0)^{-1} X_0^\top z^\dagger$, onde o vetor z é da forma $z^\dagger = g(y^\dagger)^\top = (g(y_1), \dots, g(y_{n_0}))^\top$, e $y^\dagger = (y_1, \dots, y_{n_0})$ é o sub-vetor formado pelas observações que estão em $(0, 1)$, X_0 a sub-matriz de covariadas associadas a y^\dagger e n_0 o número de observações que se encontram no intervalo $(0, 1)$. Ainda, temos que a estimativa inicial para ϕ pode ser dada por

$$\phi^{(0)} = \frac{1}{n_0} \sum_{t=1}^{n_0} \frac{\check{\mu}_t(1 - \check{\mu}_t)}{\check{\sigma}_t^2} - 1,$$

onde $\check{\mu}_t$ é obtido aplicando $g^{-1}(\cdot)$ ao t -ésimo valor ajustado da regressão linear de z^\dagger em X_0 , i.e., $\check{\mu}_t = g^{-1}(x_{0t}^\top (X_0^\top X_0)^{-1} X_0^\top z^\dagger)$, e $\check{\sigma}_t^2 = \check{e}^\top \check{e} / [(n_0 - k) \{g'(\check{\mu}_t)\}^2]$, sendo $\check{e} = z^\dagger - (X_0^\top X_0)^{-1} X_0^\top z^\dagger$ o vetor de resíduos de mínimos quadrados da regressão linear sob a variável resposta transformada.

Assumiremos como estimativa inicial para ρ a obtida pela regressão linear auxiliar no intercepto i.e., consideramos como vetor inicial $\rho_1 = h_0^{-1}(\sum_{t=1}^n \mathbb{1}_{\{0\}}(y_t)/n, \sum_{t=1}^n \mathbb{1}_{\{1\}}(y_t)/n)$, $\rho_2 = 0, \dots, \rho_{k_0} = 0$. Finalmente, consideramos como estimativa inicial para γ a obtida pela regressão linear auxiliar no intercepto, i.e., consideramos como vetor inicial $\gamma_1 = h_1^{-1}(\sum_{t=1}^n \mathbb{1}_{\{0\}}(y_t)/n, \sum_{t=1}^n \mathbb{1}_{\{1\}}(y_t)/n)$, $\gamma_2 = 0, \dots, \gamma_{k_1} = 0$. Aqui, as funções h_0 e h_1 satisfazem (3.3.1).

Recentemente o modelo RBIZU foi implementado no pacote `gamlss` no R (Stasinopoulos, Rigby & Akantziliotou, 2006). A função de log-verossimilhança do modelo é maximizada iterativamente usando os algoritmos RS ou CG (Rigby & Stasinopoulos, 2005) e as primeiras derivadas (opcionalmente pode-se utilizar o valor esperado das derivadas de segunda ordem) com respeito aos parâmetros da distribuição (2.4.7), que são $\delta_0 = \nu/(1 + \nu + \tau)$, $\delta_1 = \tau/(1 + \nu + \tau)$, $\mu = \mu$ e $\phi + 1 = 1/\sigma^2$, com $\sigma, \nu, \tau > 0$ e $0 < \mu < 1$.

3.3.3 Intervalos de confiança e testes de hipóteses

Sob adequadas condições de regularidade, temos que os estimadores de máxima verossimilhança $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_{k_0+k_1+k+1}) = (\hat{\rho}_1, \dots, \hat{\rho}_{k_0}, \hat{\gamma}_1, \dots, \hat{\gamma}_{k_1}, \hat{\beta}_1, \dots, \beta_k, \phi)$ e $K(\hat{\theta})$ são estimadores consistentes de θ e $K(\theta)$, respectivamente, em que $K(\hat{\theta})$ é a matriz de informação de Fisher (3.3.12) avaliada em $\hat{\theta}$. Assumindo que $J(\theta) = \lim_{n \rightarrow \infty} K(\hat{\theta})/n$ existe e é não-singular, temos que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_{k+k_0+k_1+1}(0, J(\theta)^{-1}),$$

3.3. Modelo de regressão beta inflacionado em zero e um

onde \xrightarrow{D} denota convergência em distribuição e $\mathcal{N}_{k_0+k_1+k+1}$ representa a distribuição normal (k_0+k_1+k+1) -variada. Assim, pela normalidade assintótica do estimador de máxima verossimilhança $\hat{\theta}$, podemos construir intervalos de confiança assintóticos para os parâmetros do modelo RBIZU. Desta forma, para $r = 1, \dots, k_0 + k_1 + k + 1$,

$$\left(\hat{\theta}_r - z_{1-\frac{\zeta}{2}}(K(\hat{\theta})^{rr})^{1/2}, \hat{\theta}_r + z_{1-\frac{\zeta}{2}}(K(\hat{\theta})^{rr})^{1/2} \right)$$

é o intervalo de confiança assintótico para θ_r com coeficiente de confiança $100(1 - \zeta)\%$ onde $K(\hat{\theta})^{rr}$ é o (r, r) -ésimo elemento da inversa da matriz de informação de Fisher $K(\theta)$ avaliada em $\hat{\theta}$.

De forma análoga ao modelo de regressão beta inflacionado em zero ou um, podemos estar interessados em realizar testes de hipóteses sobre os parâmetros do modelo RBIZU com o intuito de verificar a significância das variáveis independentes. Suponha que estamos interessados em testar um subconjunto dos vetores de parâmetros ρ , γ e β . Particionando os vetores de parâmetros $\rho = (\rho_1^\top, \rho_2^\top)^\top$, $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$ e $\beta = (\beta_1^\top, \beta_2^\top)^\top$, em que $\rho_1 = (\rho_1, \dots, \rho_{k_{0_1}})^\top$, $\rho_2 = (\rho_{k_{0_1}+1}, \dots, \rho_{k_0})^\top$, $\gamma_1 = (\gamma_1, \dots, \gamma_{k'})^\top$, $\gamma_2 = (\gamma_{k'+1}, \dots, \gamma_{k_1})^\top$, $\beta_1 = (\beta_1, \dots, \beta_{k''})^\top$ e $\beta_2 = (\beta_{k''+1}, \dots, \beta_k)^\top$, podemos estar interessados em testar a hipótese $\mathcal{H}_0^1: \rho_1 = \rho_1^{(0)}; \gamma_1 = \gamma_1^{(0)}; \beta_1 = \beta_1^{(0)}$ contra a hipótese \mathcal{H}_1^1 : violação de pelo menos uma igualdade, em que $\rho_1^{(0)}$, $\gamma_1^{(0)}$ e $\beta_1^{(0)}$ são vetores de parâmetros especificados de dimensões k_{0_1} , k_{1_1} e k_1 , respectivamente.

Assumimos que $0 \leq k_{0_1} \leq k_0$, $0 \leq k' \leq k_1$ e $0 \leq k'' \leq k$ (o caso trivial, $k_{0_1} = k' = k'' = 0$ é excluído). A estatística de teste da razão de log-verossimilhanças é

$$\Lambda = 2\{\ell(\hat{\rho}, \hat{\gamma}, \hat{\beta}, \hat{\phi}) - \ell(\tilde{\rho}, \tilde{\gamma}, \tilde{\beta}, \tilde{\phi})\},$$

onde $\ell(\rho, \gamma, \beta, \phi)$ é a função de log-verossimilhança (3.3.4) e $(\tilde{\rho}^\top, \tilde{\gamma}^\top, \tilde{\beta}^\top, \tilde{\phi})^\top$ é o estimador de máxima verossimilhança restrito de $(\rho^\top, \gamma^\top, \beta^\top, \phi)^\top$ obtido ao se impor a hipótese nula. Se o modelo restrito não se situa na fronteira do espaço paramétrico, então, assumindo condições usuais de regularidade, tem-se que sob \mathcal{H}_0 , $\Lambda \xrightarrow{D} \chi_{k_{0_1}+k'+k''}^2$. Dessa forma, o teste pode ser realizado usando valores críticos aproximados da distribuição χ^2 com $k_{0_1} + k' + k''$ graus de liberdade.

Alternativamente ao teste da razão de verossimilhanças pode-se utilizar o teste escore. Para este caso, restringiremos o modelo para o caso do modelo RLBIZU. O teste escore para o caso geral pode ser obtido através de complicadas manipulações algébricas e por isso decidimos não apresentá-lo aqui. Note que a hipótese nula induz partições sobre as respectivas matrizes de regressores do modelo RLBIZU. Sejam $V = [V_1, V_2]$, $Z = [Z_1, Z_2]$ e $X = [X_1, X_2]$ as respectivas matrizes particionadas de covariadas do modelo de regressão beta inflacionado em zero e um, em que V_1 , V_2 , Z_1 , Z_2 , X_1 e X_2 são matrizes de posto completo de dimensões $n \times k_{0_1}$, $n \times (k_0 - k_{0_1})$, $n \times k'$, $n \times (k_1 - k')$, $n \times k_1$, e $n \times (k - k'')$, respectivamente. No caso em que $k_{0_1} = k_0$ definimos $V_1 = V$, se $k' = k_1$ definimos $Z_1 = Z$. Analogamente, se $k_1 = k$ definimos $X_1 = X$. Seja $U_{1\rho}$ um vetor de dimensão k_{0_1} que contém os primeiros k_{0_1} elementos do vetor escore $U_\rho(\rho, \gamma)$ e seja $K_{11}^{\rho\rho}$ uma matriz de dimensão $k_{0_1} \times k_{0_1}$ formada pelas primeiras k_{0_1} linhas e as primeiras k_{0_1} colunas da matriz $K^{\rho\rho}$ definida em (3.3.14). Definimos $U_{1\gamma}$ o vetor de dimensão k' que contém os primeiros k'

3.3. Modelo de regressão beta inflacionado em zero e um

elementos do vetor escore $U_\gamma(\rho, \gamma)$ e seja $K_{11}^{\gamma\gamma}$ a matriz de dimensão $k' \times k'$ formada pelas primeiras k' linhas e pelas primeiras k' colunas da matriz $K^{\gamma\gamma}$ definida em (3.3.14). De forma análoga, definimos $U_{1\beta}$ o vetor de dimensão k'' que contém os primeiros k'' elementos do vetor escore $U_\beta(\beta, \phi)$ e seja $K_{11}^{\beta\beta}$ uma matriz de dimensão $k'' \times k''$ formada pelas primeiras k'' linhas e pelas primeiras k'' colunas da matriz $K^{\beta\beta}$ definida em (3.3.14).

Desta forma, podemos mostrar que a partição induzida pela hipótese \mathcal{H}_0^1 conduz a

$$\begin{aligned} U_{1\rho} &= V_1^\top (y_{\{0\}} - \delta_0), \\ U_{1\gamma} &= Z_1^\top (y_{\{1\}} - \delta_1), \\ U_{1\beta} &= \phi X_1^\top \text{mdia}(y_{(0,1)}) T(y^* - \mu^*). \end{aligned}$$

Daí, a estatística escore ξ pode ser escrita como a soma de três formas quadráticas, a saber:

$$\xi = \tilde{U}_{1\rho}^\top \tilde{K}_{11}^{\rho\rho} \tilde{U}_{1\rho} + \tilde{U}_{1\gamma}^\top \tilde{K}_{11}^{\gamma\gamma} \tilde{U}_{1\gamma} + \tilde{U}_{1\beta}^\top \tilde{K}_{11}^{\beta\beta} \tilde{U}_{1\beta}.$$

Aqui, o ‘til’ indica que as quantidades são avaliadas no estimador de máxima verossimilhança restrito (impondo-se a hipótese nula). Sob a hipótese nula e em condições usuais de regularidade, a estatística escore tem assintoticamente distribuição $\chi_{k_{01}+k'+k''}^2$.

Finalmente, a estatística de Wald para testar a hipótese \mathcal{H}_0^1 é

$$\begin{aligned} \varpi &= (\hat{\rho}_1 - \rho_1^{(0)})^\top (\hat{K}_{11}^{\rho\rho})^{-1} (\hat{\rho}_1 - \rho_1^{(0)}) + (\hat{\gamma}_1 - \gamma_1^{(0)})^\top (\hat{K}_{11}^{\gamma\gamma})^{-1} (\hat{\gamma}_1 - \gamma_1^{(0)}) \\ &\quad + (\hat{\beta}_1 - \beta_1^{(0)})^\top (\hat{K}_{11}^{\beta\beta})^{-1} (\hat{\beta}_1 - \beta_1^{(0)}), \end{aligned}$$

em que o ‘chapéu’ denota que as quantidades estão sendo avaliadas no estimador irrestrito. Sob a hipótese nula e em condições gerais de regularidade, temos que $\varpi \xrightarrow{D} \chi_{k_{01}+k'+k''}^2$. Desta forma, o teste pode ser realizado usando os valores críticos aproximados da distribuição $\chi_{k_{01}+k'+k''}^2$. Em particular, se queremos testar a hipótese nula $\mathcal{H}_0 : \beta_i = 0$ podemos utilizar a raiz quadrada sinalizada da estatística de Wald, isto é, $\hat{\beta}_i / \text{s.e.}(\hat{\beta}_i)$, em que $\text{s.e.}(\hat{\beta}_i)$ é o erro padrão assintótico do estimador de máxima verossimilhança de $\hat{\beta}_i$ obtido do elemento (i, i) -ésimo da matriz $K^{\beta\beta}$ avaliada no estimador de máxima verossimilhança. A distribuição limite da estatística de teste sob a hipótese nula é normal padrão.

3.3.4 Aplicação a dados simulados

Agora, ilustramos o uso do pacote `gamLSS` do R para estimar por máxima verossimilhança os parâmetros do modelo de regressão beta inflacionado em zero e um, utilizando a implementação proposta por Stasinopou-

3.3. Modelo de regressão beta inflacionado em zero e um

los, Rigby & Akantziliotou (2006). Consideramos um modelo RLBIZU com estrutura

$$\begin{aligned}\log(\mu_t/(1 - \mu_t)) &= \beta_0 + \beta_1 x_t, \\ \log(\delta_{0t}/(1 - \delta_{0t} - \delta_{1t})) &= \rho_0 + \rho_1 v_t, \\ \log(\delta_{1t}/(1 - \delta_{0t} - \delta_{1t})) &= \gamma_0 + \gamma_1 z_t,\end{aligned}\tag{3.3.15}$$

$t = 1, \dots, n$. Desta forma, a variável resposta $y_t \sim \text{BIZU}(\delta_{0t}, \delta_{1t}, \mu_t, \phi)$, para $t = 1, \dots, n$. Os valores das covariáveis x_t , v_t e z_t são realizações independentes de uma variável aleatória uniforme $\mathcal{U}(0, 1)$. O tamanho de amostra é $n = 500$ e os valores verdadeiros dos parâmetros são $\rho_0 = -1.0$, $\rho_1 = 0.5$, $\gamma_0 = -1.0$, $\gamma_1 = 0.7$, $\beta_0 = -1.5$, $\beta_1 = 1.5$ e $\phi = 50$. Para ajustar este modelo de regressão no pacote `gamlss` do R utilizamos

```
fit=gamlss(ydata~(-1+x), nu.formula=~(-1+v), tau.formula=~(-1+z), family=BEINF)
```

em que `y~(-1+x)` corresponde à estrutura de regressão do componente contínuo, `nu.formula=~(-1+v)` e `tau.formula=~(-1+z)` são as estruturas de regressão do componente discreto associadas à presença de zeros e uns. Finalmente, `family=BEINF` indica que a família a ser modelada é a distribuição beta inflacionada em zero e um. Neste exemplo, 16.2% das observações são zeros e 15.4% das observações são uns. A Figura 3.2 apresenta os diagramas de dispersão da variável resposta y_t , $t = 1, \dots, n$, contra as covariáveis x_t , v_t e z_t , respectivamente.

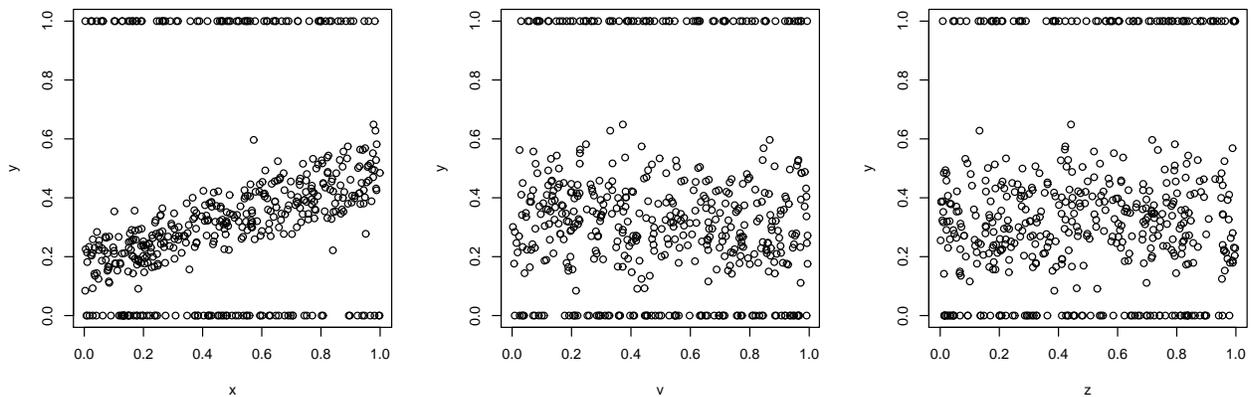


Figura 3.2: Diagrama de dispersão da resposta y_t , $t = 1, \dots, n$, contra as covariáveis x_t , v_t e z_t .

3.4. Seleção de modelos

Na Tabela 6.3 encontram-se as estimativas de máxima verossimilhança com seus respectivos erros padrão. Para mais detalhes do resumo de estatísticas do ajuste deste modelo ver o Apêndice A.4.

Tabela 3.2: Estimativas de máxima verossimilhança com erros padrão para os dados simulados; $y_t \sim \text{BIZU}(\delta_{0t}, \delta_{1t}, \mu_t, \phi)$.

Parâmetro	ρ_0	ρ_1	γ_0	γ_1	β_0	β_1	ϕ
Estimativa	-1.72401	0.51977	-1.84539	0.61558	-1.49058	1.53234	49.23541
Erro-Padrão	0.02680	0.04307	0.02772	0.04251	0.00480	0.00727	3.70233

3.4 Seleção de modelos

Um problema simples na seleção de modelos ocorre quando comparamos dois modelos que são encaixados. Como exemplo, considere os dois modelos de regressão beta inflacionados em zero com estruturas de regressão dadas na Tabela 3.3. Note que o **modelo 1** está encaixado no **modelo 2** e a seleção do melhor modelo pode ser obtida através de um teste de hipótese que confronte $\mathcal{H}_0 : \beta_2 = 0$ contra $\mathcal{H}_1 : \beta_2 \neq 0$. Neste caso podemos utilizar alguns dos testes discutidos anteriormente (razão de verossimilhanças, escore ou Wald).

Tabela 3.3: Duas estruturas de regressão para o modelo RBIZ.

modelo 1	modelo 2
$\text{logito}(\alpha_t) = \gamma + \gamma_1 z_t$	$\text{logito}(\alpha_t) = \gamma + \gamma_1 z_t$
$\text{logito}(\mu_t) = \beta + \beta_1 x_{1t}$	$\text{logito}(\mu_t) = \beta + \beta_1 x_{1t} + \beta_2 x_{2t}$

Quando os modelos não são encaixados ou quando a hipótese de interesse se situa na fronteira do espaço paramétrico, a comparação dos modelos não se reduz a um teste de hipótese padrão. Há vários critérios de seleção que podem ser usados para este propósito, inclusive critérios de informação (Hastie & Tibshirani, 1990, Cap. 3.9) tais como o o critério de informação de Akaike (AIC), o critério de informação bayesiano de Schwarz (BIC) e o critério consistente de informação de Akaike (CAIC), que são definidos como

$$\begin{aligned} \text{AIC} &= -2\hat{\ell} + 2d, \\ \text{BIC} &= -2\hat{\ell} + d \log(n), \\ \text{CAIC} &= -2\hat{\ell} + d(\log(n) + 1), \end{aligned}$$

em que $\hat{\ell}$ representa o máximo da função de log-verossimilhança, d é o número de parâmetros do modelo e n é o número de observações. Cada um destes critérios baseia-se numa penalização da verossimilhança na medida em que o modelo se torna mais complexo, i.e., cresce o número de parâmetros. Desta forma, o

3.5. Conclusões

modelo que apresentar o menor valor do critério de informação será o modelo selecionado (Akaike, 1974; Schwarz, 1978; Akaike, 1983). Maiores detalhes sobre testes de hipóteses e critérios de seleção de modelos podem ser encontrados em Buckland, Burnham, & Augustin (1997), Linhart & Zucchini (1986) e Burnham & Anderson (2002).

3.5 Conclusões

Neste capítulo tratamos de modelos de regressão beta inflacionados. Consideramos modelos de regressão em que a distribuição da variável resposta é uma mistura entre uma distribuição beta e uma distribuição de Bernoulli (modelo de regressão beta inflacionado em zero e um), degenerada em zero (modelo de regressão beta inflacionado em zero) e degenerada em um (modelo de regressão beta inflacionado em um). Aqui os parâmetros que modelam a probabilidade de ocorrer zero e/ou um são definidas através de estruturas de regressão. Também foram abordados aspectos de estimação por máxima verossimilhança para cada modelo e estudamos aspectos inferenciais, tais como a obtenção de expressões matriciais para o vetor escore e para a matriz de informação de Fisher, a construção de intervalos de confiança assintóticos, a realização de testes de hipóteses e seleção de modelos, entre outros. Apresentamos exemplos simulados para ilustrar o uso do pacote `gamlss` como ferramenta computacional para ajustar tais modelos. Estudos de simulação não apresentados aqui mostram que os estimadores de máxima verossimilhança para esta classe de modelos podem ser consideravelmente viesados em pequenas amostras. A análise do viés dos estimadores de máxima verossimilhança será o tema do próximo capítulo.

Capítulo 4

Análise de viés dos estimadores de máxima verossimilhança nos modelos de regressão beta inflacionados

4.1 Introdução

Os métodos convencionais de estimação em modelos estatísticos podem não ser viáveis ou apropriados em amostras de tamanho pequeno. Neste contexto, uma importante área de pesquisa é o estudo do comportamento de estimadores de máxima verossimilhança (EMV) em pequenas amostras, em particular a análise de viés. Usualmente, o viés é ignorado na prática sob o argumento de que este é pequeno ao ser comparado com o erro padrão do estimador do parâmetro. De fato, pode-se mostrar que o viés em geral é de ordem $O(n^{-1})$, onde n é o tamanho da amostra, ao passo que o desvio padrão assintótico é de ordem $O(n^{-1/2})$. Entretanto, para alguns modelos o viés pode ser consideravelmente maior que o correspondente desvio padrão em amostras de tamanho não “muito grande”. Portanto, é útil obtermos expressões para os vieses de segunda ordem, as quais permitem determinar a qualidade das estimativas, sobretudo para amostras de tamanho pequeno ou moderado.

Historicamente, o artigo pioneiro sobre o cálculo do viés surgiu com um trabalho de Bartlett (1953) sobre intervalos de confiança aproximados baseados nos EMV no caso uniparamétrico. Haldane (1953) e Haldane & Smith (1956) discutiram expansões assintóticas para os EMV em amostras aleatórias de um ou dois parâmetros desconhecidos, encontrando expressões de ordem $O(n^{-1})$ para os primeiros quatro cumulantes. Shenton & Wallington (1962) desenvolveram uma metodologia para obter o viés de ordem $O(n^{-1})$ dos EMV no contexto biparamétrico. Mais tarde, Shenton & Bowman (1963) desenvolveram os quatro primeiros momentos amostrais dos EMV até ordem $O(n^{-4})$. Fórmulas para o viés dos EMV até ordem $O(n^{-2})$ e covariâncias da mesma ordem, ainda para o caso multiparamétrico, foram obtidas por Bowman & Shenton (1965) e Shenton & Bowman (1977).

Posteriormente, Cox & Snell (1968) apresentaram uma expressão geral para o viés de ordem $O(n^{-1})$ dos EMV nos casos uniparamétrico e multiparamétrico. Os autores trabalham com o modelo $Y = X\beta + \epsilon$, onde X é uma matriz conhecida de dados, β um vetor de parâmetros desconhecidos e ϵ é um vetor $n \times 1$ de variáveis aleatórias não-observáveis independentes e distribuídas com média zero e variância constante. Definindo R tal que $Y = X\hat{\beta} + R$, eles examinaram as propriedades dos R 's e encontraram expressões para

4.1. Introdução

suas esperanças e covariâncias, através de uma expansão em série de Taylor de $R_i - \epsilon_i$ em termos de $\widehat{\beta} - \beta$. As fórmulas encontradas, em particular a expressão do viés de segunda ordem de β , são úteis, dado que a maioria das propriedades dos R 's são similares às dos ϵ 's quando o número de parâmetros é pequeno comparado com o tamanho da amostra. Um inconveniente na aplicação da metodologia de Cox & Snell em modelos multiparamétricos é o cálculo de inúmeros cumulantes de derivadas da função de log-verossimilhança. Assim, se p é o número de parâmetros do modelo, torna-se necessário calcular $2^p - (p + 1)$ cumulantes para encontrar os vieses dos EMV desses parâmetros.

Um método geral para o cálculo de vieses de segunda ordem numa classe de problemas de mínimos quadrados não-lineares foi apresentado por Box (1971), que considerou, entre outros aspectos, o cálculo de viés por estimação bayesiana, embora os exemplos ilustrativos tenham sido para o caso de uma distribuição a priori uniforme, na qual o estimador é de máxima verossimilhança; seu enfoque é aplicado para determinar os vieses dos estimadores de máxima verossimilhança dos parâmetros da distribuição gama. Robertson & Fryer (1970) desenvolveram um método geral para encontrar os vieses e covariâncias de ordem $O(n^{-2})$ dos estimadores de máxima verossimilhança. Sowden (1972) comparou os métodos de estimação por máxima verossimilhança, χ^2 mínimo e χ^2 mínimo modificado, num modelo de resposta quantal “dose resposta”, em termos dos seus vieses de segunda ordem, mostrando que o método de máxima verossimilhança é um procedimento mais eficiente em relação aos outros métodos. Fryer & Robertson (1972) comparam os vieses dos estimadores de momentos, de máxima verossimilhança, multinomial e χ^2 mínimo até ordem $O(n^{-1})$ e erro quadrático médio de ordem $O(n^{-2})$ para uma série de distribuições misturadas.

A partir da fórmula desenvolvida por Cox & Snell (1968), Anderson & Richardson (1979), McLanchlan (1980) e Copas (1988) encontraram os vieses dos EMV em problemas de discriminação logística. Uma análise feita por Cook, Tsai & Wei (1986) mostra o cálculo dos vieses dos EMV em modelos de regressão não-lineares com erro normal, ou seja, com função de resposta $y_i = f(x_i, \beta) + \epsilon_i$, $i = 1, \dots, n$, onde f é uma função não-linear duas vezes diferenciável em β . Eles mostram que o viés tende a zero à medida que a diferença esperada δ entre as aproximações linear e quadrática de $f(\beta)$ se aproxima de zero, de tal forma que o modelo é essencialmente linear se δ é ortogonal ao espaço coluna de V , onde V é a matriz de elementos $f_i^r = \partial f_i / \partial \beta_r$, $i = 1, \dots, n$, $r = 1, \dots, p$, e o viés b é expresso na forma $b = (V^T V)^{-1} V^T \delta$. Além disso, eles mostram como o viés afeta os métodos de diagnóstico e se estabelece uma relação entre viés e curvatura.

Young & Bakir (1987) apresentam estimadores corrigidos em modelos de regressão log-gama. Cordeiro & Klein (1994) desenvolvem fórmulas matriciais para corrigir os EMV em modelos ARMA. Paula (1992) generaliza trabalhos sobre cálculo de viés até ordem $O(n^{-1})$ e sua forma matricial encontrados na literatura estatística. Cordeiro (1993) apresenta fórmulas gerais em notação matricial para o vício de ordem $O(n^{-1})$ dos EMV de dois modelos de regressão heteroscedásticos. Cordeiro & McCullagh (1991) e Cordeiro & Cribari-Neto (1993) desenvolvem expressões simples para o cálculo do viés com aplicações em modelos econométricos. Paula & Cordeiro (1995) derivam fórmulas para os vieses de ordem $O(n^{-1})$ dos estimadores dos parâmetros em modelos não-lineares da família exponencial. Ferrari, Botter, Cordeiro & Cribari-Neto (1996) derivam o viés de ordem $O(n^{-1})$ em modelos uniparamétricos e comparam o estimador corrigido com o estimador de máxima verossimilhança usual em termos do erro quadrático médio. Cribari-Neto, Botter, Cordeiro &

4.1. Introdução

Ferrari (1998) obtêm expressões em forma fechada para o viés de segunda ordem dos EMV para várias distribuições da família exponencial uniparamétrica, mostrando graficamente como cada vício e erro quadrático do EMV e de sua versão corrigida variam de acordo com o parâmetro que indexa a distribuição. Vasconcellos & Cordeiro (1997a) generalizam os trabalhos de Box (1971) e Cook, Tsai & Wei (1986) ao obterem fórmulas gerais em notação matricial para a correção de viés em modelos multivariados não-lineares heteroscedásticos. Aubin & Cordeiro (1997) e Cordeiro & Vasconcellos (1997) obtêm fórmulas matriciais para os vieses de ordem $O(n^{-1})$ dos parâmetros da matriz de covariância num modelo de regressão linear com erros normais e matriz de covariância desconhecida.

Vasconcellos & Cordeiro (1997b) deduzem uma fórmula geral para os vieses de segunda ordem dos EMV nos modelos SUR ("Seemingly Unrelated Regressions"), os quais apresentam correlação serial entre os erros nas diferentes equações de um conjunto de equações de regressão. As fórmulas obtidas mostram como o viés pode ser expresso como o vetor das estimativas de mínimos quadrados de uma regressão linear ponderada auxiliar. Cordeiro, Vasconcellos & Santos (1998) derivam fórmulas para os vieses de segunda ordem dos EMV dos parâmetros de uma regressão não-linear com erros *t*-Student e dos parâmetros de dispersão e precisão. Cordeiro & Cribari-Neto (1998) estudam os vieses dos EMV em alguns modelos lineares generalizados e modelos não-lineares de regressão da família exponencial. Botter & Cordeiro (1998) obtêm fórmulas para os vieses de segunda ordem dos EMV em modelos lineares generalizados com parâmetros de dispersão; e tais fórmulas podem ser vistas como generalizações dos trabalhos de Cordeiro & McCullagh (1991) e Cordeiro (1993). Posteriormente, Vasconcellos & Cordeiro (1999) obtêm expressões em forma matricial para os vieses de ordem $O(n^{-1})$ dos EMV dos parâmetros do vetor de locação em modelos de regressão multivariada com erros *t*-Student. Cordeiro & Vasconcellos (1999) discutem o melhoramento dos EMV em modelos de regressão von Mises, obtendo expressões matriciais gerais para os vieses de segunda ordem dos EMV dos parâmetros que determinam a locação da distribuição.

Além do método analítico para a correção de viés dos EMV, outros métodos são conhecidos; por exemplo, Firth (1993) propõe um método "preventivo" de redução do viés dos EMV por meio de uma modificação da função escore, mostrando que este método não depende da existência de restrições na maximização da função de log-verossimilhança. Ferrari & Cribari-Neto (1998) exploram a relação entre o método analítico, baseado em expansões de Edgeworth, e a técnica de reamostragem bootstrap, sendo estes dois métodos equivalentes para a correção de viés dos EMV. Por outro lado, utilizam o método bootstrap para obter expressões gerais para correção de viés de terceira ordem dos EMV em modelos multiparamétricos. Trabalhos sobre correção de viés do EMV foram feitos por Cordeiro, Rocha, Rocha & Cribari-Neto (1997) e Cribari-Neto & Vasconcellos (2002) para os dois parâmetros que indexam a distribuição beta. Entretanto, estes resultados não são aplicados a modelos com estrutura de regressão, daí a idéia de Vasconcellos & Cribari-Neto (2003) de obter estimativas melhoradas dos EMV dos parâmetros da distribuição beta, onde tais parâmetros são modelados através de uma estrutura de regressão. Além disso, eles encontram expressões de forma fechada para o viés de segunda ordem.

4.2. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

Recentemente, Ospina, Cribari–Neto & Vasconcellos (2006) deduziram uma fórmula para o cálculo dos vieses dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta proposto por Ferrari & Cribari–Neto (2004) e forneceram estimativas corrigidas de tipo corretivo, preventivo e de bootstrap, mostrando numericamente que as estimativas corrigidas de tipo corretivo e de bootstrap apresentam desempenhos superiores em termos de viés e erro médio quadrático às suas respectivas estimativas de máxima verossimilhança. Além disso, eles apresentam intervalos de confiança do tipo assintótico, bootstrap percentil e bootstrap BCa para os parâmetros do modelo de regressão beta.

Este capítulo se encontra organizado da seguinte forma. Na Seção 4.2 serão obtidas expressões analíticas do termo de ordem n^{-1} para o vieses dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta inflacionado em zero ou um, os quais serão usadas para definir estimadores corrigidos que têm viés de ordem n^{-2} . Este tópico será estudado como uma extensão do trabalho de Ospina, Cribari–Neto & Vasconcellos (2006). Na Seção 4.3 serão obtidas expressões para os vieses de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão logístico beta inflacionado em zero e um. Na Seção 4.4 avaliamos numericamente os estimadores corrigidos através de simulação de Monte Carlo e, finalmente, as conclusões deste capítulo são apresentadas na Seção 4.5.

4.2 Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

Um dos objetivos deste trabalho é obter uma expressão para calcular o viés de segunda ordem do vetor dos estimadores de máxima verossimilhança do modelo de regressão beta inflacionado em zero ou um usando a fórmula de Cox & Snell (1968) apresentada no Apêndice A.5.

A notação a ser utilizada é introduzida a seguir. As derivadas da função de log-verossimilhança com respeito aos parâmetros desconhecidos são indicadas por índices, onde as letras R, S, \dots correspondem às derivadas em relação aos componentes γ , as letras r, s, \dots correspondem às derivadas em relação aos componentes β e a letra ϕ corresponde às derivadas com respeito ao parâmetro ϕ . Logo, $U_R = \partial\ell/\partial\gamma_R$, $U_r = \partial\ell/\partial\beta_r$, $U_\phi = \partial\ell/\partial\phi$, $U_{\phi s} = \partial^2\ell/\partial\phi\partial\beta_s$, $U_{rs\phi} = \partial^3\ell/\partial\beta_r\partial\beta_s\partial\phi$ e assim por diante. A notação para os momentos destas derivadas é dada por Lawley (1956): $\kappa_{RS} = E(U_{RS})$, $\kappa_{rs} = E(U_{rs})$, $\kappa_{\phi\phi} = E(U_\phi U_\phi)$, $\kappa_{rs,\phi} = E(U_{rs}U_\phi)$, $\kappa_{RSU} = E(U_{RSU})$, $\kappa_{rst} = E(U_{rst})$, etc., onde todos os κ 's são de ordem $O(n)$. As derivadas desses momentos são denotadas por $\kappa_{RS}^{(U)} = \partial\kappa_{RS}/\partial\gamma_U$, $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{r\phi}^{(\phi)} = \partial\kappa_{r\phi}/\partial\phi$, etc.

No modelo de regressão beta inflacionado em zero ou um temos que da separabilidade entre γ e $\vartheta = (\beta^\top, \phi)^\top$ o cálculo do viés de segunda ordem para o estimador $\hat{\gamma}$ pode ser feito de forma independente do cálculo do viés de segunda ordem do estimador $(\hat{\beta}^\top, \hat{\phi})^\top$. Supondo que as funções de ligação h e g dos componentes sistemáticos em (3.2.1) sejam arbitrárias (por exemplo, logito, probito, etc.), temos que a fórmula de Cox & Snell (1968) para calcular o viés de segunda ordem do estimador de máxima verossimilhança para

4.2. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

o b -ésimo componente do vetor $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_M)$ reduz-se a

$$B(\hat{\gamma}_b) = \sum_{R,S,U} \kappa^{bR} \kappa^{SU} \left\{ \kappa_{RS}^{(U)} - \frac{1}{2} \kappa_{RSU} \right\}. \quad (4.2.1)$$

As parcelas da fórmula (4.2.1) são deduzidas através dos momentos e cumulantes obtidos no Apêndice A.6. Usando notação matricial e a expressão (A.6.5) obtida no Apêndice A.6 obtemos o viés de segunda ordem do vetor $\hat{\gamma}$ na forma $B(\hat{\gamma}) = K^{\gamma\gamma} Z^\top W_0 \delta_{\gamma\gamma}$, isto é,

$$B(\hat{\gamma}) = (Z^\top Q Z)^{-1} Z^\top Q \tilde{\mathfrak{Z}}, \quad (4.2.2)$$

em que Z é a matriz de covariadas utilizada para modelar o componente discreto, W_0 é a matriz diagonal (A.6.3) definida no Apêndice A.6, Q é a matriz definida em (3.2.14) e $\tilde{\mathfrak{Z}} = Q^{-1} W_0 \delta_{\gamma\gamma}$ é um vetor auxiliar de dimensão $n \times 1$ onde $\delta_{\gamma\gamma}$ é o vetor formado pela diagonal da matriz $Z(Z^\top Q Z)^{-1} Z^\top$. Note que a expressão (4.2.2) corresponde a uma regressão de mínimos quadrados generalizados. Percebemos que a expressão anterior é a fórmula de correção de viés para os estimadores de máxima verossimilhança dos parâmetros de um modelo linear generalizado em que a resposta é binária (Cordeiro & McCullagh, 1991, eq. (6.3)).

Através de (4.2.2), podemos definir a estimativa de máxima verossimilhança corrigida $\tilde{\gamma}$ até segunda ordem na forma

$$\tilde{\gamma} = \hat{\gamma} - \hat{B}(\hat{\gamma}), \quad (4.2.3)$$

onde $\hat{B}(\hat{\gamma})$ denota o estimador de máxima verossimilhança de $B(\hat{\gamma})$, i.e., o vetor de parâmetros desconhecidos é substituído por sua respectiva estimativa de máxima verossimilhança.

Continuando com nossa análise, a fórmula de Cox & Snell (1968) para calcular o viés de segunda ordem do estimador de máxima verossimilhança do a -ésimo componente do vetor $\hat{\vartheta} = (\hat{\beta}_1, \dots, \hat{\beta}_m, \hat{\phi})$ reduz-se a

$$\begin{aligned} B(\hat{\vartheta}_a) &= \sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} \right\} + \kappa^{a\phi} \sum_{s,u} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} \\ &+ \sum_{r,u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} + \sum_{r,s} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} \\ &+ \kappa^{a\phi} \sum_u \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} + \kappa^{a\phi} \sum_s \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} \\ &+ \kappa^{\phi\phi} \sum_r \kappa^{ar} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} + \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\}. \end{aligned} \quad (4.2.4)$$

Note que os componentes $K_{\beta\phi} = K_{\phi\beta}^\top$ da matriz de informação de Fisher (3.2.14) não são nulos, o que leva ao cálculo de todas as parcelas de $B(\hat{\vartheta}_a)$. Para o cálculo das parcelas da fórmula (4.2.4), são obtidos no Apêndice A.7 os momentos e cumulantes de $\ell_2(\beta, \phi)$. A partir de algumas manipulações algébricas em

4.2. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

(A.7.21) derivamos o viés de segunda ordem de $\hat{\beta}$, o qual é dado por

$$B(\hat{\beta}) = K^{\beta\beta} X^\top (\Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2) X K^{\beta\phi} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi}) \\ + K^{\beta\phi} (\text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4)) X K^{\beta\phi}), \quad (4.2.5)$$

em que $\text{diagonal}(\cdot)$ é o vetor linha formado pela diagonal principal de uma matriz quadrada. Quando $\Delta = \text{diag}(\delta_1, \dots, \delta_n) = \text{diag}(1 - \alpha_1, \dots, 1 - \alpha_t)$ coincide com I_n , sendo I_n a matriz identidade de dimensão $n \times n$, temos que $\alpha_t = 0, t = 1, \dots, n$. Logo, a fórmula (4.2.5) corresponde exatamente à expressão do viés de segunda ordem $B(\hat{\beta})$ obtida em Ospina, Cribari–Neto & Vasconcellos (2006, eq. 3.1).

Agora, consideramos o vetor auxiliar $\tilde{\delta}$ de dimensão $(n + 1) \times 1$ como

$$\tilde{\delta} = \begin{pmatrix} \Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2) X K^{\beta\phi} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi} \\ \text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4)) X K^{\beta\phi} \end{pmatrix} \quad (4.2.6)$$

e definimos

$$K^{\beta*} = (K^{\beta\beta} \quad K^{\beta\phi}),$$

a matriz de dimensão $m \times (m + 1)$, obtida pelos blocos $K^{\beta\beta}$ e $K^{\beta\phi}$ da inversa da matriz de informação de Fisher (3.2.14). Desta forma, o viés de segunda ordem de $\hat{\beta}$ pode ser escrito como

$$B(\hat{\beta}) = K^{\beta*} \tilde{X}^\top \tilde{\delta}, \quad (4.2.7)$$

em que \tilde{X} é matriz de dimensão $(n + 1 \times n + 1)$ definida em (3.2.2).

De forma análoga aos cálculos anteriores, e através de algumas manipulações algébricas na equação (A.7.22) dada no Apêndice A.7, deduz-se que

$$B(\hat{\phi}) = K^{\phi\beta} X^\top (\Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2) X K^{\beta\phi} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi}) \\ + K^{\phi\phi} (\text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4)) X K^{\beta\phi}). \quad (4.2.8)$$

Se definimos

$$K^{\phi*} = (K^{\phi\beta} \quad K^{\phi\phi}),$$

a matriz de dimensão $1 \times (m + 1)$, o viés de segunda ordem de $\hat{\phi}$ pode ser escrito como

$$B(\hat{\phi}) = K^{\phi*} \tilde{X}^\top \tilde{\delta}.$$

Agora, se consideramos o vetor conjunto $\vartheta = (\beta^\top, \phi)^\top$, podemos escrever uma expressão para o viés de segunda ordem do estimador de máxima verossimilhança de ϑ na forma

$$B(\hat{\vartheta}) = K(\vartheta)^{-1} \tilde{X}^\top \tilde{\delta} = (\tilde{X}^\top \tilde{W} \tilde{X})^{-1} \tilde{X}^\top \tilde{\delta}.$$

4.2. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

Esta expressão pode ser escrita como

$$B(\hat{\vartheta}) = (\tilde{X}^T \tilde{W} \tilde{X})^{-1} \tilde{X}^T \tilde{W} \tilde{\xi}, \quad (4.2.9)$$

onde $\tilde{\xi} = \tilde{W}^{-1} \tilde{\delta}$. Desta forma, os coeficientes do vetor $B(\hat{\vartheta})$ podem ser estimados a partir de uma regressão de mínimos quadrados generalizados na variável auxiliar $\tilde{\xi}$. Novamente, se $\Delta = I_n$, a fórmula anterior coincide com a expressão do viés de segunda ordem obtida em Ospina, Cribari-Neto & Vasconcellos (2006, equação 3.4). Como uma observação, note que a expressão para correção do viés de segunda ordem do estimador $\hat{\vartheta}$ é mais interessante pois envolve os parâmetros associados aos componentes discreto e contínuo do modelo. Em vez do que acontece com a expressão para correção do viés de segunda ordem do estimador $\hat{\gamma}$. Afinal, γ , como é dito é um vetor de um modelo binário.

Finalmente, podemos definir o estimador de máxima verossimilhança corrigido $\tilde{\vartheta}$ da forma

$$\tilde{\vartheta} = \hat{\vartheta} - \hat{B}(\hat{\vartheta}), \quad (4.2.10)$$

em que $\hat{B}(\hat{\vartheta})$ denota o estimador de máxima verossimilhança de $B(\hat{\vartheta})$, i.e., os parâmetros desconhecidos são substituídos por suas respectivas estimativas de máxima verossimilhança.

4.2.1 Corrigindo os vieses de $\hat{\zeta}$ e $\hat{\eta}$

Sejam $\zeta = (\zeta_1, \dots, \zeta_n)^T = Z\gamma$ e $\eta = (\eta_1, \dots, \eta_n)^T = X\beta$ preditores lineares. Os estimadores de máxima verossimilhança de ζ e η são $\hat{\zeta} = Z\hat{\gamma}$ e $\hat{\eta} = X\hat{\beta}$ respectivamente. Assim, $\hat{\zeta} - \zeta = Z(\hat{\gamma} - \gamma)$ e $\hat{\eta} - \eta = X(\hat{\beta} - \beta)$. Conseqüentemente, o viés de segunda ordem de $\hat{\zeta}$ é

$$B(\hat{\zeta}) = ZB(\hat{\gamma}) = Z(Z^T QZ)^{-1} Z^T Q \tilde{\zeta}, \quad (4.2.11)$$

em que $B(\hat{\gamma})$ é o viés de segunda ordem de $\hat{\gamma}$ obtido em (4.2.2). Analogamente, o viés de segunda ordem de $\hat{\eta}$ é

$$B(\hat{\eta}) = XB(\hat{\beta}) = XK^{\beta*} \tilde{X}^T \tilde{\delta}, \quad (4.2.12)$$

em que $K^{\beta*}$ é a matriz de dimensão $m \times (m + 1)$, obtida pelos blocos $K^{\beta\beta}$ e $K^{\beta\phi}$ da inversa da matriz de informação de Fisher (3.2.14) e \tilde{X} é a matriz de dimensão $(n + 1 \times n + 1)$ definida em (3.2.2).

De forma análoga aos estimadores definidos em (4.2.3) e (4.2.10), podemos construir estimadores de máxima verossimilhança corrigidos para os vetores ζ e η na forma

$$\begin{aligned} \tilde{\zeta} &= \hat{\zeta} - \hat{B}(\hat{\zeta}), \\ \tilde{\eta} &= \hat{\eta} - \hat{B}(\hat{\eta}), \end{aligned} \quad (4.2.13)$$

em que $\hat{B}(\hat{\zeta})$ e $\hat{B}(\hat{\eta})$ denotam os estimadores de máxima verossimilhança de $B(\hat{\zeta})$ e $B(\hat{\eta})$, i.e., os parâmetros desconhecidos são substituídos por seus respectivos estimadores de máxima verossimilhança.

4.2. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

4.2.2 Corrigindo os vieses de $\hat{\mu}$ e $\hat{\alpha}$

Consideremos os vetores de parâmetros $\mu = (\mu_1, \dots, \mu_n)^\top$, $\alpha = (\alpha_1, \dots, \alpha_n)^\top$, onde $\alpha_t = h^{-1}(\zeta_t)$ e $\mu_t = g^{-1}(\eta_t)$ para $t = 1, \dots, n$, sendo g e h funções de ligação. Para obter o viés de segunda ordem de $\hat{\alpha}_t$ expandimos $h^{-1}(\hat{\zeta}_t)$ em série de Taylor, até segunda ordem, numa vizinhança de ζ_t . Assim,

$$h^{-1}(\hat{\zeta}_t) = h^{-1}(\zeta_t) + \frac{dh^{-1}(\zeta_t)}{d\zeta_t}(\hat{\zeta}_t - \zeta_t) + \frac{1}{2} \frac{d^2h^{-1}(\zeta_t)}{d\zeta_t^2}(\hat{\zeta}_t - \zeta_t)^2 + O_p(n^{-3/2}).$$

Arranjando os termos e tomando o valor esperado, temos que o viés de segunda ordem de $\hat{\alpha}_t$ é dado por

$$B(\hat{\alpha}_t) = B(h^{-1}(\hat{\zeta}_t)) = B(\hat{\zeta}_t) \frac{d\alpha_t}{d\zeta_t} + \frac{1}{2} V(\hat{\zeta}_t) \frac{d^2\alpha_t}{d\zeta_t^2},$$

em que $V(\hat{\zeta}_t)$ é a variância assintótica de $\hat{\zeta}_t$, obtida do (t, t) -ésimo elemento da matriz $Z(Z^\top QZ)^{-1}Z^\top$ e $B(\hat{\zeta}_t)$ é o viés de ordem n^{-1} de $\hat{\zeta}_t$. Desta forma, se definimos as matrizes diagonais $G_1 = \text{diag}(d\alpha_1/d\zeta_1, \dots, d\alpha_n/d\zeta_n)$ e $G_2 = \text{diag}(d^2\alpha_1/d\zeta_1^2, \dots, d^2\alpha_n/d\zeta_n^2)$, podemos escrever matricialmente

$$\begin{aligned} B(\hat{\alpha}) &= G_1 B(\hat{\zeta}) + \frac{1}{2} G_2 \text{diagonal}(ZK^{\gamma\gamma}Z^\top) \\ &= \frac{1}{2} (G_2 \text{diagonal}(Z(Z^\top QZ)^{-1}Z^\top) + 2G_1 Z(Z^\top QZ)^{-1}Z^\top Q\tilde{\mathfrak{Z}}), \end{aligned}$$

em que $\text{diagonal}(A)$ é um vetor formado pela diagonal principal da matriz quadrada A , $K^{\gamma\gamma}$ é a variância assintótica de $\hat{\gamma}$, obtida em (3.2.19), e $B(\hat{\zeta})$ é o viés de ordem n^{-1} de $\hat{\zeta}$, dado em (4.2.11). De forma análoga, podemos obter o viés de segunda ordem de $\hat{\mu}$. Para isto, expandimos $g^{-1}(\hat{\eta}_t)$ em série de Taylor, até segunda ordem, numa vizinhança de η_t . Assim,

$$g^{-1}(\hat{\eta}_t) = g^{-1}(\eta_t) + \frac{dg^{-1}(\eta_t)}{d\eta_t}(\hat{\eta}_t - \eta_t) + \frac{1}{2} \frac{d^2g^{-1}(\eta_t)}{d\eta_t^2}(\hat{\eta}_t - \eta_t)^2 + O_p(n^{-3/2}).$$

Ordenando os termos e tomando o valor esperado, temos que o viés de segunda ordem de $\hat{\mu}_t$ é dado por

$$B(\hat{\mu}_t) = B(g^{-1}(\hat{\eta}_t)) = B(\hat{\eta}_t) \frac{d\mu_t}{d\eta_t} + \frac{1}{2} V(\hat{\eta}_t) \frac{d^2\mu_t}{d\eta_t^2},$$

em que $V(\hat{\eta}_t)$ é a variância assintótica de $\hat{\eta}_t$ obtida do (t, t) -ésimo elemento da matriz $XK^{\beta\beta}X^\top$, sendo $K^{\beta\beta}$ a variância assintótica de $\hat{\beta}$ obtida em (3.2.19) e $B(\hat{\eta}_t)$ o viés de ordem n^{-1} de $\hat{\eta}_t$ dado em (4.2.12). Desta forma, se definimos as matrizes diagonais $T = \text{diag}(d\mu_1/d\eta_1, \dots, d\mu_n/d\eta_n)$ e $T_1 = \text{diag}(d^2\mu_1/d\eta_1^2, \dots, d^2\mu_n/d\eta_n^2)$,

4.2. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

..., $d^2\mu_n/d\eta_n^2$), podemos escrever matricialmente

$$\begin{aligned} B(\hat{\mu}) &= TB(\hat{\eta}) + \frac{1}{2}T_1 \text{diagonal}(XK^{\beta\beta}X^\top) \\ &= \frac{1}{2}(T_1 \text{diagonal}(XK^{\beta\beta}X^\top) + 2TXK^{\beta*}\tilde{X}^\top\tilde{\delta}), \end{aligned}$$

em que $K^{\beta*} = (K^{\beta\beta}, K^{\beta\phi})$ é o bloco superior da inversa da matriz de informação de Fisher definida em (3.2.19).

De forma análoga aos estimadores definidos em (4.2.3) e (4.2.10) podemos construir estimadores de máxima verossimilhança corrigidos para os vetores α e μ da forma

$$\begin{aligned} \tilde{\alpha} &= \hat{\alpha} - \hat{B}(\hat{\alpha}), \\ \tilde{\mu} &= \hat{\mu} - \hat{B}(\hat{\mu}), \end{aligned} \tag{4.2.14}$$

em que $\hat{B}(\hat{\alpha})$ e $\hat{B}(\hat{\gamma})$ denotam os estimadores de máxima verossimilhança de $B(\hat{\alpha})$ e $B(\hat{\gamma})$, i.e., os parâmetros desconhecidos são substituídos por seus respectivos estimadores de máxima verossimilhança.

4.2.3 Viés de $\hat{\sigma}$

Para o modelo de regressão beta inflacionado em zero ou um podemos considerar uma reparametrização do parâmetro de precisão ϕ . Seja $\sigma = T(\phi)$ onde $T(\cdot)$ é uma função contínua duas vezes diferenciável. Pelo princípio de invariância dos estimadores de máxima verossimilhança temos que $\hat{\sigma} = T(\hat{\phi})$. Se tomamos uma expansão em série de Taylor da função T até o terceiro termo numa vizinhança de ϕ , obtemos

$$T(\hat{\phi}) = T(\phi) + T'(\phi)(\hat{\phi} - \phi) + \frac{1}{2}T''(\phi)(\hat{\phi} - \phi)^2 + O_p(n^{-3/2}).$$

Tomando o valor esperado chega-se ao viés de segunda ordem de $\hat{\sigma}$ dado por

$$B(\hat{\sigma}) = T'(\phi)B(\hat{\phi}) + \frac{1}{2}T''(\phi)V(\hat{\phi}), \tag{4.2.15}$$

em que $B(\hat{\phi})$ é o viés de segunda ordem de $\hat{\phi}$, obtido em (4.2.8) e $V(\hat{\phi}) = K^{\phi\phi}$ é a variância assintótica de $\hat{\phi}$ dada em (3.2.19). Na Tabela 4.1 apresentamos diferentes reparametrizações de ϕ , i.e., σ com seus respectivos vieses de segunda ordem para o estimador de máxima verossimilhança $\hat{\sigma}$.

De (4.2.15) podemos definir um estimador corrigido para σ na forma

$$\tilde{\sigma} = \hat{\sigma} - \hat{B}(\hat{\sigma}), \tag{4.2.16}$$

sendo $\hat{B}(\hat{\sigma})$ o estimador de máxima verossimilhança de $B(\hat{\sigma})$, i.e., o parâmetro desconhecido é substituído por seu respectivo estimador de máxima verossimilhança.

4.2. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero ou um

Tabela 4.1: Viés de $\hat{\sigma} = \mathcal{T}(\hat{\phi})$

$\sigma = \mathcal{T}(\phi)$	$B(\hat{\sigma})$
$1/(\phi + 1)$	$-\frac{B(\hat{\phi})}{(\phi+1)^2} + \frac{V(\hat{\phi})}{(\phi+1)^3}$
$1/(\phi + 1)^2$	$-\frac{2B(\hat{\phi})}{(\phi+1)^3} + \frac{3V(\hat{\phi})}{(\phi+1)^4}$
$\log \phi$	$\frac{B(\hat{\phi})}{\phi} - \frac{V(\hat{\phi})}{2\phi^2}$

Através de dados simulados, ilustramos o comportamento do viés relativo $B(\hat{\sigma})/\sigma$ como uma função de ϕ . Consideramos um modelo regressão beta inflacionado no zero (RBIZ) com estrutura

$$\begin{aligned} g(\mu_t) &= \beta_0 + \beta_1 x_t, \\ h(\alpha_t) &= \gamma_0 + \gamma_1 z_t, \end{aligned} \tag{4.2.17}$$

$t = 1, \dots, n$, onde g e h são funções de ligação logito. Desta forma, a variável resposta $y_t \sim \text{BIZ}(\alpha_t, \mu_t, \phi)$ com $t = 1, \dots, n$. Os valores das covariáveis x_t e z_t são realizações independentes de uma variável aleatória uniforme $\mathcal{U}(0, 1)$. O tamanho de amostra é $n = 100$ e os valores verdadeiros dos parâmetros são $\gamma_0 = -0.5$, $\gamma_1 = 0.5$, $\beta_0 = -0.5$, $\beta_1 = 1.5$. Aqui, ϕ assume os valores 2, 5, 10, 20, 40, 60, 100, 150 e 500.

A Figura 4.1 apresentamos o comportamento do viés relativo de $B(\hat{\sigma})/\sigma$ como uma função de ϕ . Note que o viés relativo para as reparametrizações $\sigma = \phi$ e $\sigma = 1/(\phi + 1)^2$ tende a um valor constante distinto de zero à medida em que ϕ aumenta. Neste mesmo gráfico vemos que na parametrização $\sigma = 1/(\phi + 1)$ o viés relativo se comporta quase como a função constante zero e este resultado é válido mesmo para valores moderados de n (por exemplo, $n = 80$). Finalmente, na parametrização $\sigma = \log \phi$ observa-se que o viés relativo é uma função que decresce para zero e isto é consistente com o fato de que a informação $K_{\phi\phi}$ diminui à medida em que ϕ aumenta, dado que o viés de segunda ordem de $\hat{\phi}$ está diretamente relacionado com a informação de ϕ . Diante destes fatos, pode ser sugerida a utilização da parametrização $\sigma = 1/(\phi + 1)$, pois mesmo com valores moderados de n e valores pequenos de ϕ o viés relativo permanece estável e próximo de zero.

4.3. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero e um

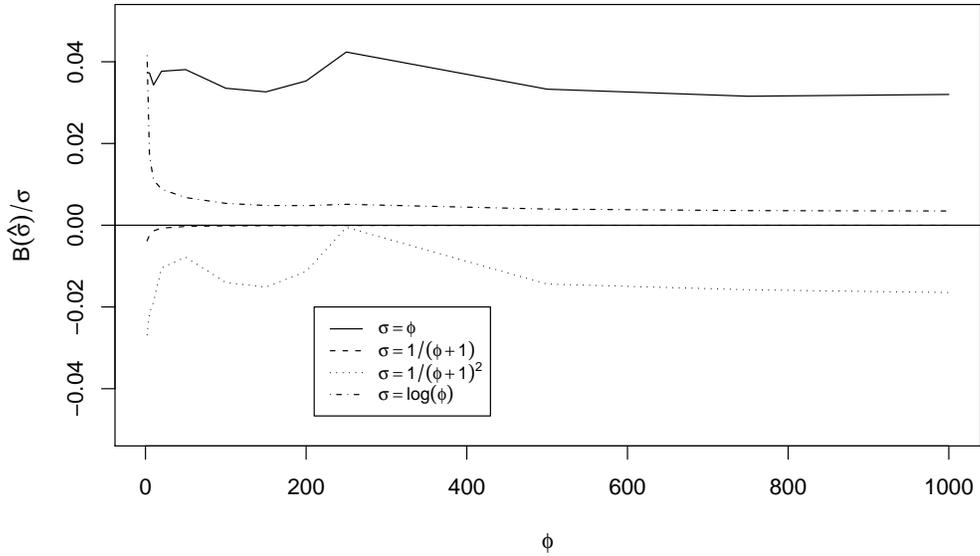


Figura 4.1: Viés relativo de segunda ordem para $\hat{\sigma} = \mathcal{T}(\hat{\phi})$.

4.3 Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero e um

Nesta seção vamos estudar as propriedades dos estimadores de máxima verossimilhança em pequenas amostras para os parâmetros do modelo de regressão beta inflacionado em zero e um com a seguinte estrutura:

$$\begin{aligned} g(\mu_t) &= \eta_t, \\ \log(\delta_{0t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{0t}, \\ \log(\delta_{1t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{1t}, \end{aligned}$$

$t = 1, \dots, n$, e o parâmetro de precisão ϕ constante para todas as observações. Note que este modelo é um caso particular do modelo (3.3.1).

No modelo de regressão beta inflacionado em zero e um temos que, da separabilidade dos vetores $\Upsilon = (\rho^\top, \gamma^\top)^\top$ e $\vartheta = (\beta^\top, \phi)^\top$, o viés de segunda ordem do estimador $\hat{\Upsilon}$ pode ser obtido de forma independente do viés de segunda ordem do estimador $\hat{\vartheta}$. Para o cálculo do viés, introduzimos a seguinte notação: As derivadas da função de log-verossimilhança com respeito aos parâmetros desconhecidos são indicadas por

4.3. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero e um

índices: r', s', \dots correspondem às derivadas em relação aos componentes ρ, r'', s'', \dots , correspondem às derivadas em relação aos componentes γ ; r, s, \dots correspondem às derivadas em relação aos componentes β e ϕ corresponde às derivadas com respeito ao parâmetro ϕ . Assim, $U_r' = \partial\ell/\partial\rho_{r'}$, $U_{r''} = \partial\ell/\partial\gamma_{r''}$, $U_r = \partial\ell/\partial\beta_r$, $U_\phi = \partial\ell/\partial\phi$, $U_{\phi s} = \partial^2\ell/\partial\phi\partial\beta_s$, $U_{rs\phi} = \partial^3\ell/\partial\beta_r\partial\beta_s\partial\phi$ e assim por diante.

A notação para os momentos destas derivadas é dada por Lawley (1956): $\kappa_{r's'} = E(U_{r's'})$, $\kappa_{r''s''} = E(U_{r''s''})$, $\kappa_{rs} = E(U_{rs})$, $\kappa_{\phi\phi} = E(U_\phi U_\phi)$, $\kappa_{rs,\phi} = E(U_{rs}U_\phi)$, $\kappa_{r's'u'} = E(U_{r's'u'})$, $\kappa_{r''s''u''} = E(U_{r''s''u''})$, $\kappa_{rst} = E(U_{rst})$, etc., e, em geral, são de ordem $O(n)$. As derivadas dos κ 's são denotadas por $\kappa_{r's'}^{(u')} = \partial\kappa_{r's'}/\partial\rho_{u'}$, $\kappa_{r''s''}^{(u'')} = \partial\kappa_{r''s''}/\partial\gamma_{u''}$, $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{r\phi}^{(\phi)} = \partial\kappa_{r\phi}/\partial\phi$, etc.

A fórmula geral de Cox & Snell (1968) para o viés de segunda ordem do estimador de máxima verossimilhança do a -ésimo componente do vetor $\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_{k_0})$ é

$$\begin{aligned} B(\hat{\rho}_a) &= \sum_{r',s',u'} \kappa^{ar'} \kappa^{s'u'} \left\{ \kappa_{r's'}^{(u')} - \frac{1}{2} \kappa_{r's'u'} \right\} + \sum_{r'',s'',u''} \kappa^{ar''} \kappa^{s''u''} \left\{ \kappa_{r''s''}^{(u'')} - \frac{1}{2} \kappa_{r''s''u''} \right\} \\ &+ \sum_{r',s'',u'} \kappa^{ar'} \kappa^{s''u'} \left\{ \kappa_{r's''}^{(u')} - \frac{1}{2} \kappa_{r's''u'} \right\} + \sum_{r',s',u''} \kappa^{ar'} \kappa^{s'u''} \left\{ \kappa_{r's'}^{(u'')} - \frac{1}{2} \kappa_{r's'u''} \right\} \\ &+ \sum_{r'',s'',u'} \kappa^{ar''} \kappa^{s''u'} \left\{ \kappa_{r''s''}^{(u')} - \frac{1}{2} \kappa_{r''s''u'} \right\} + \sum_{r'',s',u''} \kappa^{ar''} \kappa^{s'u''} \left\{ \kappa_{r''s'}^{(u'')} - \frac{1}{2} \kappa_{r''s'u''} \right\} \\ &+ \sum_{r',s'',u''} \kappa^{ar'} \kappa^{s''u''} \left\{ \kappa_{r's''}^{(u'')} - \frac{1}{2} \kappa_{r's''u''} \right\} + \sum_{r'',s',u''} \kappa^{ar''} \kappa^{s'u''} \left\{ \kappa_{r''s'}^{(u'')} - \frac{1}{2} \kappa_{r''s'u''} \right\}. \end{aligned}$$

Como as entradas da matriz $K^{\rho\gamma}$ em (3.3.14) não são nulas, todos os termos da expansão anterior devem ser calculados (ver Apêndice A.8). Das quantidades (A.8.22) e (A.8.23) obtidas no Apêndice A.8 e usando notação matricial, temos que o viés de segunda ordem do estimador de máxima verossimilhança de ρ é dado por

$$\begin{aligned} B(\hat{\rho}) &= \frac{1}{2} K^{\rho\rho} V^\top (P_1 \delta_{\rho\rho} + P_2 \delta_{\rho\gamma} + P_3 \delta_{\rho\gamma} + P_3 \delta_{\gamma\gamma}) \\ &+ \frac{1}{2} K^{\rho\gamma} Z^\top (P_2 \delta_{\rho\rho} + P_2 \delta_{\rho\gamma} + P_3 \delta_{\rho\gamma} + P_4 \delta_{\gamma\gamma}), \end{aligned}$$

em que $P_i, i = 1, \dots, 4$, são as matrizes diagonais definidas em (A.8.21) no Apêndice A.8.

Se definimos

$$K^{\rho*} = (K^{\rho\rho} \ K^{\rho\gamma}),$$

o bloco superior de dimensão $k_0 \times (k_0 + k_1)$ obtido da inversa da matriz de informação de Fisher (3.3.14),

4.3. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero e um

podemos escrever o viés de $\hat{\rho}$ como

$$B(\hat{\rho}) = \frac{1}{2} K^{\rho*} \tilde{Z}^T \tilde{\delta}, \quad (4.3.1)$$

em que

$$\tilde{\delta} = \begin{pmatrix} P_1 \delta_{\rho\rho} + P_2 \delta_{\rho\gamma} + P_3 \delta_{\rho\gamma} + P_3 \delta_{\gamma\gamma} \\ P_2 \delta_{\rho\rho} + P_2 \delta_{\rho\gamma} + P_3 \delta_{\rho\gamma} + P_4 \delta_{\gamma\gamma} \end{pmatrix}.$$

Agora, definimos as matrizes \aleph_1 e \aleph_2 de dimensão $2n \times 2n$ como

$$\aleph_1 = \begin{pmatrix} P_1 & P_1 \\ P_3 & P_3 \end{pmatrix} \odot \begin{pmatrix} VK^{\rho\rho}V^T & VK^{\rho\gamma}Z^T \\ ZK^{\gamma\rho}V^T & ZK^{\gamma\gamma}Z^T \end{pmatrix}$$

e

$$\aleph_2 = \begin{pmatrix} P_2 & P_2 \\ P_3 & P_4 \end{pmatrix} \odot \begin{pmatrix} VK^{\rho\rho}V^T & VK^{\rho\gamma}Z^T \\ ZK^{\gamma\rho}V^T & ZK^{\gamma\gamma}Z^T \end{pmatrix},$$

em que $A \odot B$ representa o produto de Hadamard, definido por $[A \odot B]_{ij} = [A]_{ij}[B]_{ij}$. Definimos ainda a matriz

$$\aleph = \begin{pmatrix} \aleph_1 & 0_{2n} \\ 0_{2n} & \aleph_2 \end{pmatrix},$$

onde 0_{2n} representa uma matriz de zeros de dimensão $2n \times 2n$, e

$$\tilde{J} = I_2 \otimes (1, 1)^T \otimes I_n = \begin{pmatrix} I_n & 0_n \\ I_n & 0_n \\ 0_n & I_n \\ 0_n & I_n \end{pmatrix},$$

em que I_n representa a matriz identidade de dimensão $n \times n$, \otimes representa o produto de Kronecker e 0_n representa uma matriz de zeros de dimensão $n \times n$. Podemos mostrar que $\tilde{\delta}$ é um vetor de dimensão $2n \times 1$ que contém a diagonal principal da matriz $\tilde{J}^T \aleph \tilde{J}$, isto é,

$$\tilde{\delta} = \text{diagonal}(\tilde{J}^T \aleph \tilde{J}).$$

Agora, o viés de segunda ordem do estimador de máxima verossimilhança de γ é obtido de forma análoga ao do estimador de ρ , a saber:

$$\begin{aligned} B(\hat{\gamma}) &= \frac{1}{2} K^{\gamma\rho} V^T (P_1 \delta_{\gamma\gamma} + P_2 \delta_{\rho\gamma} + P_3 \delta_{\rho\gamma} + P_3 \delta_{\gamma\gamma}) \\ &\quad + \frac{1}{2} K^{\rho\gamma} Z^T (P_2 \delta_{\rho\rho} + P_2 \delta_{\rho\gamma} + P_3 \delta_{\rho\gamma} + P_4 \delta_{\gamma\gamma}). \\ &= \frac{1}{2} K^{\gamma*} \tilde{Z}^T \tilde{\delta}, \end{aligned} \quad (4.3.2)$$

4.3. Correção analítica dos vieses dos estimadores de máxima verossimilhança para o modelo de regressão beta inflacionado em zero e um

em que

$$K^{\gamma^*} = (K^{\rho\rho} \ K^{\rho\gamma}),$$

é o bloco inferior de dimensão $k_1 \times (k_0 + k_1)$ da inversa da matriz de informação de Fisher (3.3.14). Assim, considerando o vetor de parâmetros $\Upsilon = (\rho^\top, \gamma^\top)^\top$ obtemos uma expressão para o viés de segunda ordem do estimador de máxima verossimilhança de Υ , a qual é dada por

$$B(\hat{\Upsilon}) = \frac{1}{2} K(\Upsilon)^{-1} \tilde{Z}^\top \tilde{\delta}, \quad (4.3.3)$$

em que

$$K(\Upsilon)^{-1} = \begin{pmatrix} K^{\rho\rho} & K^{\rho\gamma} \\ K^{\gamma\rho} & K^{\gamma\gamma} \end{pmatrix},$$

sendo que os elementos desta matriz são os apresentados em (3.3.14). Alternativamente, a expressão (4.3.3) pode ser escrita como

$$B(\hat{\Upsilon}) = \frac{1}{2} (\tilde{Z}^\top \tilde{Q} \tilde{Z})^{-1} \tilde{Z}^\top \tilde{\delta},$$

sendo \tilde{Q} a matriz de pesos definida em (3.2.16). Agora, se definimos $\tilde{\varphi} = \frac{1}{2} \tilde{Q}^{-1} \tilde{\delta}$ é possível escrever o viés de segunda ordem de $\hat{\Upsilon}$ por

$$B(\hat{\Upsilon}) = (\tilde{Z}^\top \tilde{Q} \tilde{Z})^{-1} \tilde{Z}^\top \tilde{Q} \tilde{\varphi}. \quad (4.3.4)$$

Desta forma, os coeficientes do vetor $B(\hat{\Upsilon})$ podem ser estimados a partir da regressão de mínimos quadrados generalizados dada pela equação anterior. Com este resultado, definimos uma estimativa de máxima verossimilhança corrigida $\tilde{\Upsilon}$ da forma

$$\tilde{\Upsilon} = \hat{\Upsilon} - \hat{B}(\hat{\Upsilon}), \quad (4.3.5)$$

onde $\hat{B}(\hat{\Upsilon})$ denota o estimador de máxima verossimilhança de $B(\hat{\Upsilon})$, i.e., os parâmetros desconhecidos são substituídos por seus respectivos estimadores de máxima verossimilhança. Na verdade, isto é a correção de viés em um modelo trinomial.

Finalmente, a fórmula de Cox & Snell (1968) para o viés de segunda ordem do estimador de máxima verossimilhança de β é

$$B(\hat{\beta}) = K^{\beta\beta} X^\top (\Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2) X K^{\beta\beta} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi}) \\ + K^{\beta\phi} (\text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4)) X K^{\beta\phi}), \quad (4.3.6)$$

em que $\Delta = \text{diag}\{1 - \delta_{01} - \delta_{11}, \dots, 1 - \delta_{0n} - \delta_{1n}\}$ e as matrizes W 's são as mesmas obtidas no Apêndice A.7.

Analogamente, temos que o viés de segunda ordem do estimador de máxima verossimilhança de ϕ é

$$B(\hat{\phi}) = K^{\phi\beta} X^\top (\Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2) X K^{\beta\beta} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi}) \\ + K^{\phi\phi} (\text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4)) X K^{\beta\phi}). \quad (4.3.7)$$

4.4. Avaliação numérica

Dado que no modelo de regressão beta inflacionado em zero ou um definido em (3.2.1) compartilha a mesma distribuição beta para modelar o componente contínuo que no modelo de regressão beta inflacionado em zero e um, era de se esperar que as expressões (4.3.6) e (4.3.7) fossem semelhantes às expressões (4.2.5) e (4.2.8) respectivamente. Note que nas expressões a diferença está na forma da matriz Δ .

Conseqüentemente, o viés de segunda ordem do estimador de máxima verossimilhança de $\vartheta = (\beta^\top, \phi)^\top$ é

$$B(\hat{\vartheta}) = K(\vartheta)^{-1} \tilde{X}^\top \tilde{\delta} = (\tilde{X}^\top \tilde{W} \tilde{X})^{-1} \tilde{X}^\top \tilde{\delta},$$

em que $\tilde{\delta}$ é o vetor auxiliar definido em (4.2.6) e \tilde{X} , \tilde{W} são as matrizes definidas em (3.2.16) e (3.2.2) respectivamente.

Note que

$$B(\hat{\vartheta}) = (\tilde{X}^\top \tilde{W} \tilde{X})^{-1} \tilde{X}^\top \tilde{W} \tilde{\xi}, \quad (4.3.8)$$

onde $\tilde{\xi} = \tilde{W}^{-1} \tilde{\delta}$. Desta forma, os coeficientes do vetor $B(\hat{\vartheta})$ podem ser estimados a partir de uma regressão de mínimos quadrados generalizados na variável auxiliar $\tilde{\xi}$. Assim, podemos considerar o estimador de máxima verossimilhança corrigido $\tilde{\vartheta}$ da forma

$$\tilde{\vartheta} = \hat{\vartheta} - \hat{B}(\hat{\vartheta}), \quad (4.3.9)$$

em que $\hat{B}(\hat{\vartheta})$ denota o estimador de máxima verossimilhança de $B(\hat{\vartheta})$, i.e., os parâmetros desconhecidos são substituídos por seus respectivos estimadores de máxima verossimilhança.

4.4 Avaliação numérica

Através de simulações de Monte Carlo avaliamos os desempenhos dos estimadores de máxima verossimilhança dos parâmetros dos modelos de regressão beta inflacionados e suas versões corrigidas em amostras de tamanho finito. Inicialmente consideramos um modelo de regressão beta inflacionado no ponto zero (RBIZ) com estrutura

$$\begin{aligned} h(\alpha_t) &= \gamma_0 + \gamma_1 z_t, \\ g(\mu_t) &= \beta_0 + \beta_1 x_t, \end{aligned} \quad (4.4.1)$$

$t = 1, \dots, n$, i.e., $y_t \sim \text{BIZ}(\alpha_t, \mu_t, \phi)$. Aqui h e g são funções de ligação logito e os valores das variáveis explicativas z_t e x_t para cada observação são conhecidos. Definimos valores verdadeiros dos parâmetros como $\gamma_0 = -0.5$, $\gamma_1 = 1.5$, $\beta_0 = 0.5$, $\beta_1 = 1.8$ e $\phi = 120$; os valores das covariáveis z_t e x_t são selecionados independentemente de uma variável aleatória uniforme $\mathcal{U}(0, 1)$. Para este experimento consideramos tamanhos de amostra $n = 30, 60$ e 90 , o número de réplicas de Monte Carlo é igual a 5000. Para cada réplica de Monte Carlo ajustamos o modelo de regressão (4.4.1), i.e., através de maximização da função de log-verossimilhança são obtidas as estimativas $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1)$, $\hat{\vartheta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\phi})$ e $\hat{\sigma} = 1/(\hat{\phi} + 1)$. Adicionalmente, através da fórmula de Cox & Snell (1968) calculamos os estimadores de máxima verossimilhança corrigidos $\tilde{\gamma} = (\tilde{\gamma}_0, \tilde{\gamma}_1)$, $\tilde{\vartheta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\phi})$ e $\tilde{\sigma}$ definidos por (4.2.3), (4.2.10) e (4.2.16), respectivamente.

No experimento de Monte Carlo utilizamos como gerador de números pseudo-aleatórios o “multiply-with-carry” (GM), cujo período é de 2^{60} (Marsaglia, 1997). A função de log-verossimilhança é maximizada através

4.4. Avaliação numérica

Tabela 4.2: Resultados de simulação, $\gamma_0 = -0.5, \gamma_1 = 1.5$

n	Estimador	Média	Viés	Viés Rel.	Erro Padrão	\sqrt{EQM}
30	$\hat{\gamma}_0$	-0.5617	-0.0617	0.1235	0.7470	0.7496
	$\tilde{\gamma}_0$	-0.5514	-0.0514	0.1028	0.7358	0.7376
60	$\hat{\gamma}_0$	-0.5340	-0.0340	0.0680	0.5597	0.5607
	$\tilde{\gamma}_0$	-0.5285	-0.0285	0.0571	0.5549	0.5556
90	$\hat{\gamma}_0$	-0.5113	-0.0113	0.0227	0.4002	0.4003
	$\tilde{\gamma}_0$	-0.5079	-0.0079	0.0158	0.3980	0.3980
30	$\hat{\gamma}_1$	1.6786	0.1786	0.1191	1.6391	1.6488
	$\tilde{\gamma}_1$	1.6472	0.1472	0.0981	1.6126	1.6193
60	$\hat{\gamma}_1$	1.5878	0.0878	0.0585	0.9723	0.9763
	$\tilde{\gamma}_1$	1.5732	0.0732	0.0488	0.9644	0.9671
90	$\hat{\gamma}_1$	1.5394	0.0394	0.0263	0.7432	0.7442
	$\tilde{\gamma}_1$	1.5295	0.0295	0.0197	0.7391	0.7397

do método BFGS (com derivadas analíticas), que, em geral, é o método que apresenta o melhor desempenho (Mittelhammer, Judge & Miller, 2000, p.199). Todo o procedimento de cálculo foi programado na linguagem de programação matricial Ox (Cribari-Neto & Zarkos, 2003; Doornik, 2001). Para a análise de resultados de estimação pontual calculamos para cada tamanho de amostra: a média, a estimativa do viés, o viés relativo, o desvio padrão e a raiz do erro quadrático médio das 5000 estimativas.

Observamos nas Tabelas 4.2 e 4.3 que as estimativas do viés, viés relativo, desvio padrão e erro quadrático diminuem com o aumento do tamanho de amostra, como era de se esperar. Na Tabela 4.2 notamos que os vieses dos estimadores de máxima verossimilhança dos parâmetros de regressão que modelam o componente discreto são ligeiramente acentuados para tamanhos de amostra pequeno. Por exemplo, para $n = 30$, o viés relativo estimado de $\hat{\gamma}_0$ é 0.1235. Observamos que a correção analítica melhora o desempenho dos estimadores de máxima verossimilhança quando o tamanho de amostra é pequeno, sendo que, em módulo, o viés e viés relativo dos estimadores corrigidos diminuem ligeiramente ao serem comparados com seus respectivos estimadores de máxima verossimilhança. Por exemplo, quando $n = 30$, o viés relativo de $\tilde{\gamma}_0$ é 0.1028.

Os estimadores de máxima verossimilhança e suas versões corrigidas analiticamente apresentam desempenhos similares em termos do erro padrão e da raiz do erro quadrático médio. Percebemos que, em geral, os estimadores corrigidos apresentam um ganho de precisão pouco significativo em relação aos estimadores de máxima verossimilhança dos parâmetros que modelam o componente discreto, mesmo em amostras de tamanho pequeno.

Na Tabela 4.3 observamos os resultados referentes aos estimadores de máxima verossimilhança dos parâmetros de regressão do componente contínuo e suas versões corrigidas. Podemos observar que o desempenho destes estimadores é similar considerando os diferentes tamanhos de amostra. Por exemplo, a

4.4. Avaliação numérica

média estimada de $\hat{\beta}_0$ difere da média estimada de $\tilde{\beta}_0$ na terceira cifra decimal. Já, os erros padrão e a $\sqrt{\text{EQM}}$ são também muito similares; por exemplo, para $n = 60$ os erros padrão de $\hat{\beta}_1$ e $\tilde{\beta}_1$ são 0.1756 e 0.1748, respectivamente. Note ainda que os vieses e vieses relativos estão próximos de zero.

Em geral, a correção analítica dos estimadores de máxima verossimilhança dos parâmetros de regressão do componente contínuo é efetiva. No entanto, o ganho de precisão pela correção não é grande em relação aos estimadores de máxima verossimilhança. Isto nos indica que os estimadores de máxima verossimilhança dos parâmetros de regressão do componente contínuo apresentam boas propriedades amostrais, i.e., os valores estimados dos parâmetros de regressão se aproximam dos verdadeiros valores dos parâmetros de regressão mesmo quando o tamanho de amostra é relativamente pequeno. Estes resultados são similares aos obtidos por Ospina, Cribari-Neto & Vasconcellos (2006) no modelo de regressão beta.

Na Tabela 4.3 observamos que os estimadores de máxima verossimilhança de ϕ apresentam vieses muito altos considerando os diferentes tamanhos de amostra. Por exemplo, para $n = 30$ o viés relativo estimado de $\hat{\phi}$ é 0.3526 enquanto que o viés relativo estimado de $\tilde{\phi}$ é -0.0720 . Desta forma, a correção por viés neste caso se torna muito importante, pois, se a estimativa do parâmetro ϕ é muito alta comparada com o verdadeiro valor do parâmetro, a variância da variável resposta pode estar sendo sub-estimada, o que em eventuais testes de hipóteses poderia levar o investigador a rejeitar erroneamente a hipótese de que os parâmetros da regressão são nulos, quando na realidade eles não são estatisticamente significativos. Notamos que a correção melhora o desempenho do estimador $\hat{\phi}$ considerando todos os tamanhos de amostra. Por exemplo, para $n = 30$ os erros padrão de $\hat{\phi}$ e $\tilde{\phi}$ são 75.5 e 67.7, respectivamente. Adicionalmente, o estimador corrigido $\tilde{\phi}$ apresenta as menores estimativas do desvio padrão e da raiz do erro quadrático médio considerando todos os tamanhos de amostra, o que leva a um estimador para ϕ mais concentrado e praticamente não-viesado. Portanto, para o parâmetro de precisão ϕ recomendamos a correção analítica de viés do estimador $\hat{\phi}$.

Finalmente, podemos observar que o estimador de máxima verossimilhança e sua versão corrigida do parâmetro $\sigma = 1/(\phi + 1)$ têm vieses relativos bem menores do que os respectivos estimadores de ϕ . Por exemplo, para $n = 60$, $\hat{\phi}$ e $\tilde{\phi}$ têm vieses relativos iguais a 0.1395 e 0.0071 respectivamente, enquanto que os vieses relativos de $\hat{\sigma}$ e $\tilde{\sigma}$ são respectivamente iguais a -0.0683 e -0.0199 . É, portanto, mais conveniente para o modelo de regressão beta inflacionado em zero ou um a modelagem do parâmetro de dispersão σ em vez de utilizar o parâmetro de precisão ϕ pois o estimador de máxima verossimilhança de σ é praticamente não-viesado mesmo em amostras de tamanho moderado.

Agora consideramos um modelo de regressão beta inflacionado em zero e um (RBIZU) com estrutura

$$\begin{aligned}\log(\delta_{0t}/(1 - \delta_{0t} - \delta_{1t})) &= \rho_0 + \rho_1 v_t, \\ \log(\delta_{1t}/(1 - \delta_{0t} - \delta_{1t})) &= \gamma_0 + \gamma_1 z_t, \\ g(\mu_t) &= \beta_0 + \beta_1 x_t,\end{aligned}\tag{4.4.2}$$

$t = 1, \dots, n$, i.e., $y_t \sim \text{BIZU}(\delta_{0t}, \delta_{1t}, \mu_t, \phi)$. Aqui, g é a função de ligação logito e os valores das variáveis explicativas v_t , z_t e x_t para cada observação são conhecidos. Usamos os seguintes valores verdadeiros dos parâmetros como $\rho_0 = -1.3$, $\rho_1 = 1.5$, $\gamma_0 = -1.3$, $\gamma_1 = 1.5$, $\beta_0 = 1.5$, $\beta_1 = 1.8$ e $\phi = 120$.

4.4. Avaliação numérica

Tabela 4.3: Resultados de simulação, $\beta_0 = 0.5, \beta_1 = 1.8$ e $\phi = 120$.

n	Estimador	Média	Viés	Viés Relativo	Erro Padrão	$\sqrt{\text{EQM}}$
30	$\hat{\beta}_0$	0.5004	0.0004	0.0009	0.1308	0.1308
	$\tilde{\beta}_0$	0.4999	-0.0001	-0.0001	0.1541	0.1541
60	$\hat{\beta}_0$	0.4984	-0.0016	-0.0030	0.0821	0.0821
	$\tilde{\beta}_0$	0.4976	-0.0024	-0.0046	0.0818	0.0818
90	$\hat{\beta}_0$	0.4996	-0.0004	-0.0007	0.0678	0.0678
	$\tilde{\beta}_0$	0.4992	-0.0008	-0.0015	0.0677	0.0677
30	$\hat{\beta}_1$	1.8043	0.0043	0.0024	0.2530	0.2531
	$\tilde{\beta}_1$	1.7787	-0.0213	-0.0118	1.0770	1.0772
60	$\hat{\beta}_1$	1.8060	0.0060	0.0033	0.1756	0.1757
	$\tilde{\beta}_1$	1.7995	-0.0005	-0.0002	0.1748	0.1748
90	$\hat{\beta}_1$	1.8024	0.0024	0.0013	0.1432	0.1433
	$\tilde{\beta}_1$	1.7984	-0.0016	-0.0008	0.1429	0.1429
30	$\hat{\phi}$	162.3195	42.3195	0.3526	75.5299	86.5712
	$\tilde{\phi}$	111.3515	-8.6485	-0.0720	67.7163	68.2596
60	$\hat{\phi}$	136.7394	16.7394	0.1395	37.6991	41.2450
	$\tilde{\phi}$	120.8568	0.8568	0.0071	33.6138	33.6214
90	$\hat{\phi}$	128.9504	8.9504	0.0745	25.9963	27.4915
	$\tilde{\phi}$	122.5856	2.5856	0.0215	24.6842	24.8168
30	$\hat{\sigma}$	0.0071	-0.0010	-0.1409	0.0027	0.0029
	$\tilde{\sigma}$	0.0085	0.0002	0.0285	0.0074	0.0074
60	$\hat{\sigma}$	0.0077	-0.0005	-0.0683	0.0019	0.0020
	$\tilde{\sigma}$	0.0081	-0.0001	-0.0199	0.0021	0.0021
90	$\hat{\sigma}$	0.0079	-0.0003	-0.0441	0.0015	0.0015
	$\tilde{\sigma}$	0.0081	-0.0001	-0.0199	0.0015	0.0015

Os valores das covariáveis v_t, z_t e x_t são selecionados independentemente de uma variável aleatória uniforme $\mathcal{U}(0, 1)$. Para este experimento consideramos tamanhos de amostra $n = 50, 70$ e 100 , e o número de réplicas de Monte Carlo foi fixado em 5000. Através de maximização da função de log-verossimilhança são obtidas as estimativas $\hat{\Upsilon} = (\hat{\rho}_0, \hat{\rho}_1, \hat{\gamma}_0, \hat{\gamma}_1)$ e $\hat{\vartheta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\phi})$ dos parâmetros dos componentes discreto e contínuo do modelo (4.4.2), respectivamente, assim como o estimador $\hat{\sigma} = 1/(\hat{\phi} + 1)$. Calculamos as estimativas de máxima verossimilhança corrigidas $\tilde{\Upsilon} = (\tilde{\rho}_0, \tilde{\rho}_1, \tilde{\gamma}_0, \tilde{\gamma}_1)$, $\tilde{\vartheta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\phi})$ definidas pelas equações (4.3.5) e (4.3.9), respectivamente, assim como a estimativa corrigida $\tilde{\sigma}$. Novamente, utilizamos como gerador de números pseudo-aleatórios o “multiply-with-carry” (GM) e a função de log-verossimilhança é maximizada através do método BFGS (com derivadas analíticas). Todo o procedimento de cálculo foi programado na linguagem de programação matricial Ox.

4.5. Conclusões

Para a análise de resultados de estimação pontual calculamos para cada tamanho de amostra: a média, a estimativa do viés, o viés relativo, o desvio padrão e a raiz do erro quadrático médio das 5000 estimativas.

Na Tabela 4.4 observamos que as estimativas do viés, viés relativo, desvio padrão e raiz do erro quadrático médio dos estimadores de máxima verossimilhança dos parâmetros de regressão do componente discreto e de suas respectivas versões corrigidas diminuem com o aumento do tamanho de amostra. Notamos que os vieses dos estimadores de máxima verossimilhança dos parâmetros de regressão que modelam o componente discreto são pouco acentuados para tamanhos de amostra pequenos. Por exemplo, para $n = 50$, o viés relativo estimado de $\hat{\gamma}_0$, em módulo, é de 0.1054. Observamos que, em praticamente todos os casos, a correção é efetiva e melhora o desempenho dos estimadores de máxima verossimilhança para os diferentes tamanhos de amostra. Neste sentido, vemos que as medidas do desvio padrão e da raiz do erro quadrático médio são muito similares. Por exemplo, para $n = 70$ temos que a raiz do erro quadrático médio de $\hat{\gamma}_0$ é 0.5147 enquanto que a de $\tilde{\gamma}_0$ é 0.5132, diferindo apenas na terceira cifra decimal.

Em geral, temos que os estimadores corrigidos apresentam um ganho de precisão pouco significativo em relação aos estimadores de máxima verossimilhança, no que se refere aos parâmetros que modelam o componente discreto.

Na Tabela 4.5 observamos que os vieses dos estimadores de máxima verossimilhança dos parâmetros de regressão do componente contínuo e de suas versões corrigidas estão próximos de zero. De forma análoga ao modelo de regressão beta inflacionado em zero, a correção dos estimadores de máxima verossimilhança dos parâmetros de regressão do componente contínuo é efetiva, considerando todos os tamanhos de amostra. No entanto, o ganho de precisão pela correção não é muito grande em relação aos estimadores de máxima verossimilhança. Nesta mesma tabela temos que os estimadores de máxima verossimilhança de ϕ apresentam vieses muito altos considerando os diferentes tamanhos de amostra. Por exemplo, para $n = 50$ o viés relativo estimado de $\hat{\phi}$ é 19.8055. Notamos que a correção analítica melhora muito o desempenho do estimador $\hat{\phi}$ considerando todos os tamanhos de amostra. De forma semelhante ao modelo de regressão beta inflacionado em zero, podemos concluir que para o parâmetro de precisão ϕ recomendamos a correção do estimador de máxima verossimilhança. Em relação ao parâmetro σ , os resultados da simulação apresentados na Tabela 4.5 indicam que o estimador de máxima verossimilhança $\hat{\sigma}$ é praticamente não-viesado considerando os diferentes tamanhos de amostra. Já o estimador corrigido $\tilde{\sigma}$ apresenta desempenho semelhante ao estimador $\hat{\sigma}$. Por exemplo, para $n = 50$ o erros padrão de $\hat{\sigma}$ e $\tilde{\sigma}$ são 0.0021 e 0.0022, respectivamente.

4.5 Conclusões

Neste capítulo derivamos expressões que permitem corrigir o viés de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros dos modelos de regressão beta inflacionados. Mostramos que os vieses de segunda ordem obtidos através da fórmula de Cox & Snell (1968) podem ser escritos em termos de regressões de mínimos quadrados generalizados, o que facilita o cálculo. As expressões encontradas nos permitem construir versões modificadas de maior precisão com respeito aos estimadores de máxima verossimilhança.

4.5. Conclusões

Tabela 4.4: Resultados de simulação, $\rho_0 = -1.3$, $\rho_1 = 1.5$, $\gamma_0 = -1.3$ e $\gamma_1 = 1.5$.

n	Estimador	Média	Viés	Viés Relativo	Erro Padrão	$\sqrt{\text{EQM}}$
50	$\hat{\rho}_0$	-1.4676	-0.1676	0.1289	0.8334	0.8500
	$\tilde{\rho}_0$	-1.4614	-0.1614	0.1241	0.8361	0.8515
70	$\hat{\rho}_0$	-1.4099	-0.1099	0.0845	0.5403	0.5514
	$\tilde{\rho}_0$	-1.4055	-0.1055	0.0812	0.5379	0.5481
100	$\hat{\rho}_0$	-1.3680	-0.0680	0.0523	0.3939	0.3997
	$\tilde{\rho}_0$	-1.3671	-0.0671	0.0516	0.3936	0.3992
50	$\hat{\rho}_1$	1.7111	0.2111	0.1407	0.7994	0.8267
	$\tilde{\rho}_1$	1.7066	0.2066	0.1377	0.7949	0.8213
70	$\hat{\rho}_1$	1.6365	0.1365	0.0910	0.6051	0.6203
	$\tilde{\rho}_1$	1.6323	0.1323	0.0882	0.6024	0.6167
100	$\hat{\rho}_1$	1.5876	0.0876	0.0584	0.4210	0.4300
	$\tilde{\rho}_1$	1.5868	0.0868	0.0579	0.4206	0.4294
50	$\hat{\gamma}_0$	-1.4370	-0.1370	0.1054	0.5785	0.5944
	$\tilde{\gamma}_0$	-1.4320	-0.1320	0.1015	0.5757	0.5906
70	$\hat{\gamma}_0$	-1.4163	-0.1163	0.0894	0.5014	0.5147
	$\tilde{\gamma}_0$	-1.4134	-0.1134	0.0872	0.5003	0.5132
100	$\hat{\gamma}_0$	-1.3631	-0.0631	0.0485	0.3824	0.3875
	$\tilde{\gamma}_0$	-1.3626	-0.0626	0.0481	0.3822	0.3873
50	$\hat{\gamma}_1$	1.6663	0.1663	0.1108	0.6237	0.6454
	$\tilde{\gamma}_1$	1.6621	0.1621	0.1081	0.6213	0.6421
70	$\hat{\gamma}_1$	1.6503	0.1503	0.1002	0.5726	0.5920
	$\tilde{\gamma}_1$	1.6474	0.1474	0.0983	0.5712	0.5899
100	$\hat{\gamma}_1$	1.5791	0.0791	0.0527	0.3909	0.3988
	$\tilde{\gamma}_1$	1.5787	0.0787	0.0525	0.3907	0.3986

Os resultados de simulação mostram que as correções dos estimadores de máxima verossimilhança dos parâmetros de regressão que modelam os componentes discretos e contínuos dos modelos de regressão beta inflacionados são efetivas na redução do viés, mas não apresenta fortes melhorias em desempenho comparada aos estimadores de máxima verossimilhança. No caso do parâmetro de precisão observamos que as estimativas de máxima verossimilhança se apresentam muito viesadas, e desta forma, recomendamos a sua correção. A correção de viés dos estimadores dos parâmetros que envolvem o componente contínuo do modelo estende os resultados de correção de viés obtidos por Ospina, Cribari-Neto & Vasconcellos (2006) para a classe de modelos de regressão beta inflacionados. No que se refere ao parâmetro σ observamos que o estimador de máxima verossimilhança é praticamente não-viesado e, desta forma, recomendamos sua utilização.

4.5. Conclusões

Tabela 4.5: Resultados de simulação, $\beta_0 = 1.5$, $\beta_1 = 1.8$ e $\phi = 120$.

n	Estimador	Média	Viés	Viés Relativo	Erro Padrão	\sqrt{EQM}
50	$\hat{\beta}_0$	1.5038	0.0038	0.0025	0.0608	0.0609
	$\tilde{\beta}_0$	1.4774	-0.0226	-0.0150	0.9536	0.9537
70	$\hat{\beta}_0$	1.5019	0.0019	0.0012	0.0463	0.0464
	$\tilde{\beta}_0$	1.5003	0.0003	0.0002	0.0462	0.0462
100	$\hat{\beta}_0$	1.5015	0.0015	0.0010	0.0424	0.0424
	$\tilde{\beta}_0$	1.5003	0.0003	0.0002	0.0423	0.0423
50	$\hat{\beta}_1$	1.8032	0.0032	0.0018	0.0750	0.0750
	$\tilde{\beta}_1$	1.7696	-0.0304	-0.0168	1.1558	1.1561
70	$\hat{\beta}_1$	1.8033	0.0033	0.0018	0.0581	0.0582
	$\tilde{\beta}_1$	1.8010	0.0010	0.0005	0.0579	0.0580
100	$\hat{\beta}_1$	1.8016	0.0016	0.0000	0.0484	0.0484
	$\tilde{\beta}_1$	1.8002	0.0002	0.0001	0.0483	0.0483
50	$\hat{\phi}$	139.8055	19.8055	0.1650	42.2420	46.6507
	$\tilde{\phi}$	122.5055	2.5055	0.0208	37.5170	37.5969
70	$\hat{\phi}$	134.3498	14.3498	0.1195	33.2971	36.2545
	$\tilde{\phi}$	122.8278	2.8278	0.0235	30.4231	30.5512
100	$\hat{\phi}$	130.0616	10.0616	0.0838	27.9305	29.6849
	$\tilde{\phi}$	122.1699	2.1699	0.0180	26.2738	26.3607
50	$\hat{\sigma}$	0.0076	-0.0006	-0.0726	0.0021	0.0021
	$\tilde{\sigma}$	0.0080	-0.0002	-0.0242	0.0022	0.0022
70	$\hat{\sigma}$	0.0078	-0.0004	-0.0484	0.0018	0.0018
	$\tilde{\sigma}$	0.0080	-0.0002	-0.0242	0.0018	0.0018
100	$\hat{\sigma}$	0.0079	-0.0003	-0.0373	0.0016	0.0016
	$\tilde{\sigma}$	0.0081	-0.0001	-0.0121	0.0016	0.0016

Capítulo 5

Diagnóstico em modelos de regressão beta inflacionados

5.1 Introdução

Uma etapa importante na análise de um modelo ajustado é a verificação de possíveis afastamentos das suposições assumidas. Nesta direção é importante detectar a presença de pontos extremos no conjunto de dados e avaliar seu impacto nos resultados inferenciais. Para tanto, uma análise de resíduos e, principalmente, os gráficos de resíduos podem ajudar na avaliação da estabilidade e robustez de resultados inferenciais, visto que os resíduos medem a discrepância entre o modelo ajustado e o conjunto de dados. Assim, é conveniente procurar uma definição para o resíduo que leve em consideração a contribuição que cada observação exerce sobre a adequação do modelo. A análise de resíduos num modelo estatístico pode ser baseada nos resíduos ordinários, em versões padronizadas, em resíduos construídos a partir dos componentes da função desvio (McCullagh & Nelder, 1989) ou em resíduos generalizados (Cox & Snell, 1968). Inúmeras pesquisas têm sido publicadas abordando o estudo dos resíduos visando, na definição, comportamento distribucional e correção. Nesta direção destacamos os trabalhos de Pregibon (1981), Copas (1988), McCullagh (1987, Cap. 6), Williams (1984, 1987), Davison & Gigli (1989), Pierce & Schafer (1986), Fahmeir & Tutz (1994, Cap. 4), Paula (1995), de Souza & Paula (2002), Dunn & Smyth (1996), Ferrari & Cribari-Neto (2004) e Espinheira, Ferrari & Cribari-Neto (2008a), entre outros. Técnicas gráficas usando os resíduos do modelo ajustado são freqüentemente adotadas na validação do modelo estatístico; por exemplo, Atkinson (1981) sugere que a construção de bandas de confiança (envelopes) simuladas pode ajudar a interpretar melhor o gráfico de probabilidade normal dos resíduos, de tal forma que, se o modelo estiver bem ajustado, a maioria dos pontos que representam os resíduos deve estar distribuída aleatoriamente dentro dessas bandas.

Outro aspecto importante na análise de diagnóstico é a detecção de observações influentes, i.e., pontos que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. Cook (1986) propõe avaliar a influência de pontos através do estudo da curvatura do deslocamento pela verossimilhança (“likelihood displacement”). No entanto, a medida proposta por Cook (1986) apresenta certa dificuldade de interpretação pois pode assumir qualquer valor positivo. Poon & Poon (1999) propõem uma nova medida da curvatura que é de fácil interpretação e possui boas propriedades geométricas. Essa medida de influência parece ser apropriada para avaliar os modelos de regressão beta inflacionados.

5.2. Resíduos

Este capítulo se encontra organizado da seguinte forma. Na Seção 5.2 definimos resíduos padronizados e resíduos quantis aleatorizados para os modelos de regressão beta inflacionados em zero e/ou um. Na Seção 5.3 desenvolvemos a análise de influência local baseada na curvatura normal conforme (Poon & Poon, 1999) para os modelos de regressão beta inflacionados. Para o modelo de regressão beta inflacionado em zero ou um consideramos dois esquemas de perturbação, a saber: a ponderação de casos e a perturbação individual de covariáveis. Dada a complexidade do modelo de regressão beta inflacionado em zero e um, consideramos unicamente como esquema de perturbação a ponderação de casos.

5.2 Resíduos

Podemos definir como resíduo uma medida que objetiva identificar discrepâncias entre o modelo ajustado e os dados. Assim, é compreensível que a maioria dos resíduos esteja baseada na diferença $y_t - \widehat{E}(y_t)$. No entanto, respeitado o formato da distribuição de probabilidade da variável resposta, é mais interessante pensar no resíduo como uma função $r(y_t, \widehat{E}(y_t))$, definição geral de resíduos proposta por Cox & Snell (1968). Sob este ponto de vista propomos resíduos para os modelos de regressão beta inflacionados.

5.2.1 Resíduos para o modelo de regressão beta inflacionado em zero ou um

Resíduos padronizados

Para os modelos de regressão beta, Espinheira, Ferrari & Cribari-Neto (2008a) sugerem utilizar resíduos padronizados obtidos da convergência do processo iterativo score de Fisher para estimação dos parâmetros de regressão. Neste sentido, estendemos tal idéia para o modelo de regressão beta inflacionado em zero ou um.

Para ϕ conhecido, obtemos da convergência do processo iterativo score de Fisher para γ e β as expressões (A.9.5) e (A.9.6) (Apêndice A.9). Desta forma, podemos definir os resíduos padronizados para o modelo de regressão beta inflacionado em zero ou um com as quantidades avaliadas nos estimadores de máxima verossimilhança por:

$$r_{pt}^{(1)} = \frac{\mathbb{1}_{\{c\}}(y_t) - \widehat{\alpha}_t}{\sqrt{\widehat{\alpha}_t(1 - \widehat{\alpha}_t)(1 - \widehat{h}_{1tt}^*)}} \quad (5.2.1)$$

e

$$r_{pt}^{(2)} = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{\widehat{w}_t(1 - \widehat{\alpha}_t)(1 - \widehat{h}_{2tt}^*)}}, \quad (5.2.2)$$

para $t = 1, \dots, n$, em que y_t^* e μ_t^* foram definidos em (3.2.7) e (3.2.8), respectivamente. Os componentes \widehat{h}_{1tt}^* e \widehat{h}_{2tt}^* são os t -ésimos elementos da diagonal principal das matrizes de projeção

$$\widehat{H}_1^* = \widehat{Q}^{1/2} Z (Z^\top \widehat{Q} Z)^{-1} Z^\top \widehat{Q}^{1/2}$$

5.2. Resíduos

e

$$\widehat{H}_2^* = \widehat{W}_{\beta\beta}^{1/2} X (X^\top \widehat{W}_{\beta\beta} X)^{-1} X^\top \widehat{W}_{\beta\beta}^{1/2},$$

respectivamente. Aqui, Z é uma matriz de valores fixos conhecidos ($n \times M$) cuja t -ésima linha é $z_t^\top = (z_{t1}, \dots, z_{tM})$, X é uma matriz de valores fixos conhecidos ($n \times m$) com t -ésima linha $x_t^\top = (x_{t1}, \dots, x_{tm})$ e as matrizes Q e $W_{\beta\beta}$ foram definidas em (3.2.14).

Por construção, $r_{pt}^{(1)}$ e $r_{pt}^{(2)}$ são, respectivamente, os resíduos padronizados do componente discreto e contínuo do modelo de regressão beta inflacionado em zero ou um. Em geral, os resíduos ponderados têm distribuição empírica assimétrica dificultando a utilização de medidas usuais de diagnóstico. Contudo, gráficos de $r_{pt}^{(1)}$ e $r_{pt}^{(2)}$ contra os valores ajustados $\widehat{\alpha}_t$ e $\widehat{\mu}_t$, respectivamente, podem revelar observações que são marginalmente aberrantes em cada sub-modelo (discreto ou contínuo).

Para o modelo de regressão beta inflacionado em zero ou um, podemos definir através dos resíduos do componente discreto e contínuo do modelo (equações (5.2.1) e (A.10.3)) dois tipos de resíduos que incorporam a característica discreta e contínua da distribuição da variável resposta. Baseados nas idéias do resíduo de Cox & Snell e do resíduo componente do desvio para dados censurados (Cox & Snell, 1968; Klein & Moeschberger, 1997, Cap. 11), definimos um resíduo padronizado na forma

$$r_{pt} = \begin{cases} r_{pt}^{(1)}, & \text{se } y_t = c, \\ r_{pt}^{(2)}, & \text{se } y_t \in (0, 1). \end{cases} \quad (5.2.3)$$

Podemos observar que a distribuição do resíduo padronizado é assimétrica devido à presença de massa de probabilidade em $y = c$.

Adicionalmente, definimos o resíduo ponderado

$$r_{pt}^* = \widehat{\alpha}_t r_{pt}^{(1)} + (1 - \widehat{\alpha}_t) r_{pt}^{(2)} \quad (5.2.4)$$

em que $c = 0$ ou $c = 1$ dependendo do caso. Note que r_{pt}^* é a média ponderada entre o resíduo $r_{pt}^{(1)}$ e o resíduo $r_{pt}^{(2)}$. Para estes resíduos, os gráficos de r_{pt} e r_{pt}^* contra os valores ajustados $\widehat{E}(y_t)$ podem ser informativos com respeito ao posicionamento de pontos aberrantes e influentes no conjunto de dados.

Resíduos quantis aleatorizados

No caso em que a resposta (y) é uma variável aleatória contínua, podemos construir um resíduo a partir de uma transformação da distribuição de probabilidade $F(y, \theta)$ de y . Inicialmente, definimos a variável aleatória $U = F(y, \theta)$, uniformemente distribuída no intervalo unitário.* Se $\Phi(\cdot)$ denota a função de distribuição

* Note que se $U = F(y, \theta)$, então $P(U \leq u) = P(F(y, \theta) \leq u) = P(y \leq F^{-1}(u)) = F(F^{-1}(u)) = u$ para $0 < u < 1$. Logo, U tem distribuição $\mathcal{U}(0, 1)$.

5.2. Resíduos

acumulada da normal padrão, temos que

$$V = \Phi^{-1}(F(y, \theta)) = \Phi^{-1}(U)$$

tem distribuição normal padrão. Assumindo θ conhecido, temos o resíduo de Cox & Snell (Cox & Snell, 1968). No entanto, na prática θ é desconhecido, devendo ser substituído por sua estimativa de máxima verossimilhança. Como $\hat{\theta}$ é um estimador consistente de θ temos a normalidade assintótica de V , i.e., $V \xrightarrow{D} \mathcal{N}(0, 1)$ e, assim, $E(V) \approx 0$ e $\text{Var}(V) \approx 1$.

Se F não é contínua, como é o caso da distribuição beta inflacionada em zero ou um definida em (2.3.1), é preciso dar uma definição mais geral do resíduo V . Neste caso, podemos optar por uma versão aleatorizada do resíduo de Cox & Snell e definir o resíduo quantil aleatorizado (Dunn & Smyth, 1996) para o modelo de regressão beta inflacionado em zero ou um como

$$r_t^q = \Phi^{-1}(u_t), \quad t = 1, \dots, n, \quad (5.2.5)$$

em que u_t é uma variável aleatória uniforme no intervalo $(a_t, b_t]$, sendo $a_t = \lim_{y \uparrow y_t} \text{BI}_c(y; \hat{\alpha}, \hat{\mu}, \hat{\phi})$ e $b_t = \text{BI}_c(y_t; \hat{\alpha}, \hat{\mu}, \hat{\phi})$, respectivamente. Por exemplo, para o modelo de regressão beta inflacionado em zero, temos que para $y_t = 0$ (ponto de massa em zero), $a_t = \lim_{y \uparrow 0} \text{BIZ}(y; \hat{\alpha}, \hat{\mu}, \hat{\phi}) = (1 - \hat{\alpha}) \lim_{y \uparrow 0} F(y; \hat{\mu}, \hat{\phi}) = 0$ e $b_t = \text{BIZ}(0; \hat{\alpha}, \hat{\mu}, \hat{\phi}) = \hat{\alpha}$. Logo, u_t é uma variável uniforme no intervalo $(0, \hat{\alpha}]$. Agora para $y_t \in (0, 1)$ temos que $u_t = \text{BIZ}(y_t; \hat{\alpha}, \hat{\mu}, \hat{\phi})$. Se consideramos o modelo de regressão beta inflacionado em um, temos para $y_t = 1$ (ponto de massa em um) que $a_t = \lim_{y \uparrow 1} \text{BIU}(y; \hat{\alpha}, \hat{\mu}, \hat{\phi}) = (1 - \hat{\alpha}) \lim_{y \uparrow 1} F(y; \hat{\mu}, \hat{\phi}) = (1 - \hat{\alpha})$ e $b_t = \text{BIU}(1; \hat{\alpha}, \hat{\mu}, \hat{\phi}) = 1$. Logo, u_t é uma variável uniforme no intervalo $(1 - \hat{\alpha}, 1]$. Quando $y_t \in (0, 1)$ temos que $u_t = \text{BIU}(y_t; \hat{\alpha}, \hat{\mu}, \hat{\phi})$.

O procedimento de aleatorização é introduzido com a finalidade de produzir resíduos normais contínuos. Como o resíduo quantil aleatorizado pode variar de uma realização a outra, em situações práticas, é aconselhável fazer pelo menos quatro realizações destes resíduos para detectar padrões no seu comportamento. O resíduo quantil aleatorizado constitui uma opção de resíduo quando a resposta é discreta e assume poucos valores distintos (Dunn & Smyth, 1996).

5.2.2 Resíduos para o modelo de regressão beta inflacionado em zero e um

Resíduos padronizados

Nesta subseção estendemos a idéia dos resíduos padronizados do modelo de regressão beta inflacionado em zero ou um. Por razões de simplicidade vamos construir resíduos padronizados para o modelo de regressão

5.2. Resíduos

beta inflacionado em zero e um (RBIZU) com a estrutura

$$\begin{aligned} g(\mu_t) &= \eta_t, \\ \log(\delta_{0t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{0t}, \\ \log(\delta_{1t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{1t}, \end{aligned}$$

$t = 1, \dots, n$, em que o parâmetro de precisão ϕ é constante para todas as observações. De forma análoga ao modelo de regressão beta inflacionado em zero ou um consideramos que ϕ é conhecido. Do processo de estimação por máxima verossimilhança temos da convergência do processo iterativo score de Fisher para ρ , γ e β as expressões (A.10.2) e (A.9.6) (ver Apêndices A.10 e A.9, respectivamente). Desta forma, podemos definir os resíduos padronizados para o componente discreto do modelo de regressão beta inflacionado em zero e um como:

$$r_{p_t}^{\{0\}} = \frac{\mathbb{1}_{\{0\}}(y_t) - \delta_{0t}}{\sqrt{\hat{q}_{1tt}(1 - \hat{h}_{tt}^{\{0\}})}} \quad (5.2.6)$$

e

$$r_{p_t}^{\{1\}} = \frac{\mathbb{1}_{\{1\}}(y_t) - \delta_{1t}}{\sqrt{\hat{q}_{2tt}(1 - \hat{h}_{tt}^{\{1\}})}} \quad (5.2.7)$$

em que q_{1tt} e q_{2tt} são, respectivamente, os t -ésimos elementos da diagonal principal das matrizes Q_1 e Q_2 definidas em (A.10) e $h^{\{0\}}_{tt}$, $h^{\{1\}}_{tt}$ são os t -ésimos elementos da diagonal principal das matrizes de projeção $H^{\{0\}} = \Psi_0^\top H_d^* \Psi_0$ e $H^{\{1\}} = \Psi_1^\top H_d^* \Psi_1$, respectivamente, em que $\Psi_0 = (I_n, 0_n)^\top$ e $\Psi_1 = (0_n, I_n)^\top$ são matrizes de dimensão $(2n \times n)$, I_n é a matriz identidade de ordem n e $H_d^* = Q^{1/2} \tilde{Z} (\tilde{Z}^\top Q \tilde{Z})^{-1} \tilde{Z}^\top Q^{1/2}$ é matriz de projeção. Aqui, as quantidades são avaliadas no estimador de máxima verossimilhança.

Note que o resíduo $r_{p_t}^{\{0\}}$ corresponde ao resíduo padronizado do sub-modelo do componente discreto que modela a probabilidade de ocorrência de zeros e $r_{p_t}^{\{1\}}$ corresponde ao resíduo padronizado do sub-modelo do componente discreto que modela a probabilidade de ocorrência de uns. Desta forma, parece razoável sugerir dois conjuntos de gráficos de resíduos para o componente discreto deste modelo. O primeiro é de $r_{p_t}^{\{0\}}$ contra $\hat{\delta}_{0t}$ e o segundo de $r_{p_t}^{\{1\}}$ contra $\hat{\delta}_{1t}$ para $t = 1, \dots, n$. Neste caso, é de se esperar que os gráficos revelem pontos marginalmente aberrantes para cada sub-modelo.

Para o modelo RBIZU a distribuição beta (2.4.7) que modela o componente contínuo é a mesma distribuição que modela o componente contínuo do modelo de regressão beta inflacionado em zero ou um. Desta forma, podemos construir o mesmo resíduo padronizado para o componente contínuo do modelo RBIZU. Neste caso,

$$r_{p_t}^{(c)} = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{\hat{w}_t(1 - \hat{\alpha}_t)(1 - \hat{h}_{ctt}^*)}} \quad (5.2.8)$$

5.3. Influência Local

$t = 1, \dots, n$, em que y_t^* e μ_t^* foram definidos em (3.2.7) e (3.2.8) respectivamente. Aqui, $\hat{\alpha}_t = \hat{\delta}_{0t} + \hat{\delta}_{1t}$, $\hat{w}_t = \hat{\psi}'(\hat{\mu}_t \hat{\phi}) + \hat{\psi}'((1 - \hat{\mu}_t) \hat{\phi})$, $\hat{h}_{c_{tt}}^*$ é o t -ésimo elemento da diagonal principal da matriz de projeção $\hat{H}_c^* = \hat{W}_{\beta\beta}^{1/2} X (X^\top \hat{W}_{\beta\beta} X)^{-1} X^\top \hat{W}_{\beta\beta}^{1/2}$, em que X é uma matriz de valores fixos conhecidos ($n \times m$) com t -ésima linha $x_t^\top = (x_{t1}, \dots, x_{tm})$ e $W_{\beta\beta}$ foi definida em (3.2.14). De forma análoga ao modelo de regressão beta inflacionado em zero ou um, gráficos de $r_{pt}^{(2)}$ contra $\hat{\mu}_t$ podem revelar observações que são marginalmente aberrantes no sub-modelo constituído pelo componente contínuo do modelo RBIZU.

Para o modelo de regressão beta inflacionado em zero e um, podemos definir através de (5.2.6), (5.2.7) e (5.2.8) o resíduo padronizado

$$r_{pt} = \begin{cases} r_{pt}^{\{0\}} & \text{se } y_t = 0, \\ r_{pt}^{\{1\}} & \text{se } y_t = 1, \\ r_{pt}^{(c)} & \text{se } y_t \in (0, 1). \end{cases} \quad (5.2.9)$$

Para estes resíduos, o gráfico de r_{pt} contra os valores ajustados $\widehat{E}(y_t)$ pode ajudar a detectar globalmente pontos aberrantes e/ou influentes.

Adicionalmente, definimos o resíduo ponderado

$$r_{pt}^* = \hat{\delta}_{0t} r_{pt}^{\{0\}} + \hat{\delta}_{1t} r_{pt}^{\{1\}} + (1 - \hat{\delta}_{0t} - \hat{\delta}_{1t}) r_{pt}^{(c)}. \quad (5.2.10)$$

Note que r_{pt}^* é a média ponderada entre os resíduos $r_{pt}^{\{0\}}$, $r_{pt}^{\{1\}}$ e $r_{pt}^{(2)}$. Para este resíduo o gráfico de r_{pt}^* contra os valores ajustados $\widehat{E}(y_t)$ pode ser informativo com respeito ao posicionamento de pontos aberrantes e influentes no conjunto de dados.

Resíduos quantis aleatorizados

Para o modelo de regressão beta inflacionado em zero e um consideramos o resíduo quantil aleatorizado

$$r_t^q = \Phi^{-1}(u_t), \quad t = 1, \dots, n, \quad (5.2.11)$$

em que u_t é uma variável aleatória uniforme no intervalo $(a_t, b_t]$, sendo $a_t = \lim_{y \uparrow y_t} \text{BIZU}(y; \alpha_t, \gamma_t, \mu_t, \phi)$ e $b_t = \text{BIZU}(y_t; \alpha_t, \gamma_t, \mu_t, \phi)$, respectivamente. Aqui, $\text{BIZU}(y; \alpha_t, \gamma_t, \mu_t, \phi)$ foi definida em (2.4.1) sendo $\gamma_t = \delta_{1t}/\alpha_t$ e $\alpha_t = \delta_{0t} + \delta_{1t}$. O gráfico de r_t^q contra o índice das observações ou contra os valores ajustados e os valores de variáveis explicativas pode revelar observações com alta influência no seu próprio valor predito ou alguma tendência sistemática, embora a sua interpretação não seja a mesma dada para os modelos de regressão normal linear.

5.3 Influência Local

A influência local é uma metodologia de diagnóstico que permite avaliar mudanças nos resultados da análise inferencial quando pequenas perturbações são incorporadas ao modelo e/ou aos dados de estudo.

5.3. Influência Local

Se essas perturbações causarem efeitos desproporcionais, pode haver indício de que o modelo está mal ajustado ou que pode existir algum tipo de afastamento nas suposições feitas para o mesmo. Neste trabalho, consideramos a técnica de diagnóstico de influência local desenvolvida por Poon & Poon (1999).

Naturalmente existem diversos trabalhos sobre influência local e suas extensões. Por exemplo, Galea, Paula & Bolfarine (1997), Liu (2000) e Galea, Paula & Uribe-Opazo (2003) apresentam estudos de influência local em modelos elípticos lineares, Liu (2002) apresenta um estudo da influência local em modelos elípticos multivariados. Kwan & Fung (1998) aplicam a metodologia em análise fatorial, Paula (1996) em modelos próprios de dispersão, Ortega, Bolfarine & Paula (2003) em modelos log-gama generalizados com dados censurados, Jansen, et al. (2003) utilizam influência local em um estudo psiquiátrico com dados binários e Shi & Wang (1999) em regressão “ridge”. Na classe de erros normais, Lawrence (1988) investiga a influência local em modelos lineares com parâmetros na transformação da variável resposta. Tsai & Wu (1992) investigam influência local em modelos auto-regressivos de primeira ordem e modelos heteroscedásticos. Paula (1993) aplica a influência local em modelos lineares com restrições nos parâmetros na forma de desigualdades lineares. Zhu & Lee (2001) estudam a influência local em modelos para dados incompletos e, mais recentemente, Xiang et. al (2005) utilizam a análise de influência local em um modelo de mistura de Poisson.

Especificamente, o conceito de influência local está baseado na curvatura da superfície da função de log-verossimilhança. Seja θ o vetor de parâmetros de interesse. Consideremos a influência de uma pequena perturbação ω sobre o parâmetro estimado. Desta forma, denotamos por $\ell(\theta|\omega)$ a função de log-verossimilhança do modelo perturbado. Segundo a proposta de Cook (1986), o efeito de ω sobre θ pode ser avaliado através do deslocamento pela verossimilhança $F(\omega) = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\omega)\}$, que compara o estimador de máxima verossimilhança restrito $\hat{\theta}_\omega$ (estimador sob o modelo perturbado) com respeito aos contornos da função de log-verossimilhança do modelo não perturbado. No caso em que ω_0 é o ponto de perturbação nula temos que $\ell(\theta|\omega_0) = \ell(\theta)$. Esta metodologia investiga a direção do maior afastamento na superfície $F(\omega)$ no ponto ω_0 , que corresponde à perturbação de casos localmente influentes na estimativa $\hat{\theta}$.

A direção da maior mudança em $F(\omega_0)$ pode ser obtida através do cálculo da curvatura normal na direção d da superfície do deslocamento pela verossimilhança, i.e., $C_d(\theta) = |d^\top \ddot{F} d|$, sendo d um vetor unitário e $\ddot{F} = \partial^2 F(\omega) / \partial \omega \partial \omega^\top |_{\omega=\omega_0}$. No entanto, é possível mostrar que

$$C_d(\theta) = 2|d^\top \Delta^\top \ddot{\ell}^{-1} \Delta d|, \quad (5.3.1)$$

onde $\Delta = \partial^2 \ell(\theta|\omega) / \partial \theta \partial \omega^\top |_{\theta=\hat{\theta}, \omega=\omega_0}$ e $\ddot{\ell} = \partial^2 \ell(\theta) / \partial \theta \partial \theta^\top |_{\theta=\hat{\theta}}$. Se denotamos por d_{\max} a direção da máxima curvatura normal,

$$C_{\max} = \max_d \{C_d(\theta)\},$$

então d_{\max} indica a direção em que se produzem as maiores mudanças locais no deslocamento pela verossimilhança em $F(\omega_0)$, sendo que esta quantidade por ser obtida pelo autovetor correspondente ao maior autovalor da matriz $\ddot{F} = -\Delta^\top \ddot{\ell}^{-1} \Delta$.

5.3. Influência Local

É possível também avaliar a influência local apenas para uma parte do vetor de parâmetros. Suponha que possamos particionar o vetor de parâmetros como $\theta = (\theta_1^\top, \theta_2^\top)^\top$, tal que

$$\ddot{\ell} = \begin{pmatrix} \ddot{\ell}_{\theta_1\theta_1} & \ddot{\ell}_{\theta_1\theta_2} \\ \ddot{\ell}_{\theta_2\theta_1} & \ddot{\ell}_{\theta_2\theta_2} \end{pmatrix},$$

em que $\ddot{\ell}_{\theta_1\theta_1} = \partial^2 \ell(\theta) / \partial \theta_1 \partial \theta_1^\top$, $\ddot{\ell}_{\theta_1\theta_2} = \partial^2 \ell(\theta) / \partial \theta_1 \partial \theta_2^\top$, $\ddot{\ell}_{\theta_2\theta_1} = \partial^2 \ell(\theta) / \partial \theta_2 \partial \theta_1^\top$ e $\ddot{\ell}_{\theta_2\theta_2} = \partial^2 \ell(\theta) / \partial \theta_2 \partial \theta_2^\top$. Então, a curvatura normal na direção do vetor d é dada por

$$C_d(\theta_1) = |d^\top \Delta^\top (\ddot{\ell}^{-1} - \ddot{\ell}_{22}) \Delta d|,$$

em que

$$\ddot{\ell}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\ell}_{\theta_2\theta_2}^{-1} \end{pmatrix}.$$

Neste caso, o gráfico de índices do maior autovetor de $-\Delta^\top (\ddot{\ell}^{-1} - \ddot{\ell}_{22}) \Delta$ pode revelar quais são as observações que estão influenciado a θ_2 .

Uma outra possibilidade consiste em avaliar a curvatura na direção do t -ésimo indivíduo. Neste caso, a curvatura normal é dada por

$$C_t = 2|\Delta_t^\top \ddot{\ell}^{-1} \Delta_t|,$$

em que Δ_t é a t -ésima coluna da matriz Δ . Uma vez que C_t reflete a situação em que foi atribuído o maior valor (o valor total) possível à t -ésima coordenada de d , tal que $\|d\| = 1$ tal curvatura é donotada por nfluência local total do t -ésimo indivíduo (Lesaffre & Verbeke, 1998). Desta forma, é útil considerar em uma análise de influência o estudo dos componentes de d_{\max} e adicionalmente a medida de influência total referente ao t -ésimo indivíduo, C_t . Assim, gráficos dos componentes de d_{\max} contra os índices das observações podem sugerir quais observações são conjuntamente influentes, enquanto os gráficos de C_t contra os índices das observações destacam casos individualmente influentes.

Embora o método de Cook, baseado na curvatura normal, seja de grande utilidade, ele apresenta alguns inconvenientes. Por exemplo, a curvatura normal pode assumir valores nos reais positivos e não é invariante sob reparametrizações de escala. Com o objetivo de contornar estas dificuldades e considerando a superfície do deslocamento pela verossimilhança $F(\omega)$, Poon & Poon (1999) definem a curvatura normal conforme no ponto ω_0 na direção d como

$$B_d(\theta) = \frac{|d^\top \ddot{F} d|}{\|\ddot{F}^2\|_{\mathcal{F}}} = \frac{2|d^\top \Delta^\top \ddot{\ell}^{-1} \Delta d|}{\|2\Delta^\top \ddot{\ell}^{-1} \Delta\|_{\mathcal{F}}}, \quad (5.3.2)$$

em que $\|\cdot\|_{\mathcal{F}}$ denota a norma de Fröbenius definida como $\|A\|_{\mathcal{F}} = \sqrt{\text{tr}(A^\top A)}$, sendo A uma matriz $m \times n$. Note que o cálculo da expressão (5.3.2) não requer mais esforço que o de $C_d(\theta)$. Pode-se mostrar que, para qualquer direção d , $B_d(\theta)$ satisfaz a condição $0 \leq B_d(\theta) \leq 1$, e desta forma, $B_d(\theta)$ é uma medida normalizada, i.e., $B_d(\theta)$ é uma medida geométrica da curvatura que assume valores no intervalo $[0, 1]$, o que torna

5.3. Influência Local

mais fácil a interpretação de seu valor. Adicionalmente, esta medida é invariante sob uma reparametrização conforme, i.e., uma reparametrização cuja matriz jacobiana M é tal que $M^\top M$ é um múltiplo positivo da matriz identidade. Aqui, a reparametrização pode ser vista como uma “modificação do esquema de perturbação” (Cook, 1986). Se a reparametrização é conforme em ω_0 , então a curvatura normal conforme em qualquer direção é invariante sob a reparametrização. Um exemplo de reparametrização conforme é $\phi(\omega) = M\omega + c$, tal que M é uma matriz conforme[†].

Se $\lambda_1, \dots, \lambda_r$ são autovalores da matriz \ddot{F} com correspondentes autovetores normalizados e_1, \dots, e_r , temos que o valor da curvatura normal conforme $B_{e_i}(\theta)$ é igual ao autovalor normalizado $\hat{\lambda}_i$, em que $\hat{\lambda}_i = \lambda_i / \sqrt{\sum_{i=1}^r \lambda_i^2}$. Logo, $\sum_{i=1}^r B_{e_i}^2(\theta) = 1$. Assim, se a curvatura normal conforme é a mesma para todos os autovetores, então o valor comum é $1/\sqrt{r}$ e este valor pode ser usado como referência para decidir se um autovetor é influente ou não (Poon & Poon, 1999). Anteriormente, vimos que a curvatura normal C_d e a direção d são usadas para avaliar a influência local, principalmente ao examinar o autovetor d_{\max} com a curvatura normal C_{\max} . Quando d tem a direção do autovetor d_{\max} correspondente ao maior autovalor, a curvatura normal conforme $B_{d_{\max}}$ assume seu valor máximo. Assim, a curvatura normal e a curvatura normal conforme são medidas de diagnóstico equivalentes, diferindo apenas por um fator positivo. Poon & Poon (1999) sugerem a utilização de um valor de referência para comparar os efeitos de B_{e_i} e $B_{d_{\max}}$ em vários níveis. Para tanto, é necessária a seguinte definição.

Definição 5.3.1. Um autovetor e_i será q -**influyente** se $|B_{e_i}(\theta)| \geq q/\sqrt{r}$.

Ou seja, escolhendo o valor de $q \in \mathbb{R}$ (por praticidade, $q = 1, \dots, r$) temos que uma direção é influente de acordo com a magnitude de seu corresponde autovalor normalizado. Como consequência direta da definição 5.3.1 temos que um autovetor q -**influyente** é também $(q-1)$ -**influyente**, $(q-2)$ -**influyente**, \dots , 1 -**influyente**.

Seja E_j o j -ésimo vetor da base canônica $\Theta = \{E_1, \dots, E_r\}$ de \mathbb{R}^r . Poon & Poon (1999) definem E_j como o vetor de perturbação básico do espaço de perturbações, cujo j -ésimo elemento é 1 e todos os restantes são iguais a 0. Assim, podemos escrever cada um dos autovetores de perturbação e_i como uma combinação linear dos elementos da base Θ , i.e., $e_i = \sum_{j=1}^r a_{ij} E_j$, e, como $\|e_i\| = 1$, temos que $\sum_{j=1}^r a_{ij}^2 = 1$. Daí, para cada i fixo, se a contribuição de todos os a_{ij} é a mesma, então $|a_{ij}| = 1/\sqrt{r}$ e este valor pode ser usado como um valor de referência para julgar a magnitude da curvatura de cada autovetor da matriz (5.3.2).

De forma geral, a análise dos vetores de perturbação básicos pode ser feita para todos os autovetores influentes. Se tomamos o módulo dos autovalores normalizados, $\lambda_i^* = |\hat{\lambda}_i|$, de forma que

$$\lambda_{\max} = \lambda_i^* \geq \dots \geq \lambda_k^* \geq q/\sqrt{r} > \lambda_{k+1}^* \dots \lambda_r^* \geq 0,$$

e denotamos por a_{ij} a j -ésima coordenada do autovetor correspondente a λ_i^* , podemos obter a contribuição

[†]Uma matriz M é conforme se $M^\top M = \tau I$, em que I é a matriz identidade e τ é um escalar que assume valores nos reais positivos.

5.3. Influência Local

agregada do j -ésimo vetor básico para todos os autovetores q -influentes em \mathbb{R}^r ao tomar

$$m[q]_j = \sqrt{\sum_{i=1}^k \lambda_i^* a_{ij}^2}. \quad (5.3.3)$$

Desta forma, para interpretar a contribuição agregada, podemos construir um gráfico dos valores de $m[q]_j$ contra os índices $j = 1, \dots, r$ e verificar se existem pontos que se distanciam dos demais.

Dado que $\sum_{j=1}^r m[q]_j^2 = \sum_i^k \lambda_i^*$, temos que a contribuição de todos os vetores de perturbação básicos é a mesma, então cada uma é igual a $\bar{m}[q] = \sqrt{1/r \sum_{i=1}^k \lambda_i^*}$. Desta forma, para determinar a significância da contribuição individual dos vetores de perturbação básicos usamos $\bar{m}[q]$. Neste ponto, devemos considerar dois casos. O primeiro é permitir que q seja suficientemente grande para agregar somente o maior autovalor λ_{\max} . Neste caso $\bar{m}[q]_j = \sqrt{\lambda_{\max}} |a_{1j}|$ e este método é equivalente a comparar $|a_{1j}|$ com $1/\sqrt{r}$. De outro lado, se admitimos que $q = 0$, então todos os autovalores serão incluídos na análise. Aqui, $m[0]_j$ é dita ser a contribuição total, e escrevemos $m_j = \sqrt{\sum_{i=1}^r \lambda_i^* a_{ij}^2}$. Se a contribuição de todas as perturbações básicas for a mesma, então cada uma será igual a

$$\bar{m} = \bar{m}[0] = \sqrt{\frac{\sum_{i=1}^r \lambda_i^*}{r}} = \sqrt{\frac{\sum_{i=1}^r |\lambda_i|}{r \sum_{i=1}^r \lambda_i^2}}.$$

A contribuição total m_j e a curvatura normal conforme $B_{E_j}(\theta)$ do vetor de perturbações básico E_j estão altamente relacionadas. Se todos os autovalores são não-negativos, $B_{E_j}(\theta)$ é igual ao quadrado da contribuição total do j -ésimo vetor de perturbação básico. Ou seja, dado que a matriz $-\ddot{F}$ é semi-definida positiva e todos os autovalores são não negativos, temos que

$$B_{E_j}(\theta) = m_j^2, \quad (5.3.4)$$

para todo j . Agora, se a contribuição total de todos os $B_{E_j}(\theta)$ é a mesma, então cada uma é igual a

$$b = \frac{\text{tr}(\ddot{F})}{r \sqrt{\text{tr}(\ddot{F}^2)}}. \quad (5.3.5)$$

Daí, para decidir se um determinado ponto é influente ou não, considerando a contribuição agregada dos k autovalores (em módulo) que são maiores do que q/\sqrt{r} , é utilizada a contribuição média $\bar{m}[q]_j$, e para decidir se um caso é ou não influente, utilizamos o valor b dado em (5.3.5). Poon & Poon (1999) propõem utilizar $\sqrt{2}\bar{m}[q]$ como ponto de corte para a contribuição agregada dos k autovalores e $2b$ como ponto de corte da contribuição agregada de todos os autovalores. Isto significa que se a contribuição agregada é maior que duas vezes a média, então a observação correspondente é considerada influente.

5.3. Influência Local

Dado que é possível perturbar o modelo proposto de diversas maneiras, é importante escolher perturbações que sejam de fácil interpretação. Por exemplo, podemos perturbar a função de log-verossimilhança ou as variáveis explicativas de maneira coletiva ou individualmente.

5.3.1 Influência local no modelo de regressão beta inflacionado em zero ou um

Para o modelo de regressão beta inflacionado em zero ou um consideramos dois esquemas de perturbação, a saber: ponderação de casos e perturbação individual de covariáveis.

Ponderação de casos

A ponderação de casos tem sido o esquema de perturbação mais amplamente difundido na análise de diagnóstico já que pode ser interpretada como a perturbação na variância do t -ésimo caso. Seja $\omega = (\omega_1, \dots, \omega_n)^\top$ um vetor $n \times 1$ de ponderações. O vetor de perturbação nulo é $\omega_0 = (1, \dots, 1)^\top$. Neste caso, a função de log-verossimilhança perturbada é da forma

$$\ell(\theta|\omega) = \ell_1(\gamma|\omega) + \ell_2(\beta, \phi|\omega), \quad (5.3.6)$$

em que

$$\begin{aligned} \ell_1(\gamma|\omega) &= \sum_{t=1}^n \omega_t \ell_t(\alpha_t), \\ \ell_2(\beta, \phi|\omega) &= \sum_{t:y_t \in (0,1)} \omega_t \ell_t(\mu_t, \phi), \end{aligned} \quad (5.3.7)$$

sendo $\ell_t(\alpha_t)$ e $\ell_t(\mu_t, \phi)$ as funções definidas em (3.2.5). Sendo $\omega_t = 0$, temos que o t -ésimo ponto é eliminado e $\omega_0 = (1, 1, \dots, 1)^\top$ implica que todos os pontos são considerados. Assim, $\ell(\theta|\omega_0) = \ell(\theta) = \ell_1(\gamma) + \ell_2(\beta, \phi)$. Para este esquema de perturbação temos que a curvatura normal conforme é $B_d(\theta) = 2|d^\top \Delta^\top \check{\ell}^{-1} \Delta d| / \|2\Delta^\top \check{\ell}^{-1} \Delta\|_{\mathcal{F}}$, em que $\check{\ell}$ é definido pela expressão (A.11.1) do Apêndice A.11 e $\Delta = (\Delta_\gamma^\top, \Delta_\beta^\top, \Delta_\phi^\top)^\top$ é da forma

$$\Delta = \begin{pmatrix} \Delta_\gamma \\ \Delta_\beta \\ \Delta_\phi \end{pmatrix} = \begin{pmatrix} Z^\top \widehat{P} \widehat{G} \widehat{\mathcal{E}}_1 \\ \widehat{\phi} X^\top H \widehat{T} \widehat{\mathcal{E}}_2 \\ \widehat{u} \end{pmatrix},$$

sendo P, G, H, T as matrizes definidas na Seção 3.2. Já $\widehat{\mathcal{E}}_1, \widehat{\mathcal{E}}_2$ e u foram definidas na expressão (A.11.3) da Subseção A.11.

Suponhamos que apenas são perturbados os casos em que a resposta é contínua ($y \in (0, 1)$). Neste caso, o vetor de ponderações para o modelo não perturbado é tal que $\omega_t = 1$ para $y_t \in (0, 1)$ e 0 caso contrário.

5.3. Influência Local

Então, a expressão (A.11.3) reduz-se à matriz de dimensão $(m + 1) \times n$, da forma

$$\Delta_{(\beta, \phi)} = \begin{pmatrix} \Delta_\beta \\ \Delta_\phi \end{pmatrix} = \begin{pmatrix} \hat{\phi} X^\top H \hat{T} \hat{\mathcal{E}}_2 \\ \hat{u} \end{pmatrix}.$$

De forma análoga, podemos considerar a perturbação dos casos em que a resposta assume unicamente o valor c , i.e., $y = c$. Aqui, o vetor de ponderações para o modelo não perturbado é tal que $\omega_t = 1$ para $y_t = c$, e 0 caso contrário. Desta forma, a expressão (A.11.3) fica reduzida a

$$\Delta_\gamma = \partial^2 \ell(\theta | \omega) / \partial \gamma \partial \omega^\top \Big|_{\theta = \hat{\theta}, \omega = \delta_0} = \partial \ell_1(\gamma) / \partial \gamma \Big|_{\theta = \hat{\theta}} = Z^\top \hat{P} \hat{G} \hat{\mathcal{E}}_1.$$

Perturbação nas variáveis explanatórias

Perturbação individual das covariáveis do componente discreto

Consideremos agora uma perturbação aditiva de uma variável explanatória particular do componente discreto, digamos z_p , $p = 1, \dots, M$, de forma que

$$z_{tp}(\omega) = z_{tp} + \omega_t s_{z_p}.$$

Neste caso, por exemplo, se $p \neq 2$ ou $p \neq M$

$$\zeta_t(\omega) = \gamma_1 + \dots + \gamma_p(z_{tp} + \omega_t s_{z_p}) + \dots + \gamma_M z_{tM} \quad (5.3.8)$$

e $\alpha_t(\omega)$ é tal que $h(\alpha_t(\omega)) = \zeta_t(\omega)$. Avaliando a expressão (A.11.4) da Seção A.11 em $\omega_0 = (0, \dots, 0)^\top$ e $\theta = \hat{\theta}$, temos para $R = 1, \dots, M$, que $\Delta_R = \hat{\gamma}_{p s_{z_p}} \hat{w}_{1t}(\omega) z_{tR}$. Assim, ao utilizar notação matricial obtemos

$$\Delta_\gamma = \hat{\gamma}_{p s_{z_p}} Z^\top \hat{\mathfrak{W}}_1,$$

sendo $\hat{\mathfrak{W}}_1(\omega) = \hat{\mathfrak{W}}_1$, em que $\hat{\mathfrak{W}}_1$ é a matriz \mathfrak{W}_1 definida em A.11 avaliada nos estimadores de máxima verossimilhança. Através deste esquema de perturbação é possível acessar a influência individual de cada covariável no processo de estimação do modelo. Contudo, este tipo de perturbação só faz sentido se a covariável perturbada é medida de forma contínua.

Perturbação individual das covariáveis do componente contínuo

De forma análoga à perturbação individual de covariáveis do componente discreto, podemos sugerir uma perturbação aditiva de uma variável explanatória particular do componente contínuo, digamos x_p , $p = 1, \dots, m$, de forma que

$$x_{tp}(\omega) = x_{tp} + \omega_t s_{x_p}.$$

5.3. Influência Local

Neste caso, por exemplo, se $p \neq 2$ ou $p \neq m$, temos

$$\eta_t(\omega) = \beta_1 + \cdots + \beta_p(x_{tp} + \omega_t s_{x_p}) + \cdots + \beta_m x_{tm} \quad (5.3.9)$$

e $\mu_t(\omega)$ é tal que $g(\mu_t(\omega)) = \eta_t(\omega)$.

Avaliando a expressão (A.11.5) em $\omega_0 = (0, \dots, 0)^\top$ e $\theta = \hat{\theta}$, temos, para $r = 1, \dots, m$ que

$$\Delta_r = \hat{\beta}_{p s_{x_p}} \hat{\mathfrak{w}}_{2t}(\omega) x_{tr} = \hat{\beta}_{p s_{x_p}} \hat{\mathfrak{w}}_{2t} x_{tr},$$

em que $\hat{\mathfrak{w}}_{2t}$ é o t -ésimo elemento da matriz \mathfrak{w}_2 definida na subsecção A.11 avaliada nos estimadores de máxima verossimilhança. Utilizando notação matricial obtemos

$$\Delta_\beta = \hat{\beta}_{p s_{x_p}} X^\top \widehat{\mathfrak{W}}_2, \quad (5.3.10)$$

em que $\widehat{\mathfrak{W}}_2(\omega) = \widehat{\mathfrak{W}}_2$, sendo $\widehat{\mathfrak{W}}_2$ a matriz \mathfrak{W}_2 definida em A.11 avaliada nos estimadores de máxima verossimilhança.

Adicionalmente, utilizando a expressão (A.11.6) obtida na subsecção A.11 do Apêndice A.11 e avaliando-a em $\omega_0 = (0, \dots, 0)^\top$ e $\theta = \hat{\theta}$, podemos escrever matricialmente

$$\Delta_\phi = \hat{\beta}_{p s_{x_p}} \hat{\mathfrak{w}}_3 \hat{T}. \quad (5.3.11)$$

Assim, ao utilizar, (5.3.10) e (5.3.11) temos que Δ assume a forma

$$\Delta = \begin{pmatrix} \hat{\beta}_{p s_{x_p}} X^\top \widehat{\mathfrak{W}}_2 \\ \hat{\beta}_{p s_{x_p}} \hat{\mathfrak{w}}_3 \hat{T} \end{pmatrix}. \quad (5.3.12)$$

Novamente, este tipo de perturbação só faz sentido se a covariável perturbada é medida de forma contínua.

Perturbação individual das covariáveis do componente discreto e contínuo simultaneamente

Para o modelo de regressão beta inflacionado em zero ou um suponhamos que as as matrizes das regressões do componente discreto e contínuo sejam tais que $Z \neq X$. Consideremos uma perturbação aditiva de uma variável explanatória particular do componente discreto e contínuo, digamos z_p , $p = 1, \dots, M$, e x_q , $q = 1, \dots, m$ de forma que

$$\begin{aligned} z_{tp}(\omega) &= z_{tp} + \omega_t s_{z_p}, \\ x_{tq}(\omega) &= x_{tq} + \omega_t s_{x_q}. \end{aligned}$$

Neste caso, por exemplo, se $p, q \neq 2$ ou $p, q \neq M$ temos

$$\begin{aligned} \zeta_t(\omega) &= \gamma_1 + \cdots + \gamma_p(z_{tp} + \omega_t s_{z_p}) + \cdots + \gamma_M z_{tM}, \\ \eta_t(\omega) &= \beta_1 + \cdots + \beta_q(x_{tq} + \omega_t s_{x_q}) + \cdots + \beta_m x_{tm}. \end{aligned} \quad (5.3.13)$$

5.3. Influência Local

Para este caso, temos que a função de log-verossimilhança do modelo perturbado é da forma $\ell(\theta|\omega) = \ell_1(\gamma|\omega) + \ell_2(\beta, \phi|\omega)$, em que $\ell_1(\gamma|\omega) = \sum_{t=1}^n \ell_t(\alpha_t(\omega))$ e $\ell_2(\beta, \phi|\omega) = \sum_{t=1}^n \ell_t(\alpha_t(\omega), \phi)$. Pela separabilidade entre γ e $(\beta^\top, \phi)^\top$, temos que

$$\begin{aligned}\Delta_\gamma &= \partial^2 \ell_1(\gamma|\omega) / \partial \gamma \partial \omega^\top \Big|_{\gamma=\hat{\gamma}, \delta=\omega_0}, \\ \Delta_\beta &= \partial^2 \ell_2(\beta, \phi|\omega) / \partial \beta \partial \omega^\top \Big|_{(\beta^\top, \phi)^\top = (\hat{\beta}^\top, \hat{\phi})^\top, \delta=\omega_0}, \\ \Delta_\phi &= \partial^2 \ell_2(\beta, \phi|\omega) / \partial \phi \partial \omega^\top \Big|_{(\beta^\top, \phi)^\top = (\hat{\beta}^\top, \hat{\phi})^\top, \delta=\omega_0}.\end{aligned}\tag{5.3.14}$$

Avaliando adequadamente as expressões (A.11.4), (A.11.5) e (A.11.6) da subseção A.11 em $\omega_0 = (0, \dots, 0)^\top$ e $\theta = \hat{\theta}$ temos que

$$\Delta = \begin{pmatrix} \hat{\gamma}_p s_{z_p} Z^\top \widehat{\mathfrak{W}}_1 \\ \hat{\beta}_q s_{x_q} X^\top \widehat{\mathfrak{W}}_2 \\ \hat{\beta}_q s_{x_q} \widehat{\mathfrak{w}}_3 \widehat{T} \end{pmatrix},\tag{5.3.15}$$

em que $\widehat{\mathfrak{W}}_1$ e $\widehat{\mathfrak{W}}_2$ são as matrizes $\mathfrak{W}_1, \mathfrak{W}_2$ definidas em A.11 avaliadas nos estimadores de máxima verossimilhança. Note que no caso em que as matrizes de regressões do componente discreto e contínuo são tais que $Z = X$, o cálculo da curvatura $B_d(\theta)$ vem da aplicação da expressão (5.3.15) substituindo Z por X .

5.3.2 Influência local no modelo de regressão beta inflacionado em zero e um

Para o modelo de regressão beta inflacionado em zero e um, consideramos unicamente o esquema de perturbação de ponderação de casos. Por simplicidade e interpretabilidade consideramos o modelo de regressão beta inflacionado em zero e um com estrutura

$$\begin{aligned}g(\mu_t) &= \eta_t, \\ \log(\delta_{0t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{0t}, \\ \log(\delta_{1t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{1t},\end{aligned}$$

$t = 1, \dots, n$, em que o parâmetro de precisão ϕ é constante para todas as observações.

Seja $\omega = (\omega_1, \dots, \omega_n)^\top$ um vetor $n \times 1$ de ponderações. O vetor de perturbação nulo é $\omega_0 = (1, \dots, 1)^\top$. Neste caso, a função de log-verossimilhança perturbada é da forma

$$\ell(\theta|\omega) = \ell_1(\rho, \gamma|\omega) + \ell_2(\beta, \phi|\omega),$$

em que

$$\begin{aligned}\ell_1(\rho, \gamma|\omega) &= \sum_{t=1}^n \omega_t \ell_t(\delta_{0t}, \delta_{1t}), \\ \ell_2(\beta, \phi|\omega) &= \sum_{t:y_t \in (0,1)} \omega_t \ell_t(\mu_t, \phi),\end{aligned}$$

5.3. Influência Local

sendo

$$\begin{aligned}
\ell_t(\delta_{0t}, \delta_{1t}) &= \mathbb{1}_{\{0\}}(y_t) \log \delta_{0t} + \mathbb{1}_{\{1\}}(y_t) \log \delta_{1t} \\
&\quad + (1 - \mathbb{1}_{\{0\}}(y_t) - \mathbb{1}_{\{1\}}(y_t)) \log(1 - \delta_{0t} - \delta_{1t}) \\
\ell_t(\mu_t, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t \\
&\quad + \{(1 - \mu_t)\phi - 1\} \log(1 - y_t).
\end{aligned} \tag{5.3.16}$$

Desta forma, $\ell(\theta|\omega_0) = \ell(\theta) = \ell_1(\rho, \gamma) + \ell_2(\beta, \phi)$. Utilizando as expressões (A.12.3) obtidas na Subseção A.12 do Apêndice A.12 podemos escrever matricialmente a matriz Δ como

$$\Delta = \begin{pmatrix} \Delta_\rho \\ \Delta_\gamma \\ \Delta_\beta \\ \Delta_\phi \end{pmatrix} = \begin{pmatrix} V^\top \widehat{\mathcal{E}}_0 \\ Z^\top \widehat{\mathcal{E}}_1 \\ \widehat{\phi} X^\top \text{mdia}\text{g}(y_{(0,1)}) \widehat{T} \widehat{\mathcal{E}}_2 \\ \widehat{u} \end{pmatrix},$$

sendo T a matrizes definida na Seção 3.2. Já $\widehat{\mathcal{E}}_0$, $\widehat{\mathcal{E}}_1$, $\widehat{\mathcal{E}}_2$ e u foram definidas pelas expressões da Subseção A.12. Para o calculo da curvatura $B_d(\theta)$ devemos diagonalizar a matriz $\Delta^\top \ddot{\ell}^{-1} \Delta$, em que $\ddot{\ell}$ é dada em (A.12.1). Suponhamos que apenas sejam perturbados os casos em que a resposta é contínua ($y \in (0, 1)$). Neste caso, temos que assumir que o vetor de ponderações para o modelo não perturbado é tal que $\omega_t = 1$, para $y_t \in (0, 1)$, e 0 caso contrário. Então, a expressão (A.12.3) reduz-se à matriz de dimensão $(k+1) \times n$ da forma

$$\Delta = \begin{pmatrix} \Delta_\beta \\ \Delta_\phi \end{pmatrix} = \begin{pmatrix} \widehat{\phi} X^\top \text{mdia}\text{g}(y_{(0,1)}) \widehat{T} \widehat{\mathcal{E}}_2 \\ \widehat{u} \end{pmatrix}.$$

Assim, a curvatura normal conforme do sub-modelo que compõe o componente contínuo é

$$B_d^{(0,1)}(\theta) = 2|d^\top \Delta^\top \ddot{\ell}^{-1} \Delta d| / \|2\Delta^\top \ddot{\ell}^{-1} \Delta\|_{\mathcal{F}},$$

De forma análoga, podemos considerar a perturbação dos casos em que a resposta assume unicamente o valor 0, i.e., $y = 0$. Aqui, o vetor de ponderações para o modelo não perturbado é tal que $\omega_t = 1$ para $y_t = 0$ e 0 caso contrário. Desta forma, a expressão (A.12.3) fica reduzida a

$$\Delta = \Delta_\rho = \partial^2 \ell(\theta|\omega) / \partial \rho \partial \omega^\top \Big|_{\theta=\widehat{\theta}, \omega=\omega_0} = \partial \ell_1(\rho, \gamma) / \partial \rho \Big|_{\theta=\widehat{\theta}} = V^\top \widehat{\mathcal{E}}_0.$$

Logo, a curvatura normal conforme do sub-modelo que compõe o componente discreto em zero é

$$B_d^{\{0\}}(\theta) = 2|d^\top \Delta^\top \ddot{\ell}^{-1} \Delta d| / \|2\Delta^\top \ddot{\ell}^{-1} \Delta\|_{\mathcal{F}}.$$

Similarmente, podemos assumir a perturbação dos casos em que a resposta toma o valor 1 unicamente, i.e., $y = 1$. Aqui, o vetor de ponderações para o modelo não perturbado é tal que $\omega_t = 1$, para $y_t = 1$, e 0

5.4. Conclusões

caso contrário. Desta forma, a expressão (A.12.3) fica reduzida a

$$\Delta = \Delta_\gamma = \partial^2 \ell(\theta|\omega) / \partial \gamma \partial \omega^\top \Big|_{\theta=\hat{\theta}, \omega=\omega_0} = \partial \ell_1(\rho, \gamma) / \partial \gamma \Big|_{\theta=\hat{\theta}} = Z^\top \hat{\mathcal{E}}_1.$$

Portanto, a curvatura normal conforme do sub-modelo que compõe o componente discreto em um é

$$B_d^{\{1\}}(\theta) = 2|d^\top \Delta^\top \ddot{\ell}^{-1} \Delta d| / \|2\Delta^\top \ddot{\ell}^{-1} \Delta\|_{\mathcal{F}}.$$

5.4 Conclusões

Neste capítulo propomos resíduos para os modelos de regressão beta inflacionados com base no processo iterativo score de Fisher propomos resíduos padronizados e ponderados. Adicionalmente, através de um processo de aleatorização, definimos resíduos quantis aleatorizados. Além disso, desenvolvemos medidas de influência utilizando um método de influência local que é baseado na curvatura normal conforme, sob diferentes esquemas de perturbação. Para o modelo de regressão beta inflacionado em zero ou um, derivamos expressões matriciais para a curvatura normal conforme sob dois esquemas de perturbação, a saber: a ponderação de casos e a perturbação individual de covariáveis. Finalmente, para o modelo de regressão beta inflacionado em zero e um derivamos a medida de curvatura normal conforme sob o esquema de ponderação de casos.

Capítulo 6

Aplicação

Este capítulo tem como objetivo ilustrar a metodologia desenvolvida nesta tese acerca de inferência, correção de viés e análise de diagnóstico nos modelos de regressão beta inflacionados. Consideramos um conjunto de dados referente à difusão de televisão a cabo nos EUA. As observações correspondem a 282 comunidades, que são essencialmente áreas de franquia individual com assinantes de televisão a cabo. Os dados foram coletados pela Comissão Federal de Comunicações (FCC) dos EUA em conjunção à implementação da ata de competição e proteção ao consumidor de televisão a cabo de 1992. Os dados são descritos em detalhe no Apêndice E de FCC 93-177 (Federal Communications Commission, 1993). A definição das variáveis da amostra original com informação econômica e demográfica se encontra disponível em Federal Communications Commission (1994).

Para nosso estudo consideramos as seguintes variáveis: o logaritmo da renda média na franquia* (*lin*) dada em milhares de dólares, a porcentagem de crianças na franquia (*child*), o número de canais com sinal local (*ltv*) e a idade em anos do sistema de televisão a cabo (*agehe*). O interesse recai em modelar y que representa a proporção de assinantes de televisão a cabo que adquirem serviços adicionais. Inicialmente foi realizada uma análise descritiva das variáveis envolvidas no problema. Na Tabela 6.1 estão apresentadas algumas medidas resumo. Observando esta tabela notamos que a variável resposta varia de zero a 0.92. Notamos também que a distribuição da variável resposta é nitidamente assimétrica. Em termos das covariáveis observamos que, em média, a renda das comunidades é de aproximadamente 10 mil dólares, a proporção de crianças é cerca de 36%, o número de canais com sinal local é seis e a idade do sistema de televisão a cabo encontra-se ao redor dos 12 anos.

Na Figura 6.1 apresentamos o histograma de frequência da variável resposta. A barra com o ponto acima representa a quantidade de zeros de y , correspondente a 22% das observações. Neste gráfico observamos que a distribuição dos dados em $(0,1)$ é assimétrica e com uma moda e há uma concentração de observações em zero. Uma simples inspeção visual indica que uma distribuição beta inflacionada em zero pode ser um modelo adequado para a variável resposta. Adicionalmente, o gráfico de box-plot de y na Figura 6.1 revela a presença de três valores extremos, a saber: as observações 18, 28 e 64.

*As franquias correspondem a comunidades servidas por televisão a cabo.

Tabela 6.1: Medidas resumo das variáveis

Medida	y	lin	$child$	ltv	$agehe$
Mínimo	0.00	9.12	1.53	0.00	1.00
Q_1	0.05	9.92	31.24	4.00	7.00
Mediana	0.22	10.14	35.92	6.00	11.00
Média	0.23	10.14	36.30	6.35	12.14
Q_3	0.35	10.33	40.75	8.00	16.00
Máximo	0.92	11.56	88.49	16.00	41.00
Desvio padrão	0.20	0.36	9.58	3.21	7.58

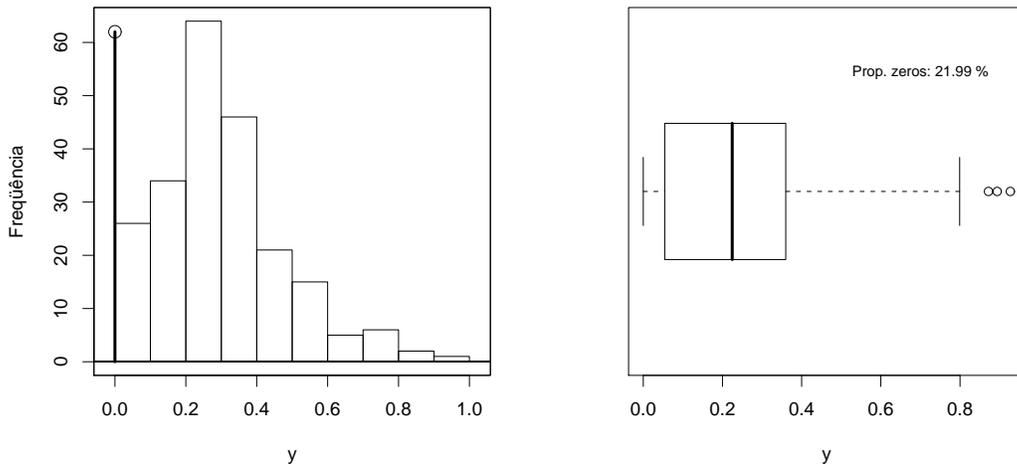


Figura 6.1: Histograma de frequência e gráfico box-plot para a proporção de assinantes de televisão a cabo que adquirem serviços adicionais.

Para este conjunto de dados assumimos um modelo de regressão beta inflacionado em zero, em que $y_t \sim \text{BIZ}(\alpha_t, \mu_t, \phi)$ são variáveis aleatórias independentes, tais que

$$\begin{aligned} \text{logit}(\alpha) &= \gamma_0 + \gamma_1 \text{lin} + \gamma_2 \text{child} + \gamma_3 \text{agehe} + \gamma_4 \text{ltv}, \\ \text{logit}(\mu) &= \beta_0 + \beta_1 \text{lin} + \beta_2 \text{child} + \beta_3 \text{agehe} + \beta_4 \text{ltv}; \end{aligned}$$

aqui suprimimos o índice t e consideramos $\phi = \exp(\nu)$. Para a estimação dos parâmetros é utilizada má-

xima verossimilhança. A função de log-verossimilhança do modelo é maximizada iterativamente usando os algoritmos RS ou CG (Rigby & Stasinopoulos, 2005). Aqui, utilizamos a implementação deste modelo proposta por Ospina (2006) no pacote `gamlss` do R. Na Tabela 6.2 apresentamos as estimativas por máxima verossimilhança com seus respectivos erros padrão.

Tabela 6.2: Estimativas de máxima verossimilhança com erros padrão.

Estimador	Estimativa	Erro Padrão
$\hat{\gamma}_0$	-1.91332	4.66688
$\hat{\gamma}_1$	0.09228	0.47763
$\hat{\gamma}_2$	0.00001	0.01585
$\hat{\gamma}_3$	0.02119	0.01852
$\hat{\gamma}_4$	-0.09077	0.05269
$\hat{\beta}_0$	-7.83031	1.58294
$\hat{\beta}_1$	0.64117	0.16248
$\hat{\beta}_2$	0.00886	0.00536
$\hat{\beta}_3$	0.00538	0.00722
$\hat{\beta}_4$	0.01673	0.01879
$\hat{\phi}$	6.05032	0.52647

Utilizando o procedimento automático de seleção de modelos `stepGAIC` no pacote `gamlss` do R selecionamos o melhor modelo. O procedimento usa o critério de Akaike (AIC) para escolher o modelo que melhor se ajusta aos dados. O modelo finalmente selecionado é

$$\begin{aligned}\text{logit}(\alpha) &= \gamma_0 + \gamma_4 ltv, \\ \text{logit}(\mu) &= \beta_0 + \beta_1 lin,\end{aligned}$$

com valor do AIC igual a 121.9194. De fato, o teste da razão de verossimilhanças de $H_0 : \beta_2 = \beta_3 = \beta_4 = \gamma_1 = \gamma_2 = \gamma_3 = 0$ contra $H_1 : (\text{pelo menos uma igualdade não é satisfeita})$ leva ao valor da estatística $\Lambda = 4.7870$ e p -valor igual a 0.5714, ou seja, H_0 não é rejeitada aos níveis de significância usuais.

Na Tabela 6.3 apresentamos para este modelo as estimativas de máxima verossimilhança e suas versões corrigidas por viés de segunda ordem com seus respectivos erros-padrão[†]. Observamos que as estimativas pontuais dos parâmetros de regressão do componente discreto e contínuo são muito próximas das respectivas estimativas corrigidas, o que era de se esperar segundo os resultados discutidos na Seção 4.4.

[†]A estimativa do erro padrão do estimador corrigido é calculada através do erro padrão assintótico do estimador de máxima verossimilhança avaliado no estimador corrigido. Por exemplo, o erro padrão assintótico de $\tilde{\beta}_r$, $r = 1, \dots, n$, é $\sqrt{\tilde{K}_{rr}^{\beta\beta}}$, em que $\tilde{K}_{rr}^{\beta\beta}$ é o (r, r) -ésimo elemento da matriz $K^{\beta\beta}$ obtida em (3.2.19) e sendo avaliada em $(\tilde{\beta}^\top, \tilde{\phi}^\top)^\top$.

Ao analisar as estimativas do parâmetro de precisão, observamos apenas uma ligeira mudança entre a estimativa de máxima verossimilhança e a estimativa corrigida, sendo que a correção pelo viés de segunda ordem do estimador $\hat{\phi}$ produz um erro padrão ligeiramente menor.

Os sinais das estimativas dos parâmetros indicam que, para comunidades em que há assinantes que adquirem serviços extras de televisão a cabo, à medida que aumenta a renda da comunidade há uma tendência de aumentar a proporção média desses assinantes. Adicionalmente, a probabilidade de não haver assinantes que adquirem serviços extras de televisão a cabo é menor para comunidades com muitos canais com sinal local.

Tabela 6.3: Estimativas de máxima verossimilhança e suas versões corrigidas com erros padrões

Estimador	Estimativa	Erro Padrão
$\hat{\gamma}_0$	-0.71435	0.31182
$\tilde{\gamma}_0$	-0.71435	0.31182
$\hat{\gamma}_4$	-0.09054	0.04722
$\tilde{\gamma}_4$	-0.09054	0.04722
$\hat{\beta}_0$	-8.29788	1.46403
$\tilde{\beta}_0$	-8.29787	1.52687
$\hat{\beta}_1$	0.73626	0.14387
$\tilde{\beta}_1$	0.73626	0.14996
$\hat{\phi}$	5.94906	0.51718
$\tilde{\phi}$	5.89983	0.46565

Com o objetivo de verificar possíveis afastamentos das suposições feitas para o modelo construímos os gráficos dos resíduos r_{p_t} , $r_{p_t}^*$ e r_t^q , definidos em (5.2.3), (5.2.4) e (5.2.11) respectivamente, contra os índices das observações. Na Figura 6.2 apresentamos os gráficos dos resíduos padronizados (Figura 6.2(a)) e dos resíduos ponderados (Figura 6.2(b)) e, dos resíduos quantis aleatorizados (Figura 6.2(c)). Em relação a estes gráficos podemos destacar a presença da observação 108 como possível ponto aberrante. Já no gráfico do resíduo quantil aleatorizado percebemos que o padrão aleatório dos resíduos é mais evidente do que aquele mostrado pelos resíduos padronizados e ponderados, indicando a não existência de uma tendência sistemática nas observações, o que sugere uma boa adequação do ajuste do modelo proposto para os dados.

Na Figura 6.3 apresentamos os gráficos de $r_{p_t}^{(1)}$ (componente discreto) e $r_{p_t}^{(2)}$ (componente contínuo) contra os valores ajustados $\hat{\alpha}_t$ e $\hat{\mu}_t$, respectivamente. Note que, para o componente discreto, as observações 21 e 54 se apresentam como marginalmente aberrantes, com destaque para a observação 21, cujo valor do resíduo é igual a 2.88 (Figura 6.3(a)). Já para o componente contínuo, o ponto que se apresenta como marginalmente aberrante é o 108 (Figura 6.3(b)), com valor de resíduo igual a -4.91 .

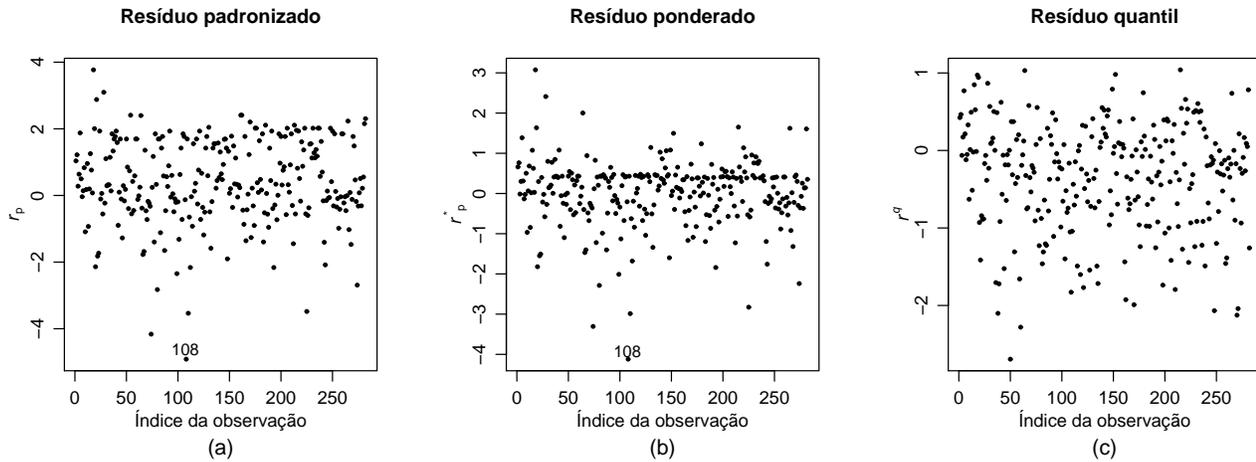


Figura 6.2: Gráficos de resíduos. Dados de difusão de televisão a cabo.

Percebemos que em ambos os gráficos não se destacam observações em comum, evidenciando ainda mais a heterogeneidade entre as observações, i.e., o comportamento discreto e contínuo do modelo.

Com relação aos gráficos normais de probabilidades (Figura 6.4), notamos que não há fortes indícios de afastamento da suposição de que o modelo de regressão beta inflacionado em zero é adequado para os dados, uma vez que, em geral, os resíduos permanecem dentro das bandas de confiança dos envelopes simulados. Destacamos que a distribuição de cada um dos resíduos tem acentuada assimetria à esquerda, como era de se esperar em virtude da presença de massa de probabilidade em zero.

Com o objetivo de verificar a existência de possíveis pontos influentes nos dados aplicamos a técnica de influência local baseada na curvatura normal conforme, sob o esquema de perturbação de casos (ver Seção 5.3.1). Tomamos valores para $q = 1, 2, 3, 4$ para determinar quais são os autovetores (direções) associados às variações máximas da função do deslocamento pela verossimilhança. Na Figura 6.5(a) podemos observar o gráfico dos autovalores normalizados em módulo e os valores de q . Lembrando que $n = 282$, note que para $q = 4$ temos apenas um autovalor acima do limiar q/\sqrt{n} (linhas horizontais) e para $q = 3$, temos apenas dois autovalores acima desse limiar, i.e., temos um autovetor 4-influente e dois autovetores 3-influentes. Os autovalores normalizados associados a estes dois autovetores são 0.9741 e 0.2263. Desta forma, a máxima curvatura normal conforme $B_{d_{\max}}$ é 0.9741 e d_{\max} , o autovetor correspondente, sintetiza as maiores variações do deslocamento da verossimilhança $F(\omega)$ quando o modelo está sendo perturbado.

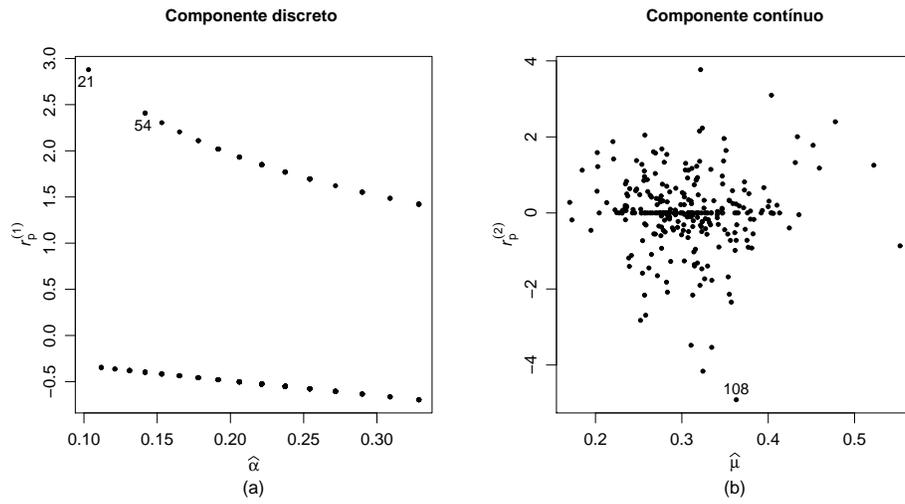


Figura 6.3: Gráficos de resíduos do componente discreto e contínuo. Dados de difusão de televisão a cabo.

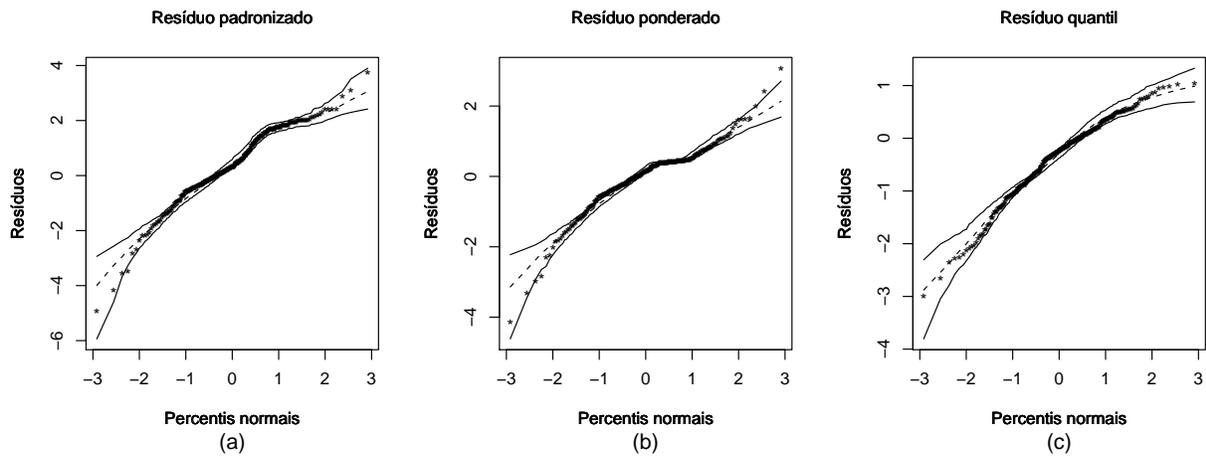


Figura 6.4: Gráficos normais de probabilidades com envelopes simulados. Dados de difusão de televisão a cabo.

Note que, como este valor se encontra entre 0 e 1 podemos interpretá-lo como a proporção do deslocamento da verossimilhança que é explicada por esse autovetor.

No gráfico da Figura 6.5(b) estão apresentados os valores da contribuição agregada de todos os autovetores ($q = 0$). Como foi sugerido em Poon & Poon (1999), as observações acima do ponto de corte $2b$ (ver (5.3.5)) são consideradas influentes. Neste caso, $2b = 0.0802$. No entanto, há muitas observações acima deste valor. Desta forma, selecionamos para nossa análise aquelas que se encontram mais distantes. Especificamente, os casos são $j = 2, 6, 8, 10, 14, 17, 19, 22, 24, 25, 26, 27, 28, 29, 36$ e 260 .

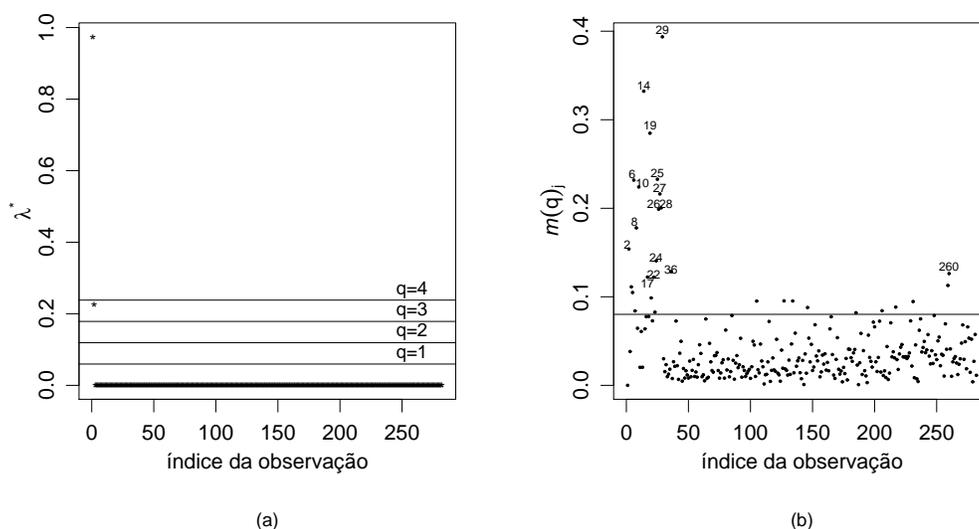


Figura 6.5: (a) Autovalores normalizados em módulo, λ_i^* , com valores de q e (b) influência devida à contribuição agregada de todos os autovetores, $m[q]_j$; esquema de ponderação de casos.

Na Tabela 6.4 estão apresentados o número de autovetores $N(e)$ que são q -influentes, os limiares críticos para B_{E_j} (ver (5.3.4)) e $m[q]_j$ (ver (5.3.3)), com $q = 3, 4$. Nesta mesma tabela se encontram os valores da curvatura normal conforme do vetor de perturbação básico B_{E_j} e a contribuição agregada $m[q]_j$ para os autovetores 3-influentes e 4-influentes. Se a diferença absoluta entre o limiar (valor crítico) e o valor $m[q]_j$ está muito próxima de zero, dizemos que a observação é marginalmente influente. Para $q = 3$ as observações 14, 17, 19, 22, 25, 29 e 260 são detectadas como influentes (ver Figura 6.6(a)) e 2, 6, 8, 10, 24, 26 27 28 36 são detectadas como marginalmente influentes. No caso de $q = 4$ as observações 10, 14, 17, 19, 22, 29

e 260 são consideradas influentes (ver Figura 6.6(b)) e 2, 6, 8, 24, 25, 26, 27, 28 e 36 são detectadas como marginalmente influentes. Assim, se selecionamos dentre as observações consideradas influentes para $q = 3$ e $q = 4$ aquelas que apresentam as maiores diferenças absolutas entre o limiar (valor crítico) e o valor $m[q]_j$ temos que as observações 14, 19 e 29 são detectadas como globalmente influentes.

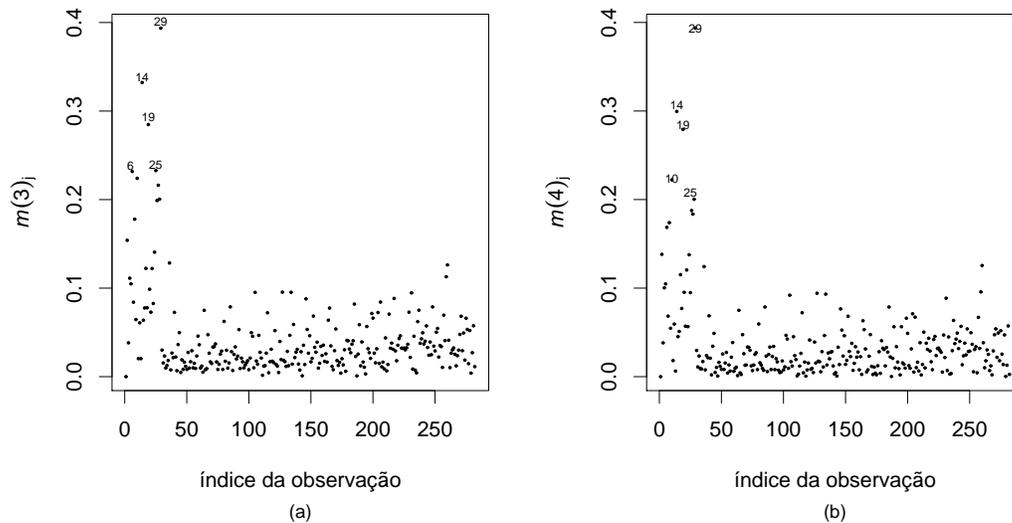


Figura 6.6: (a) Contribuição agregada do autovetor 4-influente correspondente ao maior autovalor ($q = 4$), e (b) contribuição agregada dos autovetores 3-influente correspondentes aos dois maiores autovalores ($q = 3$); esquema de ponderação de casos.

As observações influentes detectadas pelo esquema da ponderação de casos (14, 19 e 29) e a aberrante detectada pelos gráficos dos resíduos padronizados e ponderados (108) correspondem a comunidades que possuem boa renda e adquirem uma proporção relativamente baixa de serviços adicionais de televisão a cabo. Fizemos uma retirada de tais observações, uma a uma e conjuntamente, com a finalidade de verificar o impacto causado no ajuste do modelo. Para avaliar a magnitude do impacto exercido pelas observações utilizamos as seguintes medidas:

6.1. Conclusões

- Desvio relativo percentual (DRP):

$$\text{DRP} = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\theta}} \times 100\%,$$

sendo \hat{y}_j o valor ajustado de y_j e $\hat{\theta}^*$ o estimador do parâmetro θ encontrado ao retirar uma ou várias (dependendo o caso) observações atípicas.

- Desvio quadrático médio relativo percentual (DQRP):

$$\text{DQRP} = \frac{\text{DQM}^* - \text{DQM}}{\text{DQM}} \times 100\%,$$

em que $\text{DQM} = \sqrt{(1/n) \sum_{j=1}^n (y_j - \hat{y}_j)^2}$ e DQM^* é o valor de DQM obtido ao retirar uma ou várias (dependendo o caso) observações atípicas.

- Erro absoluto médio relativo percentual (EAMRP):

$$\text{EAMRP} = \frac{\text{EAM}^* - \text{EAM}}{\text{EAM}} \times 100\%,$$

em que $\text{EAM} = (1/n) \sum_{j=1}^n |y_j - \hat{y}_j|$ e EAM^* é o valor de EAM obtido ao retirar uma ou várias (dependendo o caso) observações atípicas.

A Tabela 6.5 apresenta as estimativas, o desvio padrão e o desvio relativo percentual DPR dos parâmetros estimados ao retirar uma a uma e coletivamente as observações consideradas como influentes e aberrantes. Note que as estimativas dos parâmetros não sofreram grande impacto. Como podemos ver, as maiores variações percentuais ocorrem para as estimativas de β_0 e β_1 quando as observações 19 e 108 são retiradas individualmente e quando as observações 19 e 29 são retiradas coletivamente. No entanto, estas variações não superam 10%. Realizamos testes de hipóteses para verificar as significâncias dos parâmetros do modelo e verificamos que, para todos os casos considerados de exclusões os parâmetros são significativos ao nível de confiança de 5%. Desta forma, podemos concluir que as observações detectadas como influentes e aberrantes não afetam os resultados inferenciais do modelo.

Em relação ao impacto que as observações detectadas como influentes e aberrantes causam nos valores ajustados, podemos observar na Tabela 6.6 que não existem mudanças consideráveis ao retirar individual ou conjuntamente tais observações. Isto indica um certo grau de robustez do modelo ajustado a esses dados na presença de observações influentes e aberrantes.

6.1 Conclusões

Em nosso exemplo, vimos que ao considerar a técnica de influência local proposta por Poon & Poon (1999) utilizando o esquema de perturbação de casos as observações 14, 19 e 29 foram destacadas como influentes.

6.1. Conclusões

Tabela 6.4: Medidas de influência utilizando a curvatura normal conforme para o esquema de ponderação de casos.

	BE_j	$m[4]_j$	$m[3]_j$
$N(e)$	282	2	1
obs \ limiar	0.08019	0.17865	0.23820
2	0.15405	0.15405	0.13812
6	0.23179	0.23179	0.16861
8	0.17784	0.17784	0.17391
10	0.22405	0.22405	0.22226
14	0.33222	0.33222	0.29951
17	0.12226	0.12226	0.11518
19	0.28477	0.28477	0.27921
22	0.12211	0.12211	0.12045
24	0.14057	0.14057	0.13768
25	0.23278	0.23278	0.09489
26	0.19894	0.19894	0.18762
27	0.21617	0.21617	0.18354
28	0.20047	0.20047	0.20031
29	0.39364	0.39364	0.39356
36	0.12837	0.12837	0.12428
260	0.12622	0.12622	0.12555

Verificamos através de uma análise confirmatória que tais observações não exercem muito impacto na estimação do modelo. Cabe ressaltar que o método de influência não detectou a observação aberrante 108, a qual foi determinada pela análise dos resíduos. Contudo, a análise confirmatória verificou que esta observação não exerce um impacto significativo nas estimativas no modelo. Desta forma, podemos atribuir um certo grau de robustez no modelo ajustado na presença de observações influentes e/ou aberrantes. Em relação aos gráficos normais de probabilidade com envelope simulado, observamos que em todos há assimetria na distribuição do resíduo considerado, atribuída em parte à presença de massa de probabilidade em zero. Notamos que o resíduo quantil aleatorizado parece ser mais adequado já que consegue adaptar-se melhor à característica discreta e contínua do modelo. De forma geral, o modelo de regressão beta inflacionado em zero mostrou-se adequado para descrever o comportamento da proporção de assinantes de televisão a cabo que adquirem serviços adicionais em função das covariadas consideradas.

6.1. Conclusões

Tabela 6.5: Estimativas, desvio padrão e DRP dos parâmetros do modelo com a amostra completa e tirando as observações influentes

Estimativas, erro padrão e DRP dos parâmetros do modelo quando são retiradas observações influentes.					
Observações	$\hat{\gamma}_0^*$	$\hat{\gamma}_4^*$	$\hat{\beta}_0^*$	$\hat{\beta}_1^*$	$\hat{\phi}^*$
14	-0.70673	-0.09099	-8.31490	0.73794	5.92453
	(0.02117)	(0.00048)	(0.00903)	(0.00009)	(0.51692)
	-1.066%	0.497%	0.205%	0.228%	-0.412%
19	-0.72148	-0.08877	-7.91028	0.69729	6.02219
	(0.02115)	(0.00048)	(0.00901)	(0.00009)	(0.52599)
	0.998%	-1.954%	-4.671%	-5.292%	1.229%
29	-0.71795	-0.08929	-8.29446	0.73607	5.92458
	(0.02116)	(0.00048)	(0.00903)	(0.00009)	(0.51692)
	0.503%	-1.380%	-0.041%	-0.025%	-0.411%
108	-0.71795	-0.08929	-8.82054	0.78856	6.24830
	(0.02116)	(0.00048)	(0.00890)	(0.00009)	(0.54730)
	0.503%	-1.380%	6.298%	7.103%	5.030%
14, 19	-0.71385	-0.08922	-7.92558	0.69880	5.99707
	(0.02118)	(0.00048)	(0.00907)	(0.00009)	(0.52486)
	-0.069%	-1.457%	-4.486%	-5.087%	0.807%
14, 29	-0.71032	-0.08974	-8.31129	0.73772	5.90004
	(0.02118)	(0.00048)	(0.00909)	(0.00009)	(0.51582)
	-0.564%	-0.883%	0.161%	0.198%	-0.824%
19, 29	-0.72513	-0.08750	-7.90735	0.69713	5.99714
	(0.02117)	(0.00048)	(0.00907)	(0.00009)	(0.52486)
	1.509%	-3.357%	-4.706%	-5.314%	0.808%
14, 19, 29	-0.71749	-0.08795	-7.92247	0.69863	5.97201
	(0.02119)	(0.00048)	(0.00913)	(0.00009)	(0.52373)
	0.439%	-2.860%	-4.524%	-5.110%	0.385%

Tabela 6.6: Medida de impacto DQMRP e EAMRP para os valores ajustados (em porcentagem).

Medida	observações retiradas							
	14	19	29	108	14,19	14,29	19, 29	14, 19, 29
DQMRP	0.031%	-0.128%	0.034%	-0.063%	-0.096%	0.066%	-0.093%	-0.061%
EAMRP	0.039%	-0.103%	0.050%	-0.048%	-0.063%	0.089%	-0.052%	-0.012%

Capítulo 7

Conclusões

7.1 Considerações finais

Ao longo deste trabalho desenvolvemos modelos de regressão beta inflacionados como extensões naturais do modelo de regressão beta proposto por Ferrari e Cribari Neto (2004). Os modelos propostos têm aplicabilidade prática na modelagem de dados de proporções, taxas ou frações com presença de zeros e/ou uns. Neste sentido, propomos modelos de regressão em que a distribuição da variável resposta é uma mistura entre uma distribuição beta e uma distribuição de Bernoulli (modelo de regressão beta inflacionado em zero e um), degenerada em zero (modelo de regressão beta inflacionado em zero) ou degenerada em um (modelo de regressão beta inflacionado em um). Aqui os parâmetros que modelam a probabilidade de ocorrer zero e/ou um são definidos através de estruturas de regressão.

Para cada modelo apresentado, propomos a modelagem dos parâmetros através de preditores lineares usando funções de ligação adequadas. Desenvolvemos aspectos de estimação por máxima verossimilhança tais como a obtenção de expressões matriciais para o vetor escore e a matriz de informação de Fisher, a construção de intervalos de confiança assintóticos, a realização de testes de hipóteses e a seleção de modelos, entre outros.

Com o fim de melhorar o desempenho em amostras finitas dos estimadores de máxima verossimilhança para estes modelos, derivamos expressões que permitem corrigir o viés até segunda ordem dos estimadores de máxima verossimilhança dos parâmetros dos modelos de regressão beta inflacionados. Através de estudos de simulação detectamos que os estimadores dos parâmetros de regressão possuem boas propriedades no sentido de serem pouco viesados. Já para o parâmetro de precisão, a correção se torna importante, pois melhora consideravelmente o desempenho do estimador de máxima verossimilhança. Encontramos que uma parametrização da distribuição beta inflacionada em termos de um parâmetro de dispersão (em lugar de um parâmetro de precisão) pode ser mais interessante na modelagem deste tipo de dados, permitindo que a estimação dos parâmetros seja menos comprometida em termos do viés do estimador de máxima verossimilhança.

Para validar as suposições dos modelos de regressão beta inflacionados, desenvolvemos métodos de diagnóstico. Para isto, criamos resíduos padronizados, ponderados e resíduos quantil aleatorizados Adicional-

7.2. *Trabalhos futuros*

mente, desenvolvemos medidas de influência local baseadas na curvatura normal conforme sob diferentes esquemas de perturbação. Para o modelo de regressão beta inflacionado em zero ou um, derivamos expressões matriciais para a curvatura normal conforme sob dois esquemas de perturbação, a saber: a ponderação de casos e a perturbação individual de covariáveis. Já para o modelo de regressão beta inflacionado em zero e um, derivamos a medida de curvatura normal conforme sob o esquema de ponderação de casos. Através de um exemplo com dados reais comprovamos a potencialidade da metodologia de influência local no sentido de identificar as observações que podem afetar desproporcionalmente os resultados inferenciais obtidos para o modelo ajustado aos dados.

Finalmente, sugerimos, aos usuários interessados em modelar dados na forma de frações, proporções ou taxas medidas de forma contínua, que utilizem modelos de regressão beta inflacionados sempre que zeros e/ou uns estejam presentes no conjunto de dados.

7.2 Trabalhos futuros

Dando continuidade a esta tese concluiremos e disponibilizaremos a implementação computacional da metodologia desenvolvida no pacote estatístico R.

Como foco de trabalhos futuros sugerimos os seguintes temas de pesquisa:

1. Desenvolver testes de hipóteses adaptativos para os parâmetros dos modelos de regressão beta inflacionados.
2. Estudar métodos de estimação robusta e de máxima verossimilhança penalizada para os modelos de regressão beta inflacionados.
3. Analisar e implementar a correção por viés utilizando métodos jackknife e bootstrap para os modelos de regressão beta inflacionados.
4. Desenvolver testes de influência local para fornecer um diagnóstico global da qualidade de ajuste do modelo usando uma modificação da proposta dada em Zhu & Zhang (2004).
5. Estender os resultados desta tese para os modelos de regressão beta inflacionados com dispersão variável.

Apêndice A

Apêndices

A.1 Apêndice 1

Derivadas e cumulantes de segunda ordem do logaritmo da função de verossimilhança no modelo de regressão beta inflacionado no ponto $c = 0$ ou $c = 1$

Para obter a matriz de informação de Fisher, calculamos os momentos das derivadas de segunda ordem da função de log-verossimilhança. Utilizamos a notação dada por Lawley (1956), em que $-\kappa_{rs} = \kappa_{r,s}$ representa o elemento (r, s) da matriz de informação de Fisher $K(\theta)$. Como já foi definido em (5.3.7) temos que

$$\ell(\theta) = \ell_1(\gamma) + \ell_2(\beta, \phi).$$

Para $R = 1, \dots, M$, e $r = 1, \dots, m$, os momentos de segunda ordem obtidos pela diferenciação da função de log-verossimilhança são

$$\begin{aligned} U_{RS} &= \frac{\partial^2 \ell_1(\gamma)}{\partial \gamma_R \gamma_S} = \sum_{t=1}^n \frac{\partial}{\partial \alpha_t} \left(\frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \frac{\partial \zeta_t}{\partial \gamma_R} \right) \frac{d\alpha_t}{d\zeta_t} \frac{\partial \zeta_t}{\partial \gamma_S} \\ &= \sum_{t=1}^n \left\{ \frac{\partial^2 \ell_t(\alpha_t)}{\partial \alpha_t^2} \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 + \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} z_{tS} z_{tR}. \end{aligned}$$

Logo,

$$U_{RS} = \sum_{t=1}^n \left\{ \left(\frac{-\mathbb{1}_{(0,1)}(y_t)}{(1-\alpha_t)^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^2} \right) \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 + \left(\frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1-\alpha_t)} \right) \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} z_{tS} z_{tR}. \quad (\text{A.1.1})$$

A.1. Apêndice 1

Agora,

$$\begin{aligned} U_{rs} &= \frac{\partial^2 \ell_2(\beta, \phi)}{\partial \beta_r \partial \beta_s} = \sum_{t:y_t \in (0,1)} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_s} \\ &= \sum_{t:y_t \in (0,1)} \left\{ \left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right) + \left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \right\} x_{ts} x_{tr}, \end{aligned}$$

onde

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} = \frac{\partial}{\partial \mu_t} \{ \phi [y_t^* - \mu_t^*] \} = -\phi \frac{\partial \mu_t^*}{\partial \mu_t}$$

e

$$-\phi \frac{\partial \mu_t^*}{\partial \mu_t} = -\phi \{ \phi [\psi'(\mu_t \phi) + \psi'((1 - \mu_t) \phi)] \},$$

sendo $w_t = [\psi'(\mu_t \phi) + \psi'((1 - \mu_t) \phi)]$, então $\partial \mu_t^* / \partial \mu_t = \phi w_t$ e $\partial^2 \ell_t(\mu_t, \phi) / \partial \mu_t^2 = -\phi^2 w_t$. Assim,

$$U_{rs} = \sum_{t:y_t \in (0,1)} \left\{ -\phi^2 w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \phi [y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right\} x_{ts} x_{tr}. \quad (\text{A.1.2})$$

Temos ainda

$$\begin{aligned} U_{\phi\phi} &= \frac{\partial^2 \ell_2(\beta, \phi)}{\partial \phi^2} = \sum_{t:y_t \in (0,1)} \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \phi^2} \\ &= \sum_{t:y_t \in (0,1)} \left\{ -\mu_t \frac{\partial \mu_t^*}{\partial \phi} + \psi'(\phi) - (1 - \mu_t) \psi'((1 - \mu_t) \phi) \right\}, \end{aligned}$$

onde

$$\begin{aligned} \frac{\partial \mu_t^*}{\partial \phi} &= \mu_t \psi'(\mu_t \phi) - (1 - \mu_t) \psi'((1 - \mu_t) \phi) \\ &= \mu_t \{ \psi'(\mu_t \phi) + \psi'((1 - \mu_t) \phi) \} - \psi'((1 - \mu_t) \phi) \\ &= \mu_t w_t - \psi'((1 - \mu_t) \phi). \end{aligned}$$

Desta forma,

$$\begin{aligned} U_{\phi\phi} &= \sum_{t:y_t \in (0,1)} \left\{ -\mu_t [\mu_t w_t - \psi'((1 - \mu_t) \phi)] + \psi'(\phi) - (1 - \mu_t) \psi'((1 - \mu_t) \phi) \right\} \\ &= \sum_{t:y_t \in (0,1)} \left\{ -\mu_t^2 \psi'(\mu_t \phi) + \mu_t (1 - \mu_t) \psi'((1 - \mu_t) \phi) + \psi'(\phi) - (1 - \mu_t) \psi'((1 - \mu_t) \phi) \right\} \\ &= \sum_{t:y_t \in (0,1)} - \left\{ (1 - \mu_t)^2 \psi'((1 - \mu_t) \phi) + \mu_t^2 \psi'(\mu_t \phi) - \psi'(\phi) \right\}. \end{aligned}$$

sendo $d_t = (1 - \mu_t)^2 \psi'((1 - \mu_t) \phi) + \mu_t^2 \psi'(\mu_t \phi) - \psi'(\phi)$. Logo,

$$U_{\phi\phi} = - \sum_{t:y_t \in (0,1)} d_t. \quad (\text{A.1.3})$$

A.1. Apêndice 1

Adicionalmente,

$$U_{r\phi} = \frac{\partial^2 \ell_2(\beta, \phi)}{\partial \beta_r \partial \phi} = \sum_{t: y_t \in (0,1)} \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r},$$

em que

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} = \frac{\partial}{\partial \phi} [\phi(y_t^* - \mu_t^*)] = (y_t^* - \mu_t^*) - \phi \frac{\partial \mu_t^*}{\partial \phi}.$$

Se definimos, $\mathbf{c}_t = \phi[\mu_t w_t - \psi'((1 - \mu_t)\phi)]$, temos $\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} = (y_t^* - \mu_t^*) - \mathbf{c}_t$. Assim,

$$U_{r\phi} = \sum_{t: y_t \in (0,1)} -\{\mathbf{c}_t - (y_t^* - \mu_t^*)\} \frac{d\mu_t}{d\eta_t} x_{tr}. \quad (\text{A.1.4})$$

Note que da separabilidade dos vetores de parâmetros γ e $(\beta^\top, \phi)^\top$ obtemos

$$U_{R\phi} = U_{rR} = 0.$$

Sob usuais condições de regularidade, temos que $E(\partial \ell_t(\alpha_t)/\partial \alpha_t) = 0$. Assim, para $R = 1, \dots, M$,

$$\begin{aligned} \kappa_{RS} = E(U_{RS}) = E \left(\sum_{t=1}^n \left\{ \left(\frac{-(1 - \mathbb{1}_{\{c\}}(y_t))}{(1 - \alpha_t)^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^2} \right) \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 \right. \right. \\ \left. \left. + \left(\frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t} - \frac{1 - \mathbb{1}_{\{c\}}(y_t)}{(1 - \alpha_t)} \right) \frac{\partial}{\partial \alpha_t} \left(\frac{d\alpha_t}{d\zeta_t} \right) \frac{\partial \zeta_t}{\partial \gamma_R} \right\} z_{tS} z_{tR} \right). \end{aligned} \quad (\text{A.1.5})$$

Como $E(\mathbb{1}_{\{c\}}(y_t)) = \alpha_t$ e $E(1 - \mathbb{1}_{\{c\}}(y_t)) = 1 - \alpha_t$, a expressão anterior reduz-se a

$$\kappa_{RS} = - \sum_{t=1}^n p_t \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 z_{tS} z_{tR}, \quad (\text{A.1.6})$$

em que $p_t = 1/[\alpha_t(1 - \alpha_t)]$.

Para o caso dos cumulantes envolvendo o vetor de parâmetros β e o parâmetro ϕ provaremos a seguinte proposição.

Proposição A.1.1. *Seja $(y_1, \dots, y_n)^\top$ um vetor de n variáveis aleatórias independentes com y_t seguindo a distribuição beta inflacionada (2.3.2), i.e., $y_t \sim \text{BI}_c(\alpha_t, \mu_t, \phi)$, $t = 1, \dots, n$. Se $\mathfrak{F} : (0, 1) \rightarrow \mathbb{R}$ é uma função contínua, então*

$$E \left(\sum_{t: y_t \in (0,1)} \mathfrak{F}(y_t) \right) = \sum_{t=1}^n (1 - \alpha_t) E(\mathfrak{F}(y_t) | \mathbb{1}_{\{c\}}(y_t) = 0).$$

Demonstração. Lembremos que o suporte da distribuição $\text{BI}_c(\alpha_t, \mu_t, \phi)$ é o conjunto $(0, 1) \cup \{c\}$ com $(c = 0)$ ou $(c = 1)$. Definimos

$$\mathfrak{F}^*(y_t) = \begin{cases} 0, & \text{se } y_t = c, \\ \mathfrak{F}(y_t), & \text{se } y_t \in (0, 1). \end{cases}$$

A.2. Apêndice 2

Então,

$$\mathbb{E}\left(\sum_{t:y_t \in (0,1)} \mathfrak{T}(y_t)\right) = \mathbb{E}\left(\sum_{t=1}^n \mathfrak{T}^*(y_t)\right) = \sum_{t=1}^n \mathbb{E}(\mathfrak{T}^*(y_t)).$$

Agora

$$\begin{aligned} \mathbb{E}(\mathfrak{T}^*(y_t)) &= \mathbb{E}(\mathfrak{T}^*(y_t) | \mathbb{1}_{\{c\}}(y_t) = 0)P(\mathbb{1}_{\{c\}}(y_t) = 0) + \mathbb{E}(\mathfrak{T}^*(y_t) | \mathbb{1}_{\{c\}}(y_t) = 1)P(\mathbb{1}_{\{c\}}(y_t) = 1) \\ &= \mathbb{E}(\mathfrak{T}^*(y_t) | \mathbb{1}_{\{c\}}(y_t) = 0)P(\mathbb{1}_{\{c\}}(y_t) = 0) = (1 - \alpha_t)\mathbb{E}(\mathfrak{T}(y_t) | \mathbb{1}_{\{c\}}(y_t) = 0). \end{aligned}$$

Daí, segue o resultado. ✠

Note que, sob usuais condições de regularidade, obtemos $\mathbb{E}(\partial \ell_t(\mu_t, \phi) / \partial \mu_t) = 0$ e $\mathbb{E}(\partial \ell_t(\mu_t, \phi) / \partial \phi) = 0$. Assim, para $r = 1, \dots, m$, temos que

$$\kappa_{rs} = \mathbb{E}(U_{rs}) = \mathbb{E}\left(\sum_{t:y_t \in (0,1)} \left\{ -\phi^2 w_t \left(\frac{d\mu_t}{d\eta_t}\right)^2 + \phi[y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right) \right\} x_{ts} x_{tr}\right).$$

Note que da proposição A.1.1 e do fato de $\mathbb{E}((y_t^* - \mu_t^*) | \mathbb{1}_{\{c\}}(y_t) = 0) = 0$, obtemos

$$\kappa_{rs} = -\phi^2 \sum_{t=1}^n (1 - \alpha_t) w_t \left(\frac{d\mu_t}{d\eta_t}\right)^2 x_{ts} x_{tr}. \quad (\text{A.1.7})$$

Analogamente,

$$\begin{aligned} \kappa_{r\phi} &= \mathbb{E}(U_{r\phi}) = \mathbb{E}\left(\sum_{t:y_t \in (0,1)} -\{c_t - [y_t^* - \mu_t^*]\} \frac{d\mu_t}{d\eta_t} x_{tr}\right) \\ &= -\sum_{t=1}^n (1 - \alpha_t) c_t \frac{d\mu_t}{d\eta_t} x_{tr} \end{aligned} \quad (\text{A.1.8})$$

e

$$\kappa_{\phi\phi} = \mathbb{E}(U_{\phi\phi}) = \mathbb{E}\left(\sum_{t:y_t \in (0,1)} -d_t\right) = -\sum_{t=1}^n (1 - \alpha_t) d_t. \quad (\text{A.1.9})$$

Finalmente, pela separabilidade dos vetores de parâmetros γ e $(\beta^\top, \phi)^\top$ temos

$$\kappa_{R\phi} = \kappa_{rR} = 0. \quad (\text{A.1.10})$$

A.2 Apêndice 2

Estimação por máxima verossimilhança do modelo RBIZ com dados simulados

```
library(gamlss)
library(gamlss.dist)
```

A.2. Apêndice 2

```
#Simulação do modelo RBIZ
tam=500 #tamanho de amostra

#Geração do preditor linear de mu
x0=rep(1,tam)
x1=runif(tam) #covariada x1 de mu uniforme
x.mu=cbind(x0,x1) # matriz de regressores de mu
par.mu=c(-1.5, 1.5) #parâmetros da regressão de mu
pred.mu=x.mu**par.mu #preditor linear de mu

#mu relacionado com preditor usando ligação logito
mu=exp(pred.mu)/(1+exp(pred.mu))

#Geração do preditor linear de alpha
z0=rep(1,tam)
z1=runif(tam) #covariada z1 de alpha uniforme
z.alpha=cbind(z0,z1) # matriz de regressores de alpha
par.alpha=c(-1.0, 0.5) #parâmetros da regressão de alpha
pred.alpha=z.alpha**par.alpha #preditor linear de alpha

#alpha relacionado com preditor usando ligação logito
alpha=exp(pred.alpha)/(1+exp(pred.alpha))

# parâmetro de precisão
phi = 50 # phi constante para todas as observações

set.seed(121) #semente

#----gerador da variável resposta do modelo RBIZ ----#
ydata = mapply(rBEZI,1, mu, phi, alpha)

###----- Ajuste do Modelo RBIZ usando GAMLSS -----###
# Ospina, R (2006). The zero-inflated beta distribution for fitting a
# GAMLSS. Available in the package {\tt gamlss.dist:} Extra
# distributions to be used for GAMLSS modelling.
###-----###

zeros = length(which(ydata==0)) #numero de zeros
zeros/ tam #porcentagen de zeros
(tam - zeros)/ tam #nporcentagen de observacoes distintas de zero

fit=gamlss(ydata~(-1+x.mu), nu.formula=~(-1+z.alpha), family=BEZI)

meany = meanBEZI(fit)

summary(fit) #Resumo de estatísticas

*****
Family: c("BEZI", "Zero Inflated Beta")
Call:
gamlss(formula = ydata ~ (-1 + x.mu), sigma.formula = ~1,
nu.formula = ~(-1 + z.alpha), family = BEZI)
Fitting method: RS()
-----
Mu link function: logit
Mu Coefficients:
```

A.3. Apêndice 3

```

      Estimate Std. Error t value Pr(>|t|)
x.mux0  -1.531    0.03157  -48.50 3.424e-190
x.mux1   1.532    0.05293   28.94 1.633e-108
-----
Sigma link function: log
Sigma Coefficients:
      Estimate Std. Error t value Pr(>|t|)
 4.041e+00  7.586e-02  5.327e+01 4.223e-207
-----
Nu link function: logit
Nu Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.alpha0 -0.9628    0.1973  -4.881 1.426e-06
z.alpha1  0.3752    0.3352   1.120 2.634e-01
-----
No. of observations in the fit: 500
Degrees of Freedom for the fit: 5
Residual Deg. of Freedom: 495
                        at cycle: 5

Global Deviance: -342.6152
AIC: -332.6152
SBC: -311.5421
*****

fit$sigma.fv[1] # estimativa de phi
56.87893

```

A.3 Apêndice 3

Derivadas e cumulantes do logaritmo da função de verossimilhança no modelo de regressão beta inflacionado em zero e um

Para o modelo de regressão RBIZU temos que a função de ligação $g(\cdot)$ para o parâmetro μ pode ser da forma $g(\mu) = \log(\mu/1 - \mu)$ ou $g(\mu) = \Phi^{-1}(\mu)$, $\Phi(\cdot)$ sendo a função de distribuição acumulada da variável normal padrão, entre outras. Desta forma,

$$\frac{d\mu_t}{d\eta_t} = \frac{dg^{-1}(\eta_t)}{d\eta_t} = \frac{1}{g'(\mu_t)}.$$

Adicionalmente, se de (3.3.1) o sistema de equações $(\zeta_{0t}, \zeta_{1t}) = (f_0(\delta_{0t}, \delta_{1t}), f_1(\delta_{0t}, \delta_{1t}))$ define uma transformação 1-1, podemos solucionar de forma única as equações $\zeta_{0t} = f_0(\delta_{0t}, \delta_{1t})$ e $\zeta_{1t} = f_1(\delta_{0t}, \delta_{1t})$ em termos de δ_{0t} e

A.3. Apêndice 3

δ_{1t} . Denotamos esta transformação inversa por $\delta_{0t} = f_0^*(\zeta_{0t}, \zeta_{1t})$, $\delta_{1t} = f_1^*(\zeta_{0t}, \zeta_{1t})$ e desta forma podemos definir

$$\begin{aligned}\frac{\partial \delta_{0t}}{\partial \zeta_{0t}} &= \frac{\partial f_0^*(\zeta_{0t}, \zeta_{1t})}{\partial \zeta_{0t}}, \\ \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} &= \frac{\partial f_0^*(\zeta_{0t}, \zeta_{1t})}{\partial \zeta_{1t}}, \\ \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} &= \frac{\partial f_1^*(\zeta_{0t}, \zeta_{1t})}{\partial \zeta_{0t}}, \\ \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} &= \frac{\partial f_1^*(\zeta_{0t}, \zeta_{1t})}{\partial \zeta_{1t}}.\end{aligned}$$

Particularmente, se consideramos o modelo RLBIZU temos que

$$\begin{aligned}\delta_{0t} &= f_0^*(\zeta_{0t}, \zeta_{1t}) = \frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}, \\ \delta_{1t} &= f_1^*(\zeta_{0t}, \zeta_{1t}) = \frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}\end{aligned}\tag{A.3.1}$$

e

$$1 - \delta_{0t} - \delta_{1t} = \frac{1}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}.$$

Daí,

$$\begin{aligned}\frac{\partial \delta_{0t}}{\partial \zeta_{0t}} &= \left(\frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) \left(\frac{1 + e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) = \delta_{0t}(1 - \delta_{0t}), \\ \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} &= \left(\frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) \left(\frac{1 + e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) = \delta_{1t}(1 - \delta_{1t}), \\ \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} &= - \left(\frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) \left(\frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) = -\delta_{0t}\delta_{1t}.\end{aligned}\tag{A.3.2}$$

Consideremos o vetor de parâmetros $(\rho^\top, \gamma^\top)^\top$ e definamos $\alpha_t = \delta_{0t} + \delta_{1t}$. As derivadas de primeira ordem da função $\ell_1(\rho, \gamma)$ definida em (3.3.5) com respeito aos parâmetros de interesse são

$$\begin{aligned}U_{r'} &= \frac{\partial \ell_1(\rho, \gamma)}{\partial \rho_{r'}} = \sum_{t=1}^n \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \\ &= \sum_{t=1}^n \left\{ \frac{\mathbb{1}_{\{0\}}(y_t)}{\delta_{0t}} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} v_{tr'} + \sum_{t=1}^n \left\{ \frac{\mathbb{1}_{\{1\}}(y_t)}{\delta_{1t}} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} v_{tr'}\end{aligned}\tag{A.3.3}$$

e

$$\begin{aligned}U_{r''} &= \frac{\partial \ell_1(\rho, \gamma)}{\partial \gamma_{r''}} = \sum_{t=1}^n \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \\ &= \sum_{t=1}^n \left\{ \frac{\mathbb{1}_{\{0\}}(y_t)}{\delta_{0t}} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} z_{tr''} + \sum_{t=1}^n \left\{ \frac{\mathbb{1}_{\{1\}}(y_t)}{\delta_{1t}} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} z_{tr''},\end{aligned}\tag{A.3.4}$$

A.3. Apêndice 3

em que $r' = 1, \dots, k_0$ e $r'' = 1, \dots, k_1$.

As derivadas de segunda ordem da função de $\ell_1(\rho, \gamma)$ são

$$\begin{aligned}
 U_{r's'} = \frac{\partial^2 \ell_1(\rho, \gamma)}{\partial \rho_{r'} \partial \rho_{s'}} = \sum_{t=1}^n \left\{ \left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}^2} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{s'}} + \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t} \partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{s'}} \right) \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \right. \\
 + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}} \frac{\partial^2 \delta_{0t}}{\partial \zeta_{0t}^2} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \frac{\partial \zeta_{0t}}{\partial \rho_{s'}} + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} \frac{\partial^2 \zeta_{0t}}{\partial \rho_{s'} \partial \rho_{r'}} \\
 + \left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t} \partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{s'}} + \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}^2} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{s'}} \right) \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \\
 \left. + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}} \frac{\partial^2 \delta_{1t}}{\partial \zeta_{0t}^2} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \frac{\partial \zeta_{0t}}{\partial \rho_{s'}} + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \frac{\partial^2 \zeta_{0t}}{\partial \rho_{s'} \partial \rho_{r'}} \right\}, \tag{A.3.5}
 \end{aligned}$$

$$\begin{aligned}
 U_{r''s''} = \frac{\partial^2 \ell_1(\rho, \gamma)}{\partial \gamma_{r''} \partial \gamma_{s''}} = \sum_{t=1}^n \left\{ \left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}^2} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{s''}} + \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t} \partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{s''}} \right) \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \right. \\
 + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}} \frac{\partial^2 \delta_{0t}}{\partial \zeta_{1t}^2} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \frac{\partial \zeta_{1t}}{\partial \gamma_{s''}} + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial^2 \zeta_{1t}}{\partial \gamma_{s''} \partial \gamma_{r''}} \\
 + \left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t} \partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{s''}} + \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}^2} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{s''}} \right) \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \\
 \left. + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}} \frac{\partial^2 \delta_{1t}}{\partial \zeta_{1t}^2} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \frac{\partial \zeta_{1t}}{\partial \gamma_{s''}} + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial^2 \zeta_{1t}}{\partial \gamma_{s''} \partial \gamma_{r''}} \right\}, \tag{A.3.6}
 \end{aligned}$$

$$\begin{aligned}
 U_{r'r''} = \frac{\partial^2 \ell_1(\rho, \gamma)}{\partial \rho_{r'} \partial \gamma_{r''}} = \sum_{t=1}^n \left\{ \left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}^2} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} + \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t} \partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \right) \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \right. \\
 + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}} \frac{\partial^2 \delta_{0t}}{\partial \zeta_{1t} \partial \zeta_{0t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} + \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}} \frac{\partial^2 \delta_{1t}}{\partial \zeta_{0t} \partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \\
 \left. + \left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t} \partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} + \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}^2} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \zeta_{1t}}{\partial \gamma_{r''}} \right) \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \frac{\partial \zeta_{0t}}{\partial \rho_{r'}} \right\}. \tag{A.3.7}
 \end{aligned}$$

Agora, para $r = 1, \dots, k$, as derivadas de primeira ordem da função de $\ell_2(\beta, \phi)$ definida em (3.3.5) são:

$$\begin{aligned}
 U_r = \frac{\partial \ell_2(\beta, \phi)}{\partial \beta_r} &= \sum_{t: y_t \in (0,1)} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} \\
 &= \sum_{t: y_t \in (0,1)} \phi \left[\log\left(\frac{y_t}{1-y_t}\right) - \{\psi(\mu_t \phi) - \psi((1-\mu_t)\phi)\} \right] \frac{d\mu_t}{d\eta_t} x_{tr}.
 \end{aligned}$$

A.3. Apêndice 3

Definindo

$$y_t^* = \begin{cases} \log\left(\frac{y_t}{1-y_t}\right), & \text{se } y_t \in (0, 1), \\ 0, & \text{caso contrário,} \end{cases}$$

e

$$\mu_t^* = \mathbb{E}(y_t^* | \mathbb{1}_{(0,1)}(y_t) = 1) = \psi(\mu_t \phi) - \psi((1 - \mu_t)\phi),$$

obtemos

$$U_r = \phi \sum_{t=1}^n \mathbb{1}_{(0,1)}(y_t) (y_t^* - \mu_t^*) \frac{d\mu_t}{d\eta_t} x_{tr}. \quad (\text{A.3.8})$$

Adicionalmente,

$$\begin{aligned} U_\phi &= \frac{\partial \ell_2(\beta, \phi)}{\partial \phi} = \sum_{t: y_t \in (0,1)} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \phi} \\ &= \sum_{t: y_t \in (0,1)} \left\{ \mu_t \left(\log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t \phi) - \psi((1 - \mu_t)\phi)] \right) + \log(1 - y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi) \right\}. \end{aligned}$$

Definindo

$$s(y_t) = \begin{cases} \log(1 - y_t), & \text{se } y_t \in (0, 1), \\ 0, & \text{caso contrário,} \end{cases}$$

temos

$$U_\phi = \sum_{t=1}^n (\mathbb{1}_{(0,1)}(y_t)) \{ \mu_t (y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi) \}. \quad (\text{A.3.9})$$

Finalmente, as derivadas de segunda ordem da função de $\ell_2(\beta, \phi)$, i.e., U_{rs} , $U_{\phi\phi}$ e $U_{r\phi}$, são as mesmas quantidades obtidas em A.1.2 até A.1.4 de A.1, uma vez que, $\ell_2(\beta, \phi)$ modela a componente contínua, i.e., a variável resposta condicionada ao intervalo $(0, 1)$, e $\alpha_t = \delta_{0t} + \delta_{1t}$ representa o parâmetro de mistura.

Pela separabilidade dos parâmetros $(\rho^\top, \gamma^\top)^\top$ e $(\beta^\top, \phi)^\top$ temos que

$$U_{rr'} = \frac{\partial^2 \ell_1(\rho, \gamma)}{\partial \beta_r \partial \rho_{r'}} = 0,$$

$$U_{rr''} = \frac{\partial^2 \ell_1(\rho, \gamma)}{\partial \beta_r \partial \gamma_{r''}} = 0,$$

$$U_{\phi r'} = \frac{\partial^2 \ell_1(\rho, \gamma)}{\partial \phi \partial \rho_{r'}} = 0,$$

$$U_{\phi r''} = \frac{\partial^2 \ell_1(\rho, \gamma)}{\partial \phi \partial \gamma_{r''}} = 0.$$

A.3. Apêndice 3

Agora, podemos ver que

$$\begin{aligned}\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}^2} &= -\frac{\mathbb{1}_{\{0\}}(y_t)}{\delta_{0t}^2} + \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)^2}, \\ \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}^2} &= -\frac{\mathbb{1}_{\{1\}}(y_t)}{\delta_{0t}^2} + \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)^2}, \\ \frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t} \partial \delta_{0t}} &= \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)^2}.\end{aligned}$$

Logo, sob adequadas condições de regularidade dadas em Lehmann & Casella (2002, § 6.5), temos que

$$\begin{aligned}\mathbb{E}\left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \delta_{0t}}\right) &= 0, \\ \mathbb{E}\left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \delta_{1t}}\right) &= 0, \\ \mathbb{E}\left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{0t}^2}\right) &= -\frac{1}{\delta_{0t}} + \frac{1}{(1 - \alpha_t)}, \\ \mathbb{E}\left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t}^2}\right) &= -\frac{1}{\delta_{1t}} + \frac{1}{(1 - \alpha_t)}, \\ \mathbb{E}\left(\frac{\partial^2 \ell_t(\delta_{0t}, \delta_{1t})}{\partial \delta_{1t} \partial \delta_{0t}}\right) &= \frac{1}{(1 - \alpha_t)}.\end{aligned}$$

Para o cálculo dos cumulantes envolvendo o o vetor de parâmetros $(\beta^\top, \phi)^\top$ consideramos a proposição abaixo.

Proposição A.3.1. *Seja $(y_1, \dots, y_n)^\top$ um vetor de n variáveis aleatórias independentes com y_t seguindo a distribuição beta inflacionada em zero e um (2.4.7), i.e., $y_t \sim \text{BIZU}(\delta_{0t}, \delta_{1t}, \mu_t, \phi)$, $t = 1, \dots, n$ em que $\alpha_t = \delta_{0t} + \delta_{1t}$. Se $\mathfrak{F} : (0, 1) \rightarrow \mathbb{R}$ é uma função continua. Então,*

$$\mathbb{E}\left(\sum_{t: y_t \in (0,1)} \mathfrak{F}(y_t)\right) = \sum_{t=1}^n (1 - \alpha_t) \mathbb{E}(\mathfrak{F}(y_t) | \mathbb{1}_{(0,1)}(y_t) = 1)$$

Demonstração. Definimos

$$\mathfrak{F}^*(y_t) = \begin{cases} 0, & \text{se } y_t \in \{0, 1\}, \\ \mathfrak{F}(y_t), & \text{se } y_t \in (0, 1). \end{cases}$$

Então,

$$\mathbb{E}\left(\sum_{t: y_t \in (0,1)} \mathfrak{F}(y_t)\right) = \mathbb{E}\left(\sum_{t=1}^n \mathfrak{F}^*(y_t)\right) = \sum_{t=1}^n \mathbb{E}(\mathfrak{F}^*(y_t)).$$

Agora

$$\begin{aligned}\mathbb{E}(\mathfrak{F}^*(y_t)) &= \mathbb{E}(\mathfrak{F}^*(y_t) | \mathbb{1}_{(0,1)}(y_t) = 0) P(\mathbb{1}_{(0,1)}(y_t) = 0) + \mathbb{E}(\mathfrak{F}^*(y_t) | \mathbb{1}_{(0,1)}(y_t) = 1) P(\mathbb{1}_{(0,1)}(y_t) = 1) \\ &= \mathbb{E}(\mathfrak{F}^*(y_t) | \mathbb{1}_{(0,1)}(y_t) = 1) P(\mathbb{1}_{(0,1)}(y_t) = 1) = (1 - \alpha_t) \mathbb{E}(\mathfrak{F}(y_t) | \mathbb{1}_{(0,1)}(y_t) = 1).\end{aligned}$$

Daí segue o resultado. ✠

A.3. Apêndice 3

Agora vamos calcular o valor esperado das derivadas de segunda ordem da função de log-verossimilhança. Note que

$$\begin{aligned}
\kappa_{r's'} &= E(U_{r's'}) \\
&= \sum_{t=1}^n \left\{ \left(\left[-\frac{1}{\delta_{0t}} + \frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} v_{ts'} + \left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} v_{ts'} \right) \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} v_{tr'} \right. \\
&\quad \left. + \left(\left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} v_{ts'} + \left[-\frac{1}{\delta_{1t}} + \frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} v_{ts'} \right) \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} v_{tr'} \right\} \\
&= \sum_{t=1}^n \left[-\frac{1}{\delta_{0t}} + \frac{1}{(1-\alpha_t)} \right] \left(\frac{\partial \delta_{0t}}{\partial \zeta_{0t}} \right)^2 v_{tr'} v_{ts'} + 2 \sum_{t=1}^n \left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} v_{tr'} v_{ts'} \\
&\quad + \sum_{t=1}^n \left[-\frac{1}{\delta_{1t}} + \frac{1}{(1-\alpha_t)} \right] \left(\frac{\partial \delta_{1t}}{\partial \zeta_{0t}} \right)^2 v_{tr'} v_{ts'},
\end{aligned} \tag{A.3.10}$$

$$\begin{aligned}
\kappa_{r''s''} &= E(U_{r''s''}) = \\
&= \sum_{t=1}^n \left\{ \left(\left[-\frac{1}{\delta_{0t}} + \frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} z_{ts''} + \left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} z_{ts''} \right) \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} z_{tr''} \right. \\
&\quad \left. + \left(\left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} z_{ts''} + \left[-\frac{1}{\delta_{1t}} + \frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} z_{ts''} \right) \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} z_{tr''} \right\} \\
&= \sum_{t=1}^n \left[-\frac{1}{\delta_{0t}} + \frac{1}{(1-\alpha_t)} \right] \left(\frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \right)^2 z_{tr''} z_{ts''} + 2 \sum_{t=1}^n \left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} z_{tr''} z_{ts''} \\
&\quad + \sum_{t=1}^n \left[-\frac{1}{\delta_{1t}} + \frac{1}{(1-\alpha_t)} \right] \left(\frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \right)^2 z_{tr''} z_{ts''}
\end{aligned} \tag{A.3.11}$$

e

$$\begin{aligned}
\kappa_{r'r''} &= E(U_{r'r''}) \\
&= \sum_{t=1}^n \left[-\frac{1}{\delta_{0t}} + \frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} z_{tr''} v_{r'} + \sum_{t=1}^n \left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} z_{tr''} v_{r'} \\
&\quad + \sum_{t=1}^n \left[\frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} z_{r''} v_{r'} + \sum_{t=1}^n \left[-\frac{1}{\delta_{1t}} + \frac{1}{(1-\alpha_t)} \right] \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} z_{r''} v_{r'}.
\end{aligned} \tag{A.3.12}$$

Usando a Proposição A.3.1 podemos calcular o valor esperado das derivadas de segunda ordem da função $\ell_2(\beta, \phi)$. Assim,

$$\kappa_{rs} = E(U_{rs}) = \sum_{t=1}^n E(\mathbb{1}_{(0,1)}(y_t)) \left\{ -\phi^2 w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right\} x_{ts} x_{tr} = -\sum_{t=1}^n (1-\alpha_t) \left\{ \phi^2 w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right\} x_{ts} x_{tr}, \tag{A.3.13}$$

$$\kappa_{\phi\phi} = E(U_{\phi\phi}) = \sum_{t=1}^n -E(\mathbb{1}_{(0,1)}(y_t)) d_t = -\sum_{t=1}^n (1-\alpha_t) d_t, \tag{A.3.14}$$

A.4. Apêndice 4

$$\kappa_{r\phi} = E(U_{r\phi}) = \sum_{t=1}^n -E(\mathbb{1}_{(0,1)}(y_t)) \mathbf{c}_t \frac{d\mu_t}{d\eta_t} x_{tr} = - \sum_{t=1}^n (1 - \alpha_t) \mathbf{c}_t \frac{d\mu_t}{d\eta_t} x_{tr}, \quad (\text{A.3.15})$$

sendo que $w_t = \psi'(\mu_t\phi) + \psi'(1 - \mu_t)\phi$, $d_t = (1 - \mu_t)^2\psi'((1 - \mu_t)\phi) + \mu_t^2\psi'(\mu_t\phi) - \psi'(\phi)$ e $\mathbf{c}_t = \phi[\mu_t\psi'(\mu_t\phi) - (1 - \mu_t)\psi'((1 - \mu_t)\phi)]$ são as mesmas quantidades obtidas no Apêndice A.1.

Novamente, pela separabilidade dos parâmetros $(\rho^\top, \gamma^\top)^\top$ e $(\beta^\top, \phi)^\top$, temos que

$$\kappa_{rr'} = E(U_{rr'}) = 0, \quad (\text{A.3.16})$$

$$\kappa_{rr''} = E(U_{rr''}) = 0, \quad (\text{A.3.17})$$

$$\kappa_{\phi r'} = E(U_{\phi r'}) = 0, \quad (\text{A.3.18})$$

$$\kappa_{\phi r''} = E(U_{\phi r''}) = 0. \quad (\text{A.3.19})$$

A.4 Apêndice 4

Estimação por máxima verossimilhança do modelo RBIZU com dados simulados

```

library(gamlss)
library(gamlss.dist)

# Simulação do modelo RBIZU
tam=500 #tamanho de amostra

# Geração do preditor linear de mu
x0=rep(1,tam)
x1=runif(tam) # covariada x1 de mu uniforme
x.mu=cbind(x0,x1) # matriz de regressores de mu
par.mu=c(-1.5, 1.5) # parâmetros da regressão de mu
pred.mu=x.mu%*%par.mu # preditor linear de mu

# mu relacionado com preditor usando ligação logito
mu=exp(pred.mu)/(1+exp(pred.mu))
# Geração do preditor linear de rho
z0=rep(1,tam)
z1=runif(tam) # covariada z1 uniforme
v1=runif(tam) # covariada v1 uniforme
z.rho=cbind(z0,v1) # matriz de regressores de rho
z.gama=cbind(z0,z1) # matriz de regressores de gamma

par.rho=c(-1.0, 0.5) # parâmetros da regressão de rho
par.gama=c(-1.0, 0.7) # parâmetros da regressão de gamma
pred.rho=z.rho%*%par.rho # preditor linear de rho
pred.gama=z.gama%*%par.gama # preditor linear de gamma

# delta.zero relacionado com preditor
delta.zero=exp(pred.rho)/(1+exp(pred.rho)+exp(pred.gama))
# delta.um relacionado com preditor
delta.um=exp(pred.gama)/(1+exp(pred.rho)+exp(pred.gama))

# parâmetro de precisão

```

A.4. Apêndice 4

```

phi = 50 # phi constante para todas as observações
sigma.m= 1/sqrt(phi+1)

set.seed(121) #semente

#----gerador da variável resposta do modelo RBIZU ---#
ydata = mapply(rBEINF, 1, mu, sigma.m, delta.zero, delta.um)

###----- Ajuste do Modelo RBIZU usando GAMLSS -----###
# Stasinopoulos D. M., Rigby R.A. and Akantziliotou C. (2006)
# Instructions on how to use the GAMLSS package in R. Accompanying
# documentation in the current GAMLSS help files, (see also <URL:
# http://www.londonmet.ac.uk/gamlss/>).
###-----###

fit=gamlss(ydata~(-1+x.mu),sigma.formula=~1, nu.formula=~(-1+z.rho),
tau.formula=~(-1+z.gama), family=BEINF)

muhat = fit$mu.fv
nuhat = fit$nu.fv
tauhat = fit$tau.fv

delta.zerohat = nuhat/(1+nuhat+tauhat)
delta.umhat = tauhat/(1+nuhat+tauhat)
meany = meanBEINF(fit)

summary(fit) #Resumo de estatísticas

*****
Family: c("BEINF", "Beta Inflated")

Call:  gamlss(formula = ydata ~ (-1 + x.mu), sigma.formula = ~1,
nu.formula = ~(-1 +      z.rho), tau.formula = ~(-1 + z.gama), family = BEINF)

Fitting method: RS()

-----
Mu link function:  logit
Mu Coefficients:
      Estimate Std. Error t value  Pr(>|t|)
x.mux0    -1.491    0.03620  -41.17 1.152e-161
x.mux1     1.532    0.05832   26.27 8.645e-96

-----
Sigma link function:  logit
Sigma Coefficients:
      Estimate Std. Error  t value  Pr(>|t|)
-1.806e+00   4.309e-02  -4.192e+01 1.190e-164

-----
Nu link function:  log
Nu Coefficients:
      Estimate Std. Error t value  Pr(>|t|)
z.rhoz0   -1.7240    0.2577  -6.691 6.046e-11
z.rhov1    0.5198    0.4350   1.195 2.327e-01

```

A.5. Apêndice 5

```

-----
Tau link function:  log
Tau Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.gamaz0    -1.8454    0.2668   -6.918 1.429e-11
z.gamaz1     0.6156    0.4366    1.410 1.592e-01
-----

No. of observations in the fit:  500
Degrees of Freedom for the fit:   7
      Residual Deg. of Freedom: 493
                        at cycle: 4

Global Deviance:    -80.37269
      AIC:          -66.37269
      SBC:          -36.87043
*****
(1/fit$sigma.fv[1]^2) -1 # estimativa de sigma
> 49.23541

```

A.5 Apêndice 5

Fórmula geral de Cox & Snell (1968)

Como é descrito por Cordeiro (1999), os vieses de ordem n^{-1} dos estimadores de máxima verossimilhança (EMV) supondo observações independentes mas não necessariamente identicamente distribuídas, foram deduzidos por Cox & Snell (1968). Eles desenvolveram a técnica descrita a seguir para corrigir EMV. Consideremos o modelo estatístico $f(t, \theta)$, com $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$, e seja $\hat{\theta}$ o estimador de máxima verossimilhança de θ obtido como solução da equação $U(\theta) = 0$, com $\hat{U}_r = 0$, para $r = 1, \dots, k$, sendo \hat{U}_r obtido avaliando-se U_r em $\hat{\theta}$. Supondo válidas usuais condições de regularidade, expandimos em série de Taylor a componente $\hat{U}_r = 0$ até primeira ordem em torno de θ , assim, $\hat{U}_r = U_r + \sum U_{rs}(\hat{\theta}_s - \theta_s) + O_p(1)$, que equivale a $U_r + \sum U_{rs}(\hat{\theta}_s - \theta_s) + O_p(1) = 0$. Em notação matricial, $U = J(\hat{\theta} - \theta) + O_p(1)$. Observamos que $J = K + O_p(n^{1/2})$. Desta forma $U = K(\hat{\theta} - \theta) + O_p(1)$. Daí,

$$\hat{\theta} - \theta = K^{-1}U + O_p(n^{-1}). \quad (\text{A.5.1})$$

A fórmula anterior é importante para calcular os momentos e cumulantes de ordens superiores dos EMV. Novamente, se expandimos \hat{U}_r até segunda ordem, obtemos

$$\hat{U}_r = U_r + \sum_s U_{rs}(\hat{\theta}_s - \theta_s) + \frac{1}{2} \sum_{s,t} U_{rst}(\hat{\theta}_s - \theta_s)(\hat{\theta}_t - \theta_t) + o_p(1).$$

Tomado o valor esperado,

$$\sum_s E\{U_{rs}(\hat{\theta}_s - \theta_s)\} + \frac{1}{2} \sum_{s,t} E\{U_{rst}(\hat{\theta}_s - \theta_s)(\hat{\theta}_t - \theta_t)\} \approx 0,$$

A.6. Apêndice 6

o que equivale a

$$\sum_s \kappa_{rs} E(\hat{\theta}_s - \theta_s) + \sum_s \text{Cov}(U_{rs}, \hat{\theta}_s - \theta_s) + \frac{1}{2} \sum_{s,t} \kappa_{rst} (-\kappa^{st}) \approx 0, \quad (\text{A.5.2})$$

com $-\kappa^{rs} = \kappa^{r,s}$ representando o elemento (r, s) da inversa K^{-1} da matriz de informação de Fisher. Usando (A.5.1) segue-se que, até ordem $o(1)$,

$$\text{Cov}(U_{rs}, \hat{\theta}_s - \theta_s) = \text{Cov}\left(U_{rs}, -\sum_t \kappa^{st} U_t\right) = -\sum_t \text{Cov}(U_{rs}, U_t) \kappa^{st} = -\sum_t \kappa_{rs,t} \kappa^{st}.$$

Definindo o viés de ordem $O(n^{-1})$ de $\hat{\theta}_r$ por $B(\hat{\theta}_r)$ e substituindo $\text{Cov}(U_{rs}, \hat{\theta}_s - \theta_s)$ em (A.5.2) vem

$$\sum_s \kappa_{rs} B(\hat{\theta}_s) = \sum_{s,t} \kappa^{st} \left(\kappa_{rs,t} + \frac{1}{2} \kappa_{rst} \right) + o(1),$$

cujas inversões produzem até ordem $O(n^{-1})$, para $r = 1, \dots, k$,

$$B(\hat{\theta}_r) = \sum_{s,t,u} \kappa^{rs} \kappa^{tu} \left(\kappa_{st,u} + \frac{1}{2} \kappa_{stu} \right), \quad (\text{A.5.3})$$

que é uma fórmula geral para determinar o viés de ordem $O(n^{-1})$ em EMV em modelos multiparamétricos. Em muitas situações torna-se conveniente substituir $\kappa_{st,u} + \frac{1}{2} \kappa_{stu}$ por $\kappa_{st}^{(u)} - \frac{1}{2} \kappa_{stu}$, o que facilita expressar o viés em notação matricial. A grande utilidade de (A.5.3) é definir um estimador de máxima verossimilhança corrigido até ordem $O(n^{-1})$ dado por $\tilde{\theta}_r = \hat{\theta}_r - \hat{B}(\hat{\theta}_r)$, onde $\hat{B}(\cdot)$ é o viés $B(\cdot)$ avaliado em $\hat{\theta}$. Este novo estimador $\tilde{\theta}_r$ tem viés de ordem $O(n^{-2})$, pois $E(\tilde{\theta}_r) = \theta_r + O(n^{-2})$, e pode ser preferido em relação a $\hat{\theta}_r$, cujo viés é de ordem $O(n^{-1})$. Diremos que $\tilde{\theta}$ é um estimador de máxima verossimilhança corrigido "corretivamente". O cálculo de momentos e cumulantes dos estimadores de máxima verossimilhança de ordem superior, como, por exemplo, $E(\hat{\theta})$ e $\text{Var}(\hat{\theta})$, até ordem $O(n^{-2})$, é bastante complexo, já que precisamos incluir outros termos de ordem superior na expansão (A.5.2).

Observa-se que $B(\hat{\theta})$ pode ser expresso na forma

$$B(\hat{\theta}) = K(\theta)^{-1} A \text{vec}\left(K(\theta)^{-1}\right) + O(n^{-2}),$$

onde $A = \left[A^{(1)} | \dots | A^{(k)} \right]$ é uma matriz $k \times k^2$ particionada, com $A^{(l)} = (a_{jm}^{(l)})$ tendo elemento típico

$$a_{jm}^{(l)} = \kappa_{jm}^{(l)} - \frac{1}{2} \kappa_{jml}, \quad j, m, l = 1, \dots, k,$$

e $\text{vec}(\cdot)$ é o operador que transforma uma matriz em vetor coluna sobrepondo as colunas desta matriz.

A.6 Apêndice 6

Viés do estimador de máxima verossimilhança de γ

A.6. Apêndice 6

Para $R = 1, \dots, M$, os momentos de terceira ordem com respeito a γ são obtidos pela diferenciação da função $\ell_1(\gamma)$ com respeito aos parâmetros de interesse. Assim,

$$\begin{aligned}
U_{RSU} &= \frac{\partial^3 \ell_1(\gamma)}{\partial \gamma_R \gamma_S \gamma_U} \\
&= \sum_{t=1}^n \frac{\partial}{\partial \alpha_t} \left(\left\{ \frac{\partial^2 \ell_t(\alpha_t)}{\partial \alpha_t^2} \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 + \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \frac{\partial}{\partial \alpha_t} \left(\frac{d\alpha_t}{d\zeta_t} \right) \frac{\partial \zeta_t}{\partial \gamma_R} \right\} z_{tS} z_{tR} \right) \frac{d\alpha_t}{d\zeta_t} \frac{\partial \zeta_t}{\partial \gamma_U} \\
&= \sum_{t=1}^n \left\{ \frac{\partial^3 \ell_t(\alpha_t)}{\partial \alpha_t^3} \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 + 2 \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{\partial^2 \ell_t(\alpha_t)}{\partial \alpha_t^2} \frac{d\alpha_t}{d\zeta_t} + \left(\frac{\partial^2}{\partial \alpha_t^2} \frac{d\alpha_t}{d\zeta_t} \right) \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right. \\
&\quad \left. + \frac{\partial^2 \ell_t(\alpha_t)}{\partial \alpha_t^2} \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} + \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right)^2 \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU}. \\
&= \sum_{t=1}^n \left\{ \left(\frac{-2(1 - \mathbb{1}_{\{c\}}(y_t))}{(1 - \alpha_t)^3} + \frac{2\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^3} \right) \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 + 3 \left(\frac{-(1 - \mathbb{1}_{\{c\}}(y_t))}{(1 - \alpha_t)^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^2} \right) \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right. \\
&\quad \left. + \left(\frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t} - \frac{(1 - \mathbb{1}_{\{c\}}(y_t))}{(1 - \alpha_t)} \right) \left[\left(\frac{\partial^2}{\partial \alpha_t^2} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} + \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right)^2 \right] \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU}.
\end{aligned}$$

Os cumulantes de terceira ordem com respeito a γ são obtidos pelo cálculo do valor esperado dos momentos acima. Para $R = 1, \dots, M$, temos

$$\begin{aligned}
\kappa_{RSU} &= E(U_{RSU}) = E \left(\sum_{t=1}^n \left\{ \left(\frac{-2(1 - \mathbb{1}_{\{c\}}(y_t))}{(1 - \alpha_t)^3} + \frac{2\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^3} \right) \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 \right. \right. \\
&\quad \left. \left. + 3 \left(\frac{-(1 - \mathbb{1}_{\{c\}}(y_t))}{(1 - \alpha_t)^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^2} \right) \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right. \right. \\
&\quad \left. \left. + \left(\frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t} - \frac{(1 - \mathbb{1}_{\{c\}}(y_t))}{(1 - \alpha_t)} \right) \left[\left(\frac{\partial^2}{\partial \alpha_t^2} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} + \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right)^2 \right] \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU} \right) \\
&= \sum_{t=1}^n \left\{ 2p_t^* \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 - 3p_t \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU},
\end{aligned} \tag{A.6.1}$$

em que $p_t^* = \{-1/(1 - \alpha_t)^2\} + \{1/\alpha_t^2\}$. Agora, derivamos o cumulante de segunda ordem (A.1.6) com respeito aos parâmetros de interesse, i.e., para $U = 1, \dots, M$,

$$\begin{aligned}
\kappa_{RS}^{(U)} &= \frac{\partial \kappa_{RS}}{\partial \gamma_U} = \frac{\partial}{\partial \alpha_t} \left(\sum_{t=1}^n -p_t \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 z_{tS} z_{tR} \right) \frac{d\alpha_t}{d\zeta_t} \frac{\partial \zeta_t}{\partial \gamma_U} \\
&= \sum_{t=1}^n \left\{ p_t^* \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 - 2p_t \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU}.
\end{aligned} \tag{A.6.2}$$

A.6. Apêndice 6

Novamente, pela ortogonalidade global entre γ e $(\beta^\top, \phi)^\top$ obtemos

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

Utilizando os cumulantes (A.6.1) e (A.6.2) calculamos a quantidade

$$\begin{aligned} \kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU} &= \sum_{t=1}^n \left\{ p_t^* \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 - 2p_t \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU} \\ &\quad - \frac{1}{2} \sum_{t=1}^n \left\{ 2p_t^* \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 - 3p_t \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU} \\ &= -\frac{1}{2} \sum_{t=1}^n \left\{ p_t \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU}. \end{aligned}$$

Seja $W_0 = \text{diag}\{w_{01}, \dots, w_{0n}\}$ a matriz diagonal de dimensão $n \times n$ com

$$w_{0t} = \frac{1}{2} p_t \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \left(\frac{d\alpha_t}{d\zeta_t} \right)^2. \quad (\text{A.6.3})$$

Então,

$$\kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU} = \sum_{t=1}^n w_{0t} z_{tR} z_{tS} z_{tU}. \quad (\text{A.6.4})$$

Conseqüentemente, a fórmula de Cox & Snell (1968) para calcular o viés de segunda ordem do estimador de máxima verossimilhança para a b -ésima componente do vetor γ é

$$\begin{aligned} B(\hat{\gamma}_b) &= \sum_{R,S,U} \kappa^{bR} \kappa^{SU} \left\{ \kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU} \right\} = \sum_{R,S,U} \kappa^{bR} \kappa^{SU} \sum_{t=1}^n w_{0t} z_{tR} z_{tS} z_{tU} \\ &= \sum_{t=1}^n w_{0t} \sum_R \kappa^{bR} z_{tR} \sum_{S,U} z_{tS} \kappa^{SU} z_{tU}. \end{aligned}$$

Se definimos \mathcal{e}_b como sendo o b -ésimo vetor coluna da matriz identidade de dimensão $M \times M$, temos que

$$\begin{aligned} \sum_{t=1}^n w_{0t} \sum_R \kappa^{bR} z_{tR} \sum_{S,U} z_{tS} \kappa^{SU} z_{tU} &= \sum_{t=1}^n w_{0t} \sum_R \kappa^{bR} z_{tR} (z_t^\top K^{\gamma\gamma} z_t) \\ &= \sum_{t=1}^n w_{0t} \mathcal{e}_b^\top K^{\gamma\gamma} z_t (z_t^\top K^{\gamma\gamma} z_t) \\ &= \mathcal{e}_b^\top K^{\gamma\gamma} \sum_{t=1}^n w_{0t} z_t (z_t^\top K^{\gamma\gamma} z_t), \end{aligned}$$

em que z_t é o vetor linha obtido da t -ésima linha da matriz Z . Se definimos $\delta_{\gamma\gamma}$ o vetor de dimensão $n \times 1$ obtido da

A.7. Apêndice 7

diagonal principal da matriz $ZK^{\gamma\gamma}Z^\top$, temos que

$$B(\hat{\gamma}_b) = \sum_{R,S,U} \kappa^{bR} \kappa^{SU} \left\{ \kappa_{RS}^{(U)} - \frac{1}{2} \kappa_{RSU} \right\} = \mathcal{L}_b^\top K^{\gamma\gamma} Z^\top W_0 \delta_{\gamma\gamma}. \quad (\text{A.6.5})$$

A.7 Apêndice 7

Viés do estimador de máxima verossimilhança de β e ϕ

Os momentos de terceira ordem com respeito a β e ϕ são obtidos pela diferenciação da função $\ell_2(\beta, \phi)$ com respeito aos parâmetros de interesse. Assim, para $r, s, u = 1, \dots, m$, temos

$$\begin{aligned} U_{rsu} &= \frac{\partial^3 \ell_2(\beta, \phi)}{\partial \beta_r \partial \beta_s \partial \beta_u} \\ &= \sum_{t: y_t \in (0,1)} \frac{\partial}{\partial \mu_t} \left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} \frac{\partial \eta_t}{\partial \beta_s} \frac{\partial \eta_t}{\partial \beta_u}, \\ U_{rsu} &= \sum_{t: y_t \in (0,1)} \left\{ \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^3} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + 2 \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right. \\ &\quad + \left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} + \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \\ &\quad \left. + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \right\} \frac{d\mu_t}{d\eta_t} x_{ts} x_{tr} x_{tu}. \end{aligned}$$

Note que

$$\frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^3} = \frac{\partial(-\phi^2 w_t)}{\partial \mu_t} = -\phi^2 \frac{\partial w_t}{\partial \mu_t},$$

com

$$\frac{\partial w_t}{\partial \mu_t} = \phi[\psi''(\mu_t \phi) - \psi''((1 - \mu_t)\phi)].$$

Se definimos $m_t = [\psi''(\mu_t \phi) - \psi''((1 - \mu_t)\phi)]$, obtemos $\partial^3 \ell_t(\mu_t, \phi) / \partial \mu_t^3 = -\phi^3 m_t$. Desta forma,

$$\begin{aligned} U_{rsu} &= \sum_{t: y_t \in (0,1)} \left\{ -\phi^3 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 - 2\phi^2 w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \phi[y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} \right. \\ &\quad \left. - \phi^2 w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \phi[y_t^* - \mu_t^*] \right\} \frac{d\mu_t}{d\eta_t} x_{ts} x_{tr} x_{tu}, \end{aligned}$$

A.7. Apêndice 7

$$\begin{aligned}
U_{rsu} &= \sum_{t:y_t \in (0,1)} \left\{ -\phi^3 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 - 3\phi^2 w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right. \\
&\quad \left. + \phi [y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} \left[\left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \right] \right\} x_{ts} x_{tr} x_{tu}, \\
U_{rsu} &= -\phi \sum_{t:y_t \in (0,1)} \left\{ \phi^2 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + 3\phi w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right. \\
&\quad \left. - [y_t^* - \mu_t^*] \frac{d\mu_t}{d\eta_t} \left[\left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \right] \right\} x_{ts} x_{tr} x_{tu}.
\end{aligned}$$

Sejam

$$\begin{aligned}
a_t &= 3 \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2, \\
b_t &= \frac{d\mu_t}{d\eta_t} \left[\left(\frac{\partial^2}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)^2 \right].
\end{aligned} \tag{A.7.1}$$

Portanto

$$U_{rsu} = -\phi \sum_{t:y_t \in (0,1)} \left\{ \phi^2 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \phi w_t a_t - [y_t^* - \mu_t^*] b_t \right\} x_{ts} x_{tr} x_{tu}.$$

Também,

$$\begin{aligned}
U_{rs\phi} &= \frac{\partial^3 \ell_2(\beta, \phi)}{\partial \beta_r \partial \beta_s \partial \phi} = \sum_{t:y_t \in (0,1)} \frac{\partial}{\partial \phi} \left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right) \frac{\partial \eta_t}{\partial \beta_r} \frac{\partial \eta_t}{\partial \beta_s} \\
&= \sum_{t:y_t \in (0,1)} \left\{ \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^2 \partial \phi} \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \right\} x_{tr} x_{ts}, \\
\frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t^2 \partial \phi} &= \frac{\partial}{\partial \phi} \{-\phi^2 w_t\} = -2\phi w_t - \phi^2 \frac{\partial w_t}{\partial \phi} \\
&= -2\phi w_t - \phi^2 [\mu_t m_t + \psi''((1 - \mu_t)\phi)] \\
&= -\phi \{2w_t + \phi [\mu_t m_t + \psi''((1 - \mu_t)\phi)]\}.
\end{aligned}$$

Definimos

$$u_t = -\phi \{2w_t + \phi [\mu_t m_t + \psi''((1 - \mu_t)\phi)]\}. \tag{A.7.2}$$

Logo,

$$U_{rs\phi} = \sum_{t:y_t \in (0,1)} \left\{ (u_t \left(\frac{d\mu_t}{d\eta_t} \right) + [y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right)) \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts} \right\}.$$

A.7. Apêndice 7

Agora,

$$U_{r\phi\phi} = \frac{\partial^3 \ell_2(\beta, \phi)}{\partial \beta_r \partial \phi^2} = \sum_{t:y_t \in (0,1)} \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi^2} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r},$$

sendo

$$\begin{aligned} \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi^2} &= \frac{\partial}{\partial \phi} \left\{ [y_t^* - \mu_t^*] - \phi \frac{\partial \mu_t^*}{\partial \phi} \right\} = -2 \frac{\partial \mu_t^*}{\partial \phi} - \phi \frac{\partial^2 \mu_t^*}{\partial \phi^2} \\ &= -2 \frac{\partial \mu_t^*}{\partial \phi} - \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \\ &= -2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] - \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)]. \end{aligned}$$

Definimos

$$\tau_t = 2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \frac{d\mu_t}{d\eta_t}. \quad (\text{A.7.3})$$

Daí,

$$U_{r\phi\phi} = - \sum_{t:y_t \in (0,1)} \tau_t x_{tr}.$$

Finalmente,

$$U_{\phi\phi\phi} = \frac{\partial^3 \ell_2(\beta, \phi)}{\partial \phi^3} = \sum_{t:y_t \in (0,1)} \frac{\partial^3 \ell_t(\mu_t, \phi)}{\partial \phi^3} = - \sum_{t:y_t \in (0,1)} \{ (1 - \mu_t)^3 \psi''((1 - \mu_t)\phi) + \mu_t^3 \psi''(\mu_t \phi) - \psi''(\phi) \}.$$

Seja

$$s_t = (1 - \mu_t)^3 \psi''((1 - \mu_t)\phi) + \mu_t^3 \psi''(\mu_t \phi) - \psi''(\phi). \quad (\text{A.7.4})$$

Logo, $s_t = \frac{\partial d_t}{\partial \phi}$. Assim,

$$U_{\phi\phi\phi} = - \sum_{t:y_t \in (0,1)} s_t.$$

Os cumulantes de terceira ordem com respeito a β e ϕ são obtidos pelo cálculo do valor esperado dos momentos acima. Consideramos U_{rsu} , $U_{r\phi\phi}$, $U_{rs\phi}$, $U_{\phi\phi\phi}$ etc. funções contínuas da variável aleatória y . Assim, ao utilizarmos a Proposição A.1.1, temos

$$\begin{aligned} \kappa_{rsu} &= E(U_{rsu}) = E \left(-\phi \sum_{t:y_t \in (0,1)} \left\{ \phi^2 m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \phi w_t a_t - [y_t^* - \mu_t^*] b_t \right\} x_{ts} x_{tr} x_{tu} \right) \\ &= -\phi^2 \sum_{t=1}^n (1 - \alpha_t) \left\{ \phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + w_t a_t \right\} x_{ts} x_{tr} x_{tu}, \\ \kappa_{\phi\phi\phi} &= E(U_{\phi\phi\phi}) = E \left(\sum_{t:y_t \in (0,1)} -s_t \right) = - \sum_{t=1}^n (1 - \alpha_t) s_t. \end{aligned}$$

A.7. Apêndice 7

Finalmente,

$$\begin{aligned}\kappa_{rs\phi} &= E(U_{rs\phi}) = E\left(\sum_{t:y_t \in (0,1)} \left\{u_t \left(\frac{d\mu_t}{d\eta_t}\right) + [y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right) - \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right)\right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts}\right) \\ &= \sum_{t=1}^n (1 - \alpha_t) \left\{u_t \left(\frac{d\mu_t}{d\eta_t}\right) - \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right)\right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts}\end{aligned}$$

e

$$\kappa_{r\phi\phi} = E(U_{r\phi\phi}) = E\left(\sum_{t:y_t \in (0,1)} -\mathbf{r}_t x_{tr}\right) = -\sum_{t=1}^n (1 - \alpha_t) \mathbf{r}_t x_{tr}.$$

Derivando os cumulantes de segunda ordem de $\ell_2(\beta, \phi)$ com respeito aos parâmetros de interesse, temos

$$\begin{aligned}\kappa_{rs}^{(u)} &= \frac{\partial \kappa_{rs}}{\partial \beta_u} = \frac{\partial}{\partial \mu_t} \left(-\phi^2 \sum_{t=1}^n (1 - \alpha_t) w_t \left(\frac{d\mu_t}{d\eta_t}\right)^2 x_{ts} x_{tr}\right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_u} \\ &= -\phi^2 \sum_{t=1}^n (1 - \alpha_t) \left(\frac{\partial w_t}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_t}\right)^2 + 2w_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right) \frac{d\mu_t}{d\eta_t}\right) \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts} x_{tu} \\ &= -\phi^2 \sum_{t=1}^n (1 - \alpha_t) \left(\phi m_t \left(\frac{d\mu_t}{d\eta_t}\right)^3 + \frac{2}{3} w_t a_t\right) x_{tr} x_{ts} x_{tu},\end{aligned}$$

$$\begin{aligned}\kappa_{rs}^{(\phi)} &= \frac{\partial \kappa_{rs}}{\partial \phi} = \frac{\partial}{\partial \phi} \left(-\phi^2 \sum_{t=1}^n (1 - \alpha_t) w_t \left(\frac{d\mu_t}{d\eta_t}\right)^2 x_{ts} x_{tr}\right) \\ &= -2\phi \sum_{t=1}^n (1 - \alpha_t) w_t \left(\frac{d\mu_t}{d\eta_t}\right)^2 x_{ts} x_{tr} - \phi^2 \sum_{t=1}^n (1 - \alpha_t) [\mu_t m_t + \psi''((1 - \mu_t)\phi)] \left(\frac{d\mu_t}{d\eta_t}\right)^2 x_{ts} x_{tr} \\ &= -\phi \sum_{t=1}^n (1 - \alpha_t) \{2w_t + \phi[\mu_t m_t + \psi''((1 - \mu_t)\phi)]\} \left(\frac{d\mu_t}{d\eta_t}\right)^2 x_{ts} x_{tr}.\end{aligned}$$

Logo,

$$\kappa_{rs}^{(\phi)} = \sum_{t=1}^n (1 - \alpha_t) \left\{u_t \left(\frac{d\mu_t}{d\eta_t}\right)^2\right\} x_{ts} x_{tr},$$

$$\begin{aligned}\kappa_{r\phi}^{(u)} &= \frac{\partial \kappa_{r\phi}}{\partial \beta_u} = \frac{\partial}{\partial \mu_t} \left(-\sum_{t=1}^n (1 - \alpha_t) \mathbf{c}_t \frac{d\mu_t}{d\eta_t} x_{tr}\right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_u} \\ &= -\sum_{t=1}^n (1 - \alpha_t) \left\{\frac{\partial \mathbf{c}_t}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} + \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t}\right)\right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu}.\end{aligned}$$

A.7. Apêndice 7

Como

$$\begin{aligned}\frac{\partial \mathbf{c}_t}{\partial \mu_t} &= \frac{\partial}{\partial \mu_t} \phi[\mu_t w_t - \psi'((1 - \mu_t)\phi)] \\ &= \phi \left(w_t + \mu_t \frac{\partial w_t}{\partial \mu_t} + \phi \psi''((1 - \mu_t)\phi) \right) \\ &= \phi(w_t + \mu_t \phi m_t + \phi \psi''((1 - \mu_t)\phi)),\end{aligned}$$

temos

$$\kappa_{r\phi}^{(u)} = - \sum_{t=1}^n (1 - \alpha_t) \left\{ [\phi(w_t + \mu_t \phi m_t + \phi \psi''((1 - \mu_t)\phi))] \frac{d\mu_t}{d\eta_t} + \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu}.$$

Temos ainda

$$\kappa_{r\phi}^{(\phi)} = \frac{\partial \kappa_{r\phi}}{\partial \phi} = \frac{\partial}{\partial \phi} \left(- \sum_{t=1}^n (1 - \alpha_t) \mathbf{c}_t \frac{d\mu_t}{d\eta_t} x_{tr} \right) = - \sum_{t=1}^n (1 - \alpha_t) \frac{\partial \mathbf{c}_t}{\partial \phi} \frac{d\mu_t}{d\eta_t} x_{tr}.$$

Seja

$$\begin{aligned}\frac{\partial \mathbf{c}_t}{\partial \phi} &= \frac{\partial}{\partial \phi} \{ \phi[\mu_t w_t - \psi'((1 - \mu_t)\phi)] \} \\ &= [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi \left\{ \mu_t \frac{\partial w_t}{\partial \phi} - (1 - \mu_t) \psi''((1 - \mu_t)\phi) \right\} \\ &= [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi \{ \mu_t [\mu_t m_t + \psi''((1 - \mu_t)\phi)] - (1 - \mu_t) \psi''((1 - \mu_t)\phi) \}. \\ &= [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)].\end{aligned}$$

Definimos

$$\mathfrak{z}_t = [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)]. \quad (\text{A.7.5})$$

Então,

$$\begin{aligned}\kappa_{r\phi}^{(\phi)} &= - \sum_{t=1}^n (1 - \alpha_t) \mathfrak{z}_t \frac{d\mu_t}{d\eta_t} x_{tr}, \\ \kappa_{\phi\phi}^{(u)} &= \frac{\partial \kappa_{\phi\phi}}{\partial \beta_u} = \frac{\partial}{\partial \mu_t} \left(- \sum_{t=1}^n (1 - \alpha_t) d_t \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_u} = - \sum_{t=1}^n (1 - \alpha_t) \frac{\partial d_t}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} x_{tu}, \\ \frac{\partial d_t}{\partial \mu_t} &= \frac{\partial}{\partial \mu_t} \{ (1 - \mu_t)^2 \psi'((1 - \mu_t)\phi) + \mu_t^2 \psi'(\mu_t \phi) - \psi'(\phi) \} \\ &= -2(1 - \mu_t) \psi'((1 - \mu_t)\phi) - \phi (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi) \\ &\quad + 2\mu_t \psi'(\mu_t \phi) + \phi \mu_t^2 \psi''(\mu_t \phi). \\ &= 2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi [\mu_t^2 \psi''(\mu_t \phi) \\ &\quad - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] = \mathbf{r}_t.\end{aligned}$$

Desta forma,

$$\kappa_{\phi\phi}^{(u)} = \sum_{t=1}^n -(1 - \alpha_t) \mathbf{r}_t \frac{d\mu_t}{d\eta_t} x_{tu}.$$

A.7. Apêndice 7

Finalmente,

$$\kappa_{\phi\phi}^{(\phi)} = \frac{\partial \kappa_{\phi\phi}}{\partial \phi} = \frac{\partial}{\partial \phi} \left(- \sum_{t=1}^n (1 - \alpha_t) d_t \right) = - \sum_{t=1}^n (1 - \alpha_t) \frac{\partial d_t}{\partial \phi} = - \sum_{t=1}^n (1 - \alpha_t) s_t.$$

Utilizando os cumulantes de terceira ordem de $\ell_2(\beta, \phi)$ e as derivadas dos cumulantes de segunda ordem de $\ell_2(\beta, \phi)$, calculamos as seguintes quantidades

$$\begin{aligned} \kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} &= -\phi^2 \sum_{t=1}^n (1 - \alpha_t) \left\{ \phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \frac{2}{3} w_t a_t \right\} x_{tr} x_{ts} x_{tu} \\ &+ \frac{1}{2} \phi^2 \sum_{t=1}^n (1 - \alpha_t) \left\{ \phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + w_t a_t \right\} x_{ts} x_{tr} x_{tu} \\ &= \sum_{t=1}^n (1 - \alpha_t) \frac{-\phi^2}{2} \left\{ \phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \frac{1}{3} w_t a_t \right\} x_{ts} x_{tr} x_{tu}. \end{aligned}$$

Seja $W_1 = \text{diag}\{w_{11}, \dots, w_{1n}\}$ uma matriz diagonal de dimensão $n \times n$ com elemento típico

$$w_{1t} = -\frac{\phi^2}{2} \left[\phi m_t \left(\frac{d\mu_t}{d\eta_t} \right)^3 + \frac{1}{3} w_t a_t \right],$$

e seja uma matriz diagonal $\Delta = \text{diag}\{\delta_1, \dots, \delta_n\}$, denominada de ponderações, em que $\delta_t = 1 - \alpha_t$ é a probabilidade de y_t ser observada no intervalo $(0,1)$. Então

$$\kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} = \sum_{t=1}^n \delta_t w_{1t} x_{tr} x_{ts} x_{tu}. \quad (\text{A.7.6})$$

Agora, calculamos

$$\begin{aligned} \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} &= \sum_{t=1}^n (1 - \alpha_t) \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 \right\} x_{ts} x_{tr} \\ &- \frac{1}{2} \sum_{t=1}^n (1 - \alpha_t) \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) - c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts} \\ &= \sum_{t=1}^n \frac{1}{2} (1 - \alpha_t) \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) + c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{ts}. \end{aligned}$$

Se definimos $W_2 = \text{diag}\{w_{21}, \dots, w_{2n}\}$ com

$$w_{2t} = \frac{1}{2} \left\{ u_t \left(\frac{d\mu_t}{d\eta_t} \right) + c_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t},$$

temos

$$\kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} = \sum_{t=1}^n \delta_t w_{2t} x_{tr} x_{ts}. \quad (\text{A.7.7})$$

A.7. Apêndice 7

Adicionalmente,

$$\begin{aligned}
\kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u} &= \\
&\sum_{t=1}^n (1 - \alpha_t) \left\{ -\phi \{w_t + \phi[\mu_t m_t + \psi''((1 - \mu)\phi)]\} \frac{d\mu_t}{d\eta_t} - \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu} \\
&- \frac{1}{2} \sum_{t=1}^n (1 - \alpha_t) \left\{ -\phi \{2w_t + \phi[\mu_t m_t + \psi''((1 - \mu)\phi)]\} \frac{d\mu_t}{d\eta_t} - \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu} \\
&= \sum_{t=1}^n (1 - \alpha_t) \left\{ -\frac{1}{2} \phi^2 [\mu_t m_t + \psi''((1 - \mu)\phi)] \frac{d\mu_t}{d\eta_t} - \frac{1}{2} \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{tr} x_{tu}.
\end{aligned}$$

Definimos $W_3 = \text{diag}\{w_{31}, \dots, w_{3n}\}$ a matriz diagonal de dimensão $n \times n$ com

$$w_{3t} = -\frac{1}{2} \left\{ \phi^2 [\mu_t m_t + \psi''((1 - \mu)\phi)] \frac{d\mu_t}{d\eta_t} + \mathbf{c}_t \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t},$$

podemos escrever

$$\kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u} = \sum_{t=1}^n \delta_t w_{3t} x_{tr} x_{tu}. \quad (\text{A.7.8})$$

De forma análoga,

$$\kappa_{\phi r}^{(u)} - \frac{1}{2}\kappa_{\phi r u} = \kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u} = \sum_{t=1}^n \delta_t w_{3t} x_{tr} x_{tu}. \quad (\text{A.7.9})$$

Agora,

$$\begin{aligned}
\kappa_{r\phi}^{(\phi)} - \frac{1}{2}\kappa_{r\phi\phi} &= \\
&= \sum_{t=1}^n -(1 - \alpha_t) \left\{ [\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi[\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \right\} \frac{d\mu_t}{d\eta_t} x_{tr} \\
&- \frac{1}{2} \sum_{t=1}^n -(1 - \alpha_t) \left\{ 2[\mu_t w_t - \psi'((1 - \mu_t)\phi)] + \phi[\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \right\} \frac{d\mu_t}{d\eta_t} x_{tr} \\
&= \sum_{t=1}^n -\frac{1}{2} (1 - \alpha_t) \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \frac{d\mu_t}{d\eta_t} x_{tr}.
\end{aligned}$$

Definimos $W_4 = \text{diag}\{w_{41}, \dots, w_{4n}\}$ a matriz diagonal de dimensão $n \times n$ com

$$w_{4t} = \frac{-1}{2} \phi [\mu_t^2 \psi''(\mu_t \phi) - (1 - \mu_t)^2 \psi''((1 - \mu_t)\phi)] \frac{d\mu_t}{d\eta_t}.$$

A.7. Apêndice 7

Logo,

$$\kappa_{r\phi}^{(\phi)} - \frac{1}{2}\kappa_{r\phi\phi} = \sum_{t=1}^n \delta_t w_{4t} x_{tr}. \quad (\text{A.7.10})$$

Finalmente,

$$\begin{aligned} \kappa_{\phi\phi}^{(u)} - \frac{1}{2}\kappa_{\phi\phi u} &= \sum_{t=1}^n -(1 - \alpha_t) \mathbf{r}_t \frac{d\mu_t}{d\eta_t} x_{tu} - \frac{1}{2} \sum_{t=1}^n -(1 - \alpha_t) \mathbf{r}_t \frac{d\mu_t}{d\eta_t} x_{tu} \\ &= \sum_{t=1}^n -\frac{1}{2} (1 - \alpha_t) \mathbf{r}_t \frac{d\mu_t}{d\eta_t} x_{tu}, \end{aligned}$$

Se definimos $W_5 = \text{diag}\{w_{51}, \dots, w_{5n}\}$ a matriz diagonal de dimensão $n \times n$ com

$$w_{5t} = -\frac{1}{2} \mathbf{r}_t \frac{d\mu_t}{d\eta_t},$$

temos que

$$\kappa_{\phi\phi}^{(u)} - \frac{1}{2}\kappa_{\phi\phi u} = \sum_{t=1}^n \delta_t w_{5t} x_{tu}, \quad (\text{A.7.11})$$

$$\kappa_{\phi\phi}^{(\phi)} - \frac{1}{2}\kappa_{\phi\phi\phi} = \sum_{t=1}^n -(1 - \alpha_t) s_t - \frac{1}{2} \sum_{t=1}^n -(1 - \alpha_t) s_t = \sum_{t=1}^n -\frac{1}{2} \delta_t s_t. \quad (\text{A.7.12})$$

Com as quantidades calculadas acima podemos obter a fórmula de Cox & Snell (1968) para o cálculo do viés de segunda ordem do estimador de máxima verossimilhança do a -ésimo componente do vetor $\widehat{\beta}$, a saber:

$$\begin{aligned} B(\widehat{\beta}_a) &= \sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(t)} - \frac{1}{2} \kappa_{rsu} \right\} + \kappa^{a\phi} \sum_{s,u} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} \\ &+ \sum_{r,u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} + \sum_{r,s} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} \\ &+ \kappa^{a\phi} \sum_u \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} + \kappa^{a\phi} \sum_s \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} \\ &+ \kappa^{\phi\phi} \sum_r \kappa^{ar} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} + \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\}. \end{aligned}$$

Pode-se observar de 3.2.19 que as entradas da matriz $K^{\beta\phi}$ não são nulas, o que leva a uma complicação do cálculo, indicando que todos os termos da expansão anterior devem ser calculados.

A.7. Apêndice 7

Assim, para $r = 1, \dots, m$ temos

$$\begin{aligned} \sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} \right\} &= \sum_{r,s,u} \kappa^{ar} \kappa^{su} \sum_{t=1}^n \delta_t w_{1t} x_{tr} x_{ts} x_{tu} \\ &= \sum_{t=1}^n \delta_t w_{1t} \sum_r \kappa^{ar} x_{tr} \sum_{s,u} x_{ts} \kappa^{su} x_{tu}. \end{aligned}$$

Se definimos $\underline{\ell}_a$ como sendo o a -ésimo vetor coluna da matriz identidade de dimensão $m \times m$, temos que

$$\begin{aligned} \sum_{t=1}^n \delta_t w_{1t} \sum_r \kappa^{ar} x_{tr} \sum_{s,u} x_{ts} \kappa^{su} x_{tu} &= \sum_{t=1}^n \delta_t w_{1t} \sum_r \kappa^{ar} x_{tr} (x_t^\top K^{\beta\beta} x_t) \\ &= \sum_{t=1}^n \delta_t w_{1t} \underline{\ell}_a^\top K^{\beta\beta} x_t (x_t^\top K^{\beta\beta} x_t) \\ &= \underline{\ell}_a^\top K^{\beta\beta} \sum_{t=1}^n \delta_t w_{1t} x_t (x_t^\top K^{\beta\beta} x_t), \end{aligned}$$

em que x_t é o vetor linha obtido da t -ésima linha da matriz X , $K^{\beta\beta}$ é a inversa da informação de Fisher de β obtida de 3.2.19 e $\delta_{\beta\beta}$ é o vetor de dimensão $n \times 1$ obtido da diagonal principal da matriz $X K^{\beta\beta} X^\top$. Assim,

$$\sum_{r,s,u} \kappa^{ar} \kappa^{su} \left\{ \kappa_{rs}^{(u)} - \frac{1}{2} \kappa_{rsu} \right\} = \underline{\ell}_a^\top K^{\beta\beta} X^\top \Delta W_1 \delta_{\beta\beta}. \quad (\text{A.7.13})$$

Agora,

$$\begin{aligned} \sum_{\phi,s,u} \kappa^{a\phi} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} &= \kappa^{a\phi} \sum_{s,u} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} \\ &= \kappa^{a\phi} \sum_{s,u} \kappa^{su} \sum_{t=1}^n \delta_t w_{3t} x_{ts} x_{tu} = \kappa^{a\phi} \sum_{t=1}^n \delta_t w_{3t} \sum_{s,u} x_{ts} \kappa^{su} x_{tu} \\ &= \kappa^{a\phi} \sum_{t=1}^n \delta_t w_{3t} (x_t^\top K^{\beta\beta} x_t) = \kappa^{a\phi} \text{tr}(\Delta W_3 X K^{\beta\beta} X^\top), \end{aligned}$$

onde $\text{tr}(\cdot)$ é o operador traço de uma matriz quadrada. Logo,

$$\sum_{\phi,s,u} \kappa^{a\phi} \kappa^{su} \left\{ \kappa_{\phi s}^{(u)} - \frac{1}{2} \kappa_{\phi su} \right\} = \underline{\ell}_a^\top K^{\beta\beta} \text{tr}(\Delta W_3 X K^{\beta\beta} X^\top). \quad (\text{A.7.14})$$

A.7. Apêndice 7

Tomemos

$$\begin{aligned}
& \sum_{r,\phi,u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} = \sum_{r,u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} \\
& = \sum_{r,u} \kappa^{ar} \kappa^{\phi u} \sum_{t=1}^n \delta_t w_{3t} x_{tr} x_{tu} = \sum_{t=1}^n \delta_t w_{3t} \sum_r \kappa^{ar} x_{tr} \sum_u \kappa^{\phi u} x_{tu} \\
& = \sum_{t=1}^n \delta_t w_{3t} (\underline{\rho}_a^\top K^{\beta\beta} x_t) (x_t^\top K^{\beta\phi}) = \underline{\rho}_a^\top K^{\beta\beta} \sum_{t=1}^n (x_t \delta_t w_{3t} x_t^\top) K^{\beta\phi}.
\end{aligned}$$

Porém,

$$\sum_{r,\phi,u} \kappa^{ar} \kappa^{\phi u} \left\{ \kappa_{r\phi}^{(u)} - \frac{1}{2} \kappa_{r\phi u} \right\} = \underline{\rho}_a^\top K^{\beta\beta} (X^\top \Delta W_3 X) K^{\beta\phi}. \quad (\text{A.7.15})$$

De forma similar,

$$\begin{aligned}
& \sum_{r,s,\phi} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} = \sum_{r,s} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} \\
& = \sum_{r,s} \kappa^{ar} \kappa^{s\phi} \sum_{t=1}^n \delta_t w_{2t} x_{tr} x_{ts} = \sum_{t=1}^n \delta_t w_{2t} \sum_r \kappa^{ar} x_{tr} \sum_s \kappa^{s\phi} x_{ts} \\
& = \sum_{t=1}^n \delta_t w_{2t} (\underline{\rho}_a^\top K^{\beta\beta} x_t) (x_t^\top K^{\beta\phi}) = \underline{\rho}_a^\top K^{\beta\beta} (X^\top \Delta W_2 X) K^{\beta\phi}.
\end{aligned}$$

Portanto,

$$\sum_{r,s,\phi} \kappa^{ar} \kappa^{s\phi} \left\{ \kappa_{rs}^{(\phi)} - \frac{1}{2} \kappa_{rs\phi} \right\} = \underline{\rho}_a^\top K^{\beta\beta} (X^\top \Delta W_2 X) K^{\beta\phi}. \quad (\text{A.7.16})$$

Agora, calculamos

$$\begin{aligned}
& \sum_{\phi,\phi,u} \kappa^{a\phi} \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} = \kappa^{a\phi} \sum_u \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} \\
& = \kappa^{a\phi} \sum_u \kappa^{\phi u} \sum_{t=1}^n \delta_t w_{5t} x_{tu} = \kappa^{a\phi} \sum_{t=1}^n \delta_t w_{5t} \sum_u \kappa^{\phi u} x_{tu} \\
& = \kappa^{a\phi} \sum_{t=1}^n \delta_t w_{5t} (\rho_t^\top X K^{\beta\phi}) = \kappa^{a\phi} \left(\sum_{t=1}^n \delta_t w_{5t} \rho_t^\top \right) X K^{\beta\phi} \\
& = \underline{\rho}_a^\top K^{\beta\phi} \text{diagonal}(\Delta W_5) X K^{\beta\phi},
\end{aligned}$$

onde $\text{diagonal}(\cdot)$ é o vetor linha formado pela diagonal principal de uma matriz quadrada e ρ_t é o t -ésimo vetor coluna da matriz identidade de dimensão $n \times n$. Conseqüentemente,

$$\sum_{\phi,\phi,u} \kappa^{a\phi} \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2} \kappa_{\phi\phi u} \right\} = \underline{\rho}_a^\top K^{\beta\phi} \text{diagonal}(\Delta W_5) X K^{\beta\phi}. \quad (\text{A.7.17})$$

A.7. Apêndice 7

Analogamente,

$$\begin{aligned}
& \sum_{\phi, s, \phi} \kappa^{a\phi} \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} = \kappa^{a\phi} \sum_s \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} \\
& = \kappa^{a\phi} \sum_s \kappa^{s\phi} \sum_{t=1}^n \delta_t w_{4t} x_{ts} = \kappa^{a\phi} \sum_{t=1}^n \delta_t w_{4t} \sum_s \kappa^{s\phi} x_{ts} \\
& = \kappa^{a\phi} \sum_{t=1}^n \delta_t w_{4t} (\rho_t^\top X K^{\beta\phi}) = \kappa^{a\phi} \left(\sum_{t=1}^n \delta_t w_{4t} \rho_t^\top \right) X K^{\beta\phi} \\
& = \underline{\mathcal{L}}_a^\top K^{\beta\phi} \text{diagonal}(\Delta W_4) X K^{\beta\phi}.
\end{aligned}$$

Logo,

$$\sum_{\phi, s, \phi} \kappa^{a\phi} \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2} \kappa_{\phi s\phi} \right\} = \underline{\mathcal{L}}_a^\top K^{\beta\phi} \text{diagonal}(W_4) X K^{\beta\phi}. \quad (\text{A.7.18})$$

Da mesma forma,

$$\begin{aligned}
& \sum_{r, \phi, \phi} \kappa^{ar} \kappa^{\phi\phi} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} = \kappa^{\phi\phi} \sum_r \kappa^{ar} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} \\
& = \kappa^{\phi\phi} \sum_r \kappa^{ar} \sum_{t=1}^n \delta_t w_{4t} x_{tr} = \kappa^{\phi\phi} \sum_{t=1}^n \delta_t w_{4t} \sum_r \kappa^{ar} x_{tr} \\
& = \kappa^{\phi\phi} \sum_{t=1}^n \delta_t w_{4t} (\rho_t^\top X K^{\beta\beta}) \underline{\mathcal{L}}_a = \kappa^{\phi\phi} \left(\sum_{t=1}^n \delta_t w_{4t} \rho_t^\top \right) X K^{\beta\beta} \underline{\mathcal{L}}_a \\
& = \underline{\mathcal{L}}_a^\top K^{\beta\beta} X^\top \text{diagonal}(\Delta W_4)^\top K^{\phi\phi}.
\end{aligned}$$

Portanto,

$$\sum_{r, \phi, \phi} \kappa^{ar} \kappa^{\phi\phi} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2} \kappa_{r\phi\phi} \right\} = \underline{\mathcal{L}}_a^\top K^{\beta\beta} X^\top \text{diagonal}(W_4)^\top K^{\phi\phi}. \quad (\text{A.7.19})$$

Finalmente, temos

$$\begin{aligned}
& \sum_{\phi, \phi, \phi} \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\} = \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\} \\
& = \kappa^{a\phi} \kappa^{\phi\phi} \sum_{t=1}^n -\frac{1}{2} \delta_t s_t = \kappa^{a\phi} \kappa^{\phi\phi} \text{tr}(\Delta S) \\
& = \underline{\mathcal{L}}_a^\top K^{\beta\phi} K^{\phi\phi} \text{tr}(\Delta S),
\end{aligned}$$

onde $S = \text{diag}\{-s_1/2, \dots, -s_n/2\}$ é uma matriz diagonal de dimensão $n \times n$. Daí,

$$\sum_{\phi, \phi, \phi} \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2} \kappa_{\phi\phi\phi} \right\} = \underline{\mathcal{L}}_a^\top K^{\beta\phi} K^{\phi\phi} \text{tr}(\Delta S). \quad (\text{A.7.20})$$

A.8. Apêndice 8

Usando notação matricial derivamos o viés de segunda ordem como

$$\begin{aligned}
B(\widehat{\beta}_a) &= \underline{\varrho}_a^\top K^{\beta\beta} X^\top \Delta W_1 \delta_{\beta\beta} + \underline{\varrho}_a^\top K^{\beta\phi} \text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) \\
&\quad + \underline{\varrho}_a^\top K^{\beta\beta} (X^\top \Delta W_3 X) K^{\beta\phi} + \underline{\varrho}_a^\top K^{\beta\beta} (X^\top \Delta W_2 X) K^{\beta\phi} \\
&\quad + \underline{\varrho}_a^\top K^{\beta\phi} \text{diagonal}(\Delta W_5) X K^{\beta\phi} + \underline{\varrho}_a^\top K^{\beta\phi} \text{diagonal}(\Delta W_4) X K^{\beta\phi} \\
&\quad + \underline{\varrho}_a^\top K^{\beta\beta} X^\top \text{diagonal}(\Delta W_4)^\top K^{\phi\phi} + \underline{\varrho}_a^\top K^{\beta\phi} K^{\phi\phi} \text{tr}(\Delta S).
\end{aligned} \tag{A.7.21}$$

Fatorando adequadamente, $B(\widehat{\beta}_a)$ se simplifica na forma

$$\begin{aligned}
B(\widehat{\beta}_a) &= \underline{\varrho}_a^\top K^{\beta\beta} X^\top (\Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2) X K^{\beta\phi} + \text{diagonal}(\Delta W_4)^\top) \\
&\quad + \underline{\varrho}_a^\top K^{\beta\phi} (\text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) + K^{\phi\phi} \text{tr}(\Delta S) + \text{diagonal}(\Delta\{W_5 + W_4\}) X K^{\beta\phi}).
\end{aligned}$$

De forma análoga aos cálculos anteriores, deduz-se que

$$\begin{aligned}
B(\widehat{\phi}) &= K^{\phi\beta} X^\top \Delta W_1 \delta_{\beta\beta} + K^{\phi\phi} \text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) \\
&\quad + K^{\phi\beta} (X^\top \Delta W_3 X) K^{\beta\phi} + K^{\phi\beta} (X^\top \Delta W_2 X) K^{\beta\phi} \\
&\quad + K^{\phi\phi} \text{diagonal}(\Delta W_5) X K^{\beta\phi} + K^{\phi\phi} \text{diagonal}(\Delta W_4) X K^{\beta\phi} \\
&\quad + K^{\phi\beta} X^\top \text{diagonal}(\Delta W_4)^\top K^{\phi\phi} + K^{\phi\phi} K^{\phi\phi} \text{tr}(\Delta S),
\end{aligned} \tag{A.7.22}$$

e $B(\widehat{\phi})$ se simplifica na forma

$$\begin{aligned}
B(\widehat{\phi}) &= K^{\phi\beta} (X^\top \Delta W_1 \delta_{\beta\beta} + (X^\top \Delta W_3 X) K^{\beta\phi} + (X^\top \Delta W_2 X) K^{\beta\phi} + X^\top \text{diagonal}(\Delta W_4)^\top K^{\phi\phi}) \\
&\quad + K^{\phi\phi} (\text{tr}(\Delta W_3 X K^{\beta\beta} X^\top) + \text{diagonal}(\Delta W_5) X K^{\beta\phi} + \text{diagonal}(\Delta W_4) X K^{\beta\phi} + K^{\phi\phi} \text{tr}(\Delta S)).
\end{aligned}$$

A.8 Apêndice 8

Viés do estimador de máxima verossimilhança de $\Upsilon = (\rho^\top, \gamma^\top)^\top$.

Para o cálculo do viés de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta inflacionado em zero e um consideramos a seguinte estrutura

$$\begin{aligned}
g(\mu_t) &= \eta_t, \\
\log(\delta_{0t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{0t}, \\
\log(\delta_{1t}/(1 - \delta_{0t} - \delta_{1t})) &= \zeta_{1t},
\end{aligned}$$

$t = 1, \dots, n$, com parâmetro de precisão ϕ constante para todas as observações.

A.8. Apêndice 8

Note que

$$\begin{aligned}\delta_{0t} &= \frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}, \\ \delta_{1t} &= \frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}, \\ 1 - \delta_{0t} - \delta_{1t} &= \frac{1}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}}.\end{aligned}$$

Daí,

$$\begin{aligned}\frac{\partial \delta_{0t}}{\partial \zeta_{0t}} &= \left(\frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) \left(\frac{1 + e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) = \delta_{0t}(1 - \delta_{0t}), \\ \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} &= \left(\frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) \left(\frac{1 + e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) = \delta_{1t}(1 - \delta_{1t}), \\ \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} &= - \left(\frac{e^{\zeta_{0t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) \left(\frac{e^{\zeta_{1t}}}{1 + e^{\zeta_{0t}} + e^{\zeta_{1t}}} \right) = -\delta_{0t}\delta_{1t}.\end{aligned}\tag{A.8.1}$$

Desta forma, as expressões (A.3.3) e (A.3.4) se reduzem a

$$\begin{aligned}U_{r'} &= \sum_{t=1}^n \left\{ \frac{\mathbb{1}_{\{0\}}(y_t)}{\delta_{0t}} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \delta_{0t}(1 - \delta_{0t})v_{tr'} - \sum_{t=1}^n \left\{ \frac{\mathbb{1}_{\{1\}}(y_t)}{\delta_{1t}} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \delta_{0t}\delta_{1t}v_{tr'} \\ &= \sum_{t=1}^n (\mathbb{1}_{\{0\}}(y_t) - \delta_{0t})v_{tr'}\end{aligned}\tag{A.8.2}$$

e

$$\begin{aligned}U_{r''} &= \sum_{t=1}^n - \left\{ \frac{\mathbb{1}_{\{0\}}(y_t)}{\delta_{0t}} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \delta_{0t}\delta_{1t}z_{tr''} + \sum_{t=1}^n \left\{ \frac{\mathbb{1}_{\{1\}}(y_t)}{\delta_{1t}} + \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t)} \right\} \delta_{1t}(1 - \delta_{1t})z_{tr''} \\ &= \sum_{t=1}^n (\mathbb{1}_{\{1\}}(y_t) - \delta_{1t})z_{tr''},\end{aligned}\tag{A.8.3}$$

respectivamente. Seguidamente, as expressões (A.3.5), (A.3.6) e (A.3.7) se reduzem a

$$\begin{aligned}U_{r's'} &= \sum_{t=1}^n \left\{ \frac{\partial[(\mathbb{1}_{\{0\}}(y_t) - \delta_{0t})v_{tr'}]}{\partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{0t}} v_{ts'} \right\} + \left\{ \frac{\partial[(\mathbb{1}_{\{0\}}(y_t) - \delta_{0t})v_{tr'}]}{\partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{0t}} v_{ts'} \right\} \\ &= - \sum_{t=1}^n \delta_{0t}(1 - \delta_{0t})v_{tr'}v_{ts'},\end{aligned}\tag{A.8.4}$$

$$\begin{aligned}U_{r''s''} &= \sum_{t=1}^n \left\{ \frac{\partial[(\mathbb{1}_{\{1\}}(y_t) - \delta_{1t})z_{tr'']]}{\partial \delta_{0t}} \frac{\partial \delta_{0t}}{\partial \zeta_{1t}} z_{ts''} \right\} + \left\{ \frac{\partial[(\mathbb{1}_{\{1\}}(y_t) - \delta_{1t})z_{tr'']]}{\partial \delta_{1t}} \frac{\partial \delta_{1t}}{\partial \zeta_{1t}} z_{ts''} \right\} \\ &= - \sum_{t=1}^n \delta_{1t}(1 - \delta_{1t})z_{tr''}z_{ts''},\end{aligned}\tag{A.8.5}$$

A.8. Apêndice 8

$$\begin{aligned}
 U_{r's''} &= \sum_{t=1}^n \left\{ \frac{\partial[(\mathbb{1}_{\{0\}}(y_t) - \delta_{0t})v_{tr'}]}{\partial\delta_{0t}} \frac{\partial\delta_{0t}}{\partial\zeta_{1t}} z_{ts''} \right\} + \left\{ \frac{\partial[(\mathbb{1}_{\{0\}}(y_t) - \delta_{0t})v_{tr'}]}{\partial\delta_{1t}} \frac{\partial\delta_{1t}}{\partial\zeta_{1t}} z_{ts''} \right\} \\
 &= \sum_{t=1}^n \delta_{0t}\delta_{1t}v_{tr'}z_{ts''}.
 \end{aligned} \tag{A.8.6}$$

Agora, os momentos de terceira ordem com respeito a ρ e γ são obtidos pela diferenciação da função de log-verossimilhança $\ell_1(\rho, \gamma)$ com respeito aos parâmetros de interesse. Assim, para $r', s', u' = 1, \dots, k_0$, e $r'', s'', u'' = 1, \dots, k_1$, temos

$$\begin{aligned}
 U_{r's'u'} &= \sum_{t=1}^n \left\{ \frac{\partial[\delta_{0t}(1 - \delta_{0t})v_{tr'}v_{ts'}]}{\partial\delta_{0t}} \frac{\partial\delta_{0t}}{\partial\zeta_{0t}} v_{tu'} \right\} + \left\{ \frac{\partial[\delta_{0t}(1 - \delta_{0t})v_{tr'}v_{ts'}]}{\partial\delta_{1t}} \frac{\partial\delta_{1t}}{\partial\zeta_{0t}} v_{tu'} \right\} \\
 &= - \sum_{t=1}^n \delta_{0t}(1 - \delta_{0t})(1 - 2\delta_{0t})v_{tr'}v_{ts'}v_{tu'},
 \end{aligned} \tag{A.8.7}$$

$$\begin{aligned}
 U_{r's''u'} &= \sum_{t=1}^n \left\{ \frac{\partial[\delta_{0t}\delta_{1t}v_{tr'}z_{ts''}]}{\partial\delta_{0t}} \frac{\partial\delta_{0t}}{\partial\zeta_{0t}} v_{tu'} \right\} + \left\{ \frac{\partial[\delta_{0t}\delta_{1t}v_{tr'}z_{ts''}]}{\partial\delta_{1t}} \frac{\partial\delta_{1t}}{\partial\zeta_{0t}} v_{tu'} \right\} \\
 &= \sum_{t=1}^n \delta_{0t}\delta_{1t}(1 - 2\delta_{0t})v_{tr'}z_{ts''}v_{tu'},
 \end{aligned} \tag{A.8.8}$$

$$\begin{aligned}
 U_{r''s''u'} &= \sum_{t=1}^n \left\{ \frac{\partial[\delta_{1t}(1 - \delta_{1t})z_{tr''}z_{ts''}]}{\partial\delta_{0t}} \frac{\partial\delta_{0t}}{\partial\zeta_{0t}} v_{tu'} \right\} + \left\{ \frac{\partial[\delta_{1t}(1 - \delta_{1t})z_{tr''}z_{ts''}]}{\partial\delta_{1t}} \frac{\partial\delta_{1t}}{\partial\zeta_{0t}} v_{tu'} \right\} \\
 &= \sum_{t=1}^n \delta_{0t}\delta_{1t}(1 - 2\delta_{0t})z_{tr''}z_{ts''}v_{tu'},
 \end{aligned} \tag{A.8.9}$$

$$\begin{aligned}
 U_{r's'u''} &= \sum_{t=1}^n \left\{ \frac{\partial[\delta_{0t}(1 - \delta_{0t})v_{tr'}v_{ts'}]}{\partial\delta_{0t}} \frac{\partial\delta_{0t}}{\partial\zeta_{1t}} z_{tu''} \right\} + \left\{ \frac{\partial[\delta_{0t}(1 - \delta_{0t})v_{tr'}v_{ts'}]}{\partial\delta_{1t}} \frac{\partial\delta_{1t}}{\partial\zeta_{1t}} z_{tu''} \right\} \\
 &= \sum_{t=1}^n \delta_{0t}\delta_{1t}(1 - 2\delta_{1t})v_{tr'}v_{ts'}z_{tu''},
 \end{aligned} \tag{A.8.10}$$

$$\begin{aligned}
 U_{r's''u''} &= \sum_{t=1}^n \left\{ \frac{\partial[\delta_{0t}\delta_{1t}v_{tr'}z_{ts''}]}{\partial\delta_{0t}} \frac{\partial\delta_{0t}}{\partial\zeta_{1t}} z_{tu''} \right\} + \left\{ \frac{\partial[\delta_{0t}\delta_{1t}v_{tr'}z_{ts''}]}{\partial\delta_{1t}} \frac{\partial\delta_{1t}}{\partial\zeta_{1t}} z_{tu''} \right\} \\
 &= \sum_{t=1}^n \delta_{0t}\delta_{1t}(1 - 2\delta_{1t})v_{tr'}v_{ts'}z_{tu''},
 \end{aligned} \tag{A.8.11}$$

A.8. Apêndice 8

$$\begin{aligned}
 U_{r''s''u''} &= \sum_{t=1}^n \left\{ \frac{\partial[\delta_{1t}(1-\delta_{1t})z_{tr''}z_{ts''}]}{\partial\delta_{0t}} \frac{\partial\delta_{0t}}{\partial\zeta_{1t}} z_{tu''} \right\} + \left\{ \frac{\partial[\delta_{1t}(1-\delta_{1t})z_{tr''}z_{ts''}]}{\partial\delta_{1t}} \frac{\partial\delta_{1t}}{\partial\zeta_{1t}} z_{tu''} \right\} \\
 &= - \sum_{t=1}^n \delta_{1t}(1-\delta_{1t})(1-2\delta_{1t})z_{tr''}z_{ts''}z_{tu''}.
 \end{aligned} \tag{A.8.12}$$

Tomando o valor esperado dos momentos acima, temos

$$\kappa_{r's'} = E(U_{r's'}) = U_{r's'},$$

$$\kappa_{r''s''} = E(U_{r''s''}) = U_{r''s''},$$

$$\kappa_{r's''} = E(U_{r's''}) = U_{r's''},$$

$$\kappa_{r's'u'} = E(U_{r's'u'}) = U_{r's'u'},$$

$$\kappa_{r''s''u''} = E(U_{r''s''u''}) = U_{r''s''u''},$$

$$\kappa_{r's'u''} = E(U_{r's'u''}) = U_{r's'u''},$$

$$\kappa_{r's''u''} = E(U_{r's''u''}) = U_{r's''u''}.$$

Derivando os cumulantes de segunda ordem com respeito aos parâmetros de interesse, temos que

$$\kappa_{r's'}^{(u')} = \frac{\partial\kappa_{r's'}}{\partial\rho_{u'}} = U_{r's'u'},$$

$$\kappa_{r's'}^{(u'')} = \frac{\partial\kappa_{r's'}}{\partial\gamma_{u''}} = U_{r's'u''},$$

$$\kappa_{r's''}^{(u')} = \frac{\partial\kappa_{r's''}}{\partial\rho_{u'}} = U_{r's''u'},$$

$$\kappa_{r's''}^{(u'')} = \frac{\partial\kappa_{r's''}}{\partial\gamma_{u''}} = U_{r's''u''},$$

$$\kappa_{r''s''}^{(u')} = \frac{\partial \kappa_{r's''}}{\partial \rho_{u'}} = U_{r''s''u'},$$

$$\kappa_{r''s''}^{(u'')} = \frac{\partial \kappa_{r's''}}{\partial \gamma_{u''}} = U_{r''s''u''}.$$

Usando os resultados acima obtemos as seguintes quantidades:

$$\kappa_{r's'}^{(u')} - \frac{1}{2}\kappa_{r's'u'} = -\frac{1}{2}\sum_{t=1}^n \delta_{0t}(1-\delta_{0t})(1-2\delta_{0t})v_{tr'}v_{ts'}v_{tu'}, \quad (\text{A.8.13})$$

$$\kappa_{r''s'}^{(u')} - \frac{1}{2}\kappa_{r''s'u'} = -\frac{1}{2}\sum_{t=1}^n \delta_{0t}\delta_{1t}(1-2\delta_{0t})v_{ts'}v_{tu'}z_{tr''}, \quad (\text{A.8.14})$$

$$\kappa_{r's''}^{(u')} - \frac{1}{2}\kappa_{r's''u'} = -\frac{1}{2}\sum_{t=1}^n \delta_{0t}\delta_{1t}(1-2\delta_{0t})v_{tr'}v_{tu'}z_{ts''}, \quad (\text{A.8.15})$$

$$\kappa_{r's'}^{(u'')} - \frac{1}{2}\kappa_{r's'u''} = -\frac{1}{2}\sum_{t=1}^n \delta_{0t}\delta_{1t}(1-2\delta_{1t})v_{tr'}v_{ts'}z_{tu''}, \quad (\text{A.8.16})$$

$$\kappa_{r''s''}^{(u')} - \frac{1}{2}\kappa_{r''s''u'} = -\frac{1}{2}\sum_{t=1}^n \delta_{0t}\delta_{1t}(1-2\delta_{0t})z_{ts''}v_{tu'}z_{tr''}, \quad (\text{A.8.17})$$

$$\kappa_{r''s'}^{(u'')} - \frac{1}{2}\kappa_{r''s'u''} = -\frac{1}{2}\sum_{t=1}^n \delta_{0t}\delta_{1t}(1-2\delta_{1t})z_{tr''}v_{ts'}z_{tu''}, \quad (\text{A.8.18})$$

$$\kappa_{r's''}^{(u'')} - \frac{1}{2}\kappa_{r's''u''} = -\frac{1}{2}\sum_{t=1}^n \delta_{0t}\delta_{1t}(1-2\delta_{1t})v_{tr'}z_{ts''}z_{tu''}, \quad (\text{A.8.19})$$

$$\kappa_{r''s''}^{(u'')} - \frac{1}{2}\kappa_{r''s''u''} = -\frac{1}{2}\sum_{t=1}^n \delta_{1t}(1-\delta_{1t})(1-2\delta_{1t})z_{tr''}z_{ts''}z_{tu''}. \quad (\text{A.8.20})$$

Se definimos as matrizes diagonais

$$\begin{aligned} P_1 &= \text{diag}\{-\delta_{01}(1-\delta_{01})(1-2\delta_{01}), \dots, -\delta_{0n}(1-\delta_{0n})(1-2\delta_{0n})\}, \\ P_2 &= \text{diag}\{\delta_{01}\delta_{11}(1-2\delta_{01}), \dots, \delta_{0n}\delta_{1n}(1-2\delta_{0n})\}, \\ P_3 &= \text{diag}\{-\delta_{01}\delta_{11}(1-2\delta_{11}), \dots, -\delta_{0n}\delta_{1n}(1-2\delta_{1n})\}, \\ P_4 &= \text{diag}\{-\delta_{11}(1-\delta_{11})(1-2\delta_{11}), \dots, -\delta_{1n}(1-\delta_{1n})(1-2\delta_{1n})\}, \end{aligned} \quad (\text{A.8.21})$$

A.8. Apêndice 8

de dimensão $(n \times n)$, temos para $r' = 1, \dots, k_0$

$$\sum_{r',s',u'} \kappa^{ar'} \kappa^{s'u'} \left\{ \kappa_{r's'}^{(u')} - \frac{1}{2} \kappa_{r's'u'} \right\} = \frac{1}{2} \sum_{r',s',u'} \kappa^{ar'} \kappa^{s'u'} \sum_{t=1}^n p_{1t} v_{tr'} v_{ts'} v_{tu'},$$

em que p_{1t} é o t -ésimo elemento da matriz diagonal P_1 . Note que

$$\begin{aligned} \sum_{r',s',u'} \kappa^{ar'} \kappa^{s'u'} \left\{ \kappa_{r's'}^{(u')} - \frac{1}{2} \kappa_{r's'u'} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{1t} \sum_{r'} \kappa^{ar'} v_{tr'} \sum_{s',u'} v_{ts'} \kappa^{s'u'} v_{tu'} \\ &= -\frac{1}{2} \sum_{t=1}^n p_{1t} \sum_{r'} \kappa^{ar'} v_{tr'} v_t K^{\rho\rho} v_t^\top \\ &= \frac{1}{2} \sum_{t=1}^n p_{1t} (v_t K^{\rho\rho} v_t^\top) \mathcal{E}_a^\top K^{\rho\rho} v_t^\top, \end{aligned}$$

sendo v_t o vetor linha obtido da t -ésima linha da matriz V , $K^{\rho\rho}$ a matriz de dimensão $k_0 \times k_0$ obtida da inversa da matriz de informação de Fisher (3.3.14) e \mathcal{E}_a representando o a -ésimo vetor coluna da matriz identidade de dimensão $k_0 \times k_0$. Se definimos $\delta_{\rho\rho}$ como o vetor de dimensão $n \times 1$ obtido da diagonal principal de matriz $V K^{\rho\rho} V^\top$, temos que

$$\sum_{r',s',u'} \kappa^{ar'} \kappa^{s'u'} \left\{ \kappa_{r's'}^{(u')} - \frac{1}{2} \kappa_{r's'u'} \right\} = \frac{1}{2} \mathcal{E}_a^\top K^{\rho\rho} V^\top P_1 \delta_{\rho\rho}. \quad (\text{A.8.22})$$

Agora,

$$\begin{aligned} \sum_{r'',s',u'} \kappa^{ar''} \kappa^{s'u'} \left\{ \kappa_{r''s'}^{(u')} - \frac{1}{2} \kappa_{r''s'u'} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{2t} \sum_{r''} \kappa^{ar''} z_{tr''} \sum_{s',u'} v_{ts'} \kappa^{s'u'} v_{tu'} \\ &= -\frac{1}{2} \sum_{t=1}^n p_{2t} \sum_{r''} \kappa^{ar''} z_{tr''} v_t K^{\rho\rho} v_t^\top \\ &= \frac{1}{2} \sum_{t=1}^n p_{2t} (v_t K^{\rho\rho} v_t^\top) \mathcal{E}_a^\top K^{\rho\rho} z_t^\top \\ &= \frac{1}{2} \mathcal{E}_a^\top K^{\rho\rho} Z^\top P_2 \delta_{\rho\rho}, \end{aligned} \quad (\text{A.8.23})$$

sendo z_t o vetor linha obtido da t -ésima linha da matriz Z e $K^{\rho\rho}$ a matriz de dimensão $k_0 \times (k_1 - k_0)$ obtida da inversa da matriz de informação de Fisher (3.3.14).

A.8. Apêndice 8

Analogamente,

$$\begin{aligned}
\sum_{r',s'',u'} \kappa^{ar'} \kappa^{s''u'} \left\{ \kappa_{r's''}^{(u')} - \frac{1}{2} \kappa_{r's''u'} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{2t} \sum_{r'} \kappa^{ar'} v_{tr'} \sum_{s'',u'} z_{ts''} \kappa^{s''u'} v_{tu'} \\
&= -\frac{1}{2} \sum_{t=1}^n p_{2t} \sum_{r'} \kappa^{ar'} v_{tr'} v_t K^{\rho\gamma} z_t^\top \\
&= \frac{1}{2} \sum_{t=1}^n p_{2t} (v_t K^{\rho\gamma} z_t^\top) \mathcal{L}_a^\top K^{\rho\rho} v_t^\top \\
&= \frac{1}{2} \mathcal{L}_a^\top K^{\rho\rho} V^\top P_2 \delta_{\rho\gamma},
\end{aligned} \tag{A.8.24}$$

em que $\delta_{\rho\gamma}$ é o vetor de dimensão $n \times 1$ obtido da diagonal principal de matriz $V K^{\rho\rho} Z^\top$.

Temos agora que

$$\begin{aligned}
\sum_{r',s',u''} \kappa^{ar'} \kappa^{s'u''} \left\{ \kappa_{r's'}^{(u'')} - \frac{1}{2} \kappa_{r's'u''} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{3t} \sum_{r'} \kappa^{ar'} v_{tr'} \sum_{s',u''} v_{ts'} \kappa^{s'u''} z_{tu''} \\
&= -\frac{1}{2} \sum_{t=1}^n p_{3t} \sum_{r'} \kappa^{ar'} v_{tr'} v_t K^{\rho\gamma} z_t^\top \\
&= \frac{1}{2} \sum_{t=1}^n p_{3t} (v_t K^{\rho\gamma} z_t^\top) \mathcal{L}_a^\top K^{\rho\rho} v_t^\top \\
&= \frac{1}{2} \mathcal{L}_a^\top K^{\rho\rho} V^\top P_3 \delta_{\rho\gamma}.
\end{aligned} \tag{A.8.25}$$

De forma análoga,

$$\begin{aligned}
\sum_{r'',s'',u'} \kappa^{ar''} \kappa^{s''u'} \left\{ \kappa_{r''s''}^{(u')} - \frac{1}{2} \kappa_{r''s''u'} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{2t} \sum_{r''} \kappa^{ar''} z_{tr''} \sum_{s'',u'} z_{ts''} \kappa^{s''u'} v_{tu'} \\
&= -\frac{1}{2} \sum_{t=1}^n p_{2t} \sum_{r''} \kappa^{ar''} z_{tr''} v_t K^{\rho\gamma} z_t^\top \\
&= \frac{1}{2} \sum_{t=1}^n p_{2t} (v_t K^{\rho\gamma} z_t^\top) \mathcal{L}_a^\top K^{\rho\gamma} v_t^\top \\
&= \frac{1}{2} \mathcal{L}_a^\top K^{\rho\gamma} V^\top P_2 \delta_{\rho\gamma}.
\end{aligned} \tag{A.8.26}$$

A.9. Apêndice 9

Similarmente,

$$\begin{aligned}
\sum_{r'',s',u''} \kappa^{ar''} \kappa^{s'u''} \left\{ \kappa_{r''s'}^{(u'')} - \frac{1}{2} \kappa_{r''s'u''} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{3t} \sum_{r''} \kappa^{ar''} z_{tr''} \sum_{s',u''} v_{ts'} \kappa^{s'u''} z_{tu''} \\
&= -\frac{1}{2} \sum_{t=1}^n p_{3t} \sum_{r''} \kappa^{ar''} z_{tr''} v_t K^{\rho\gamma} z_t^\top \\
&= \frac{1}{2} \sum_{t=1}^n p_{3t} (v_t K^{\rho\gamma} z_t^\top) \mathcal{L}_a^\top K^{\rho\gamma} z_t^\top \\
&= \frac{1}{2} \mathcal{L}_a^\top K^{\rho\gamma} Z^\top P_3 \delta_{\rho\gamma}.
\end{aligned} \tag{A.8.27}$$

Agora, temos que

$$\begin{aligned}
\sum_{r',s'',u''} \kappa^{ar'} \kappa^{s''u''} \left\{ \kappa_{r's''}^{(u'')} - \frac{1}{2} \kappa_{r's''u''} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{3t} \sum_{r'} \kappa^{ar'} v_{tr'} \sum_{s'',u''} z_{ts''} \kappa^{s''u''} z_{tu''} \\
&= -\frac{1}{2} \sum_{t=1}^n p_{3t} \sum_{r'} \kappa^{ar'} v_{tr'} z_t K^{\gamma\gamma} z_t^\top \\
&= \frac{1}{2} \sum_{t=1}^n p_{3t} (z_t K^{\gamma\gamma} z_t^\top) \mathcal{L}_a^\top K^{\rho\rho} v_t^\top \\
&= \frac{1}{2} \mathcal{L}_a^\top K^{\rho\rho} V^\top P_3 \delta_{\gamma\gamma},
\end{aligned} \tag{A.8.28}$$

sendo $K^{\gamma\gamma}$ a matriz de dimensão $k_1 \times k_1$ obtida da inversa da matriz de informação de Fisher (3.3.14) e $\delta_{\gamma\gamma}$ o vetor de dimensão $n \times 1$ obtido da diagonal principal de matriz $ZK^{\gamma\gamma}Z^\top$. Finalmente

$$\begin{aligned}
\sum_{r'',s'',u''} \kappa^{ar''} \kappa^{s''u''} \left\{ \kappa_{r''s''}^{(u'')} - \frac{1}{2} \kappa_{r''s''u''} \right\} &= \frac{1}{2} \sum_{t=1}^n p_{4t} \sum_{r''} \kappa^{ar''} z_{tr''} \sum_{s'',u''} z_{ts''} \kappa^{s''u''} z_{tu''} \\
&= -\frac{1}{2} \sum_{t=1}^n p_{4t} \sum_{r''} \kappa^{ar''} z_{tr''} z_t K^{\gamma\gamma} z_t^\top \\
&= \frac{1}{2} \sum_{t=1}^n p_{4t} (z_t K^{\gamma\gamma} z_t^\top) \mathcal{L}_a^\top K^{\rho\gamma} z_t^\top \\
&= \frac{1}{2} \mathcal{L}_a^\top K^{\rho\gamma} Z^\top P_4 \delta_{\gamma\gamma}.
\end{aligned} \tag{A.8.29}$$

A.9 Apêndice 9

Resíduos padronizados do modelo de regressão beta inflacionado em zero ou um

Respeitando o formato da distribuição de probabilidades da variável resposta e usando a definição geral de resíduos

A.9. Apêndice 9

dada por Cox & Snell (1968), podemos definir resíduos padronizados para o modelo de regressão beta inflacionado em zero ou um. Neste sentido, se consideramos ϕ conhecido, temos que o algoritmo escore de Fisher é expresso por

$$\begin{aligned}\gamma^{(m+1)} &= \gamma^{(m)} + (Z^\top Q^{(m)} Z)^{-1} Z^\top P^{(m)} G^{(m)} (y^c - \alpha^{*(m)}) \\ &= (Z^\top Q^{(m)} Z)^{-1} Z^\top Q^{(m)} z^{(m)},\end{aligned}\tag{A.9.1}$$

onde $z^{(m)} = Z\gamma^{(m)} + (Q^{(m)})^{-1}P^{(m)}G^{(m)}(y^c - \alpha^{*(m)})$. Já o algoritmo escore de Fisher para calcular a estimativa de máxima verossimilhança de β é expresso por

$$\beta^{(m+1)} = (X^\top W_{\beta\beta}^{(m)} X)^{-1} X^\top W_{\beta\beta}^{(m)} z_1^{*(m)}\tag{A.9.2}$$

sendo

$$z_1^{*(m)} = X\beta^{(m)} + (W_{\beta\beta}^{(m)})^{-1}T^{(m)}H^{(m)}(y^* - \mu^{*(m)})$$

um vetor de dimensão $n \times 1$. Da convergência do processo iterativo escore de Fisher para γ e β temos

$$\begin{aligned}\hat{\gamma} &= (Z^\top \hat{Q} Z)^{-1} Z^\top \hat{Q} \hat{\tau}_1, \\ \hat{\beta} &= (X^\top \hat{W}_{\beta\beta} X)^{-1} X^\top \hat{W}_{\beta\beta} \hat{\tau}_2,\end{aligned}$$

sendo $\hat{\tau}_1 = \hat{\zeta} + G^{-1}(y^c - \hat{\alpha}^*)$ e $\hat{\tau}_2 = \hat{\eta} + \phi \hat{W}_{\beta\beta}^{-1} \hat{T} H(y^* - \hat{\mu}^*)$ em que $\hat{\zeta} = (\hat{\zeta}_1, \dots, \hat{\zeta}_n) = Z\hat{\gamma}$ e $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_m) = X\hat{\beta}$. Se definimos $\tau_1^* = \hat{Q}^{1/2} \hat{\tau}_1$, $\tau_2^* = \hat{W}_{\beta\beta}^{1/2} \hat{\tau}_2$, $Z^* = \hat{Q}^{1/2} Z$ e $X^* = \hat{W}_{\beta\beta}^{1/2} X$, temos que

$$\begin{aligned}\hat{\gamma} &= (Z^{*\top} Z^*)^{-1} Z^{*\top} \tau_1^*, \\ \hat{\beta} &= (X^{*\top} X^*)^{-1} X^{*\top} \tau_2^*,\end{aligned}\tag{A.9.3}$$

ou seja, $\hat{\gamma}$ é a solução de mínimos quadrados de uma regressão linear de τ_1^* contra as colunas de Z^* e $\hat{\beta}$ é a solução de mínimos quadrados de uma regressão linear de τ_2^* contra as colunas de X^* . Da equação (A.9.3) temos que

$$\begin{aligned}\hat{\tau}_1^* &= Z^* \hat{\gamma} = \hat{Q}^{1/2} Z (Z^\top \hat{Q} Z)^{-1} Z^\top \hat{Q}^{1/2} \hat{Q}^{1/2} \hat{\tau}_1 \\ &= \hat{H}_1^* \tau_1^*, \\ \hat{\tau}_2^* &= X^* \hat{\beta} = \hat{W}_{\beta\beta}^{1/2} X (X^\top \hat{W}_{\beta\beta} X)^{-1} X^\top \hat{W}_{\beta\beta}^{1/2} \hat{W}_{\beta\beta}^{1/2} \hat{\tau}_2 \\ &= \hat{H}_2^* \tau_2^*,\end{aligned}$$

em que $\hat{H}_1^* = \hat{Q}^{1/2} Z (Z^\top \hat{Q} Z)^{-1} Z^\top \hat{Q}^{1/2}$ e $\hat{H}_2^* = \hat{W}_{\beta\beta}^{1/2} X (X^\top \hat{W}_{\beta\beta} X)^{-1} X^\top \hat{W}_{\beta\beta}^{1/2}$ são matrizes de projeção. Assim, de (A.9.3) os resíduos ordinários podem ser expressos por

$$\begin{aligned}e_1^* &= \tau_1^* - \hat{\tau}_1^* = (I_n - \hat{H}_1^*) \tau_1^*, \\ e_2^* &= \tau_2^* - \hat{\tau}_2^* = (I_n - \hat{H}_2^*) \tau_2^*,\end{aligned}\tag{A.9.4}$$

A.10. Apêndice 10

em que I_n é a matriz identidade de dimensão $n \times n$. Alternativamente, podemos escrever

$$\begin{aligned} e_1^* &= (I_n - \widehat{H}_1^*)\tau_1^* = \widehat{Q}^{1/2}\widehat{\tau}_1 - \widehat{Q}^{1/2}Z(Z^\top\widehat{Q}Z)^{-1}Z^\top\widehat{Q}^{1/2}\widehat{Q}^{1/2}\widehat{\tau}_1, \\ &= \widehat{Q}^{1/2}G^{-1}(y^c - \widehat{\alpha}^*) \\ &= P^{1/2}(y^c - \widehat{\alpha}^*) \end{aligned} \quad (\text{A.9.5})$$

e

$$\begin{aligned} e_2^* &= (I_n - \widehat{H}_2^*)\tau_2^* = \widehat{W}_{\beta\beta}^{1/2}\widehat{\tau}_2 - \widehat{W}_{\beta\beta}^{1/2}X(X^\top\widehat{W}_{\beta\beta}X)^{-1}X^\top\widehat{W}_{\beta\beta}^{1/2}\widehat{W}_{\beta\beta}^{1/2}\widehat{\tau}_2, \\ &= \phi\widehat{W}_{\beta\beta}^{1/2}\widehat{W}_{\beta\beta}^{-1}\widehat{T}(y^* - \widehat{\mu}^*) \\ &= \Delta^{-1/2}W^{-1/2}(y^* - \widehat{\mu}^*). \end{aligned} \quad (\text{A.9.6})$$

As matrizes de variâncias-covariâncias assintóticas dos resíduos e_1^* e e_2^* , com as quantidades avaliadas nos parâmetros verdadeiros são

$$\begin{aligned} \text{Var}(e_1^*) &= (I_n - H_1^*)\text{Var}(\tau_1^*), \\ &= (I_n - H_1^*)\text{Var}(\widehat{Q}^{1/2}\tau_1)(I_n - H_1^*) \\ &= (I_n - H_1^*)\widehat{Q}^{1/2}\text{Var}(\tau_1)\widehat{Q}^{1/2}(I_n - H_1^*) \\ &= (I_n - H_1^*)\widehat{Q}^{1/2}\widehat{Q}^{-1}\widehat{Q}^{1/2}(I_n - H_1^*) \\ &= (I_n - H_1^*) \end{aligned} \quad (\text{A.9.7})$$

e

$$\begin{aligned} \text{Var}(e_2^*) &= (I_n - H_2^*)\text{Var}(\tau_2^*), \\ &= (I_n - H_2^*)\text{Var}(\widehat{W}_{\beta\beta}^{1/2}\tau_2)(I_n - H_2^*) \\ &= (I_n - H_2^*)\widehat{W}_{\beta\beta}^{1/2}\text{Var}(\tau_2)\widehat{W}_{\beta\beta}^{1/2}(I_n - H_2^*) \\ &= (I_n - H_2^*)\widehat{W}_{\beta\beta}^{1/2}\widehat{W}_{\beta\beta}^{-1}\widehat{W}_{\beta\beta}^{1/2}(I_n - H_2^*) \\ &= (I_n - H_2^*). \end{aligned} \quad (\text{A.9.8})$$

Adicionalmente, é fácil ver que $E(e_1^*) = 0$ e $E(e_2^*) = 0$.

A.10 Apêndice 10

Resíduos padronizados do modelo de regressão beta inflacionado em zero e um

De forma análoga ao modelo de regressão beta inflacionado em zero ou um consideramos que ϕ é conhecido. Da convergência do processo iterativo escore de Fisher de $\Upsilon = (\rho^\top, \gamma^\top)^\top$ temos que

$$\widehat{\Upsilon} = (\widetilde{Z}^\top\widehat{Q}\widetilde{Z})^{-1}Z^\top\widehat{Q}\widehat{\tau}_d,$$

sendo $\widehat{\tau}_d = \widetilde{Z}\widehat{\Upsilon} + \widehat{Q}^{-1}(y_d - \widehat{\Delta}_d)$, em que $y_d = (y_{\{0\}}^\top, y_{\{1\}}^\top)^\top$, $\Delta_d = (\delta_0^\top, \delta_1^\top)^\top$ são vetores de dimensão $2n \times 1$. Aqui

A.10. Apêndice 10

$\delta_0 = (\delta_{01}, \dots, \delta_{0n})^\top$, $\delta_1 = (\delta_{11}, \dots, \delta_{1n})^\top$, $y_{\{0\}} = (\mathbb{1}_{\{0\}}(y_1), \dots, \mathbb{1}_{\{0\}}(y_n))^\top$ e $y_{\{1\}} = (\mathbb{1}_{\{1\}}(y_1), \dots, \mathbb{1}_{\{1\}}(y_n))^\top$ são vetores de dimensão $n \times 1$. Adicionalmente,

$$\tilde{Z} = \begin{pmatrix} V & 0 \\ 0 & Z \end{pmatrix}, \quad Q = \begin{pmatrix} Q_1 & Q_3 \\ Q_3 & Q_2 \end{pmatrix},$$

são matrizes de dimensão $(2n \times (k_0 + k_1))$ e $(2n \times 2n)$, respectivamente. V é uma matriz $n \times k_0$ de covariadas cuja t -ésima linha é v_t e Z é uma matriz $n \times k_1$ de covariadas cuja t -ésima linha é z_t . Os blocos de Q são: $Q_1 = V^\top \text{diag}\{\delta_{01}(1 - \delta_{01}), \dots, \delta_{0n}(1 - \delta_{0n})\}V$, $Q_2 = Z^\top \text{diag}\{\delta_{11}(1 - \delta_{11}), \dots, \delta_{1n}(1 - \delta_{1n})\}Z$ e $Q_3 = Z^\top \text{diag}\{-\delta_{01}\delta_{11}, \dots, -\delta_{0n}\delta_{1n}\}V$.

Tomando $\tau_d^* = Q^{1/2}\hat{\tau}_d$ e $Z^* = Q^{1/2}\tilde{Z}$ temos que

$$\hat{Y} = (Z^{*\top}Z^*)^{-1}Z^{*\top}\tau_d^*. \quad (\text{A.10.1})$$

Portanto, \hat{Y} é a solução de mínimos quadrados da regressão auxiliar de τ_d^* contra as colunas de Z^* . Da equação anterior deduz-se que

$$\begin{aligned} \hat{\tau}_d^* &= Z^*\hat{Y} = \hat{Q}^{1/2}\tilde{Z}(\tilde{Z}^\top\hat{Q}\tilde{Z})^{-1}\tilde{Z}^\top\hat{Q}^{1/2}\hat{Q}^{1/2}\hat{\tau}_d \\ &= \hat{H}_d^*\tau_d^*. \end{aligned}$$

O resíduo ordinário da regressão (A.10.1) pode ser expresso por

$$e_d^* = \tau_d^* - \hat{\tau}_d^* = \tau_d^* - \hat{H}_d^*\tau_d^* = (I_{2n} - \hat{H}_d^*)\tau_d^*,$$

em que I_{2n} é a matriz identidade de dimensão $(2n \times 2n)$ e $\hat{H}_d^* = \hat{Q}^{1/2}\tilde{Z}(\tilde{Z}^\top\hat{Q}\tilde{Z})^{-1}\tilde{Z}^\top\hat{Q}^{1/2}$ é a matriz de projeção. Note que e_d^* também pode ser escrito como

$$\begin{aligned} e_d^* &= (I_{2n} - \hat{H}_d^*)\tau_d^* = \hat{Q}^{1/2}\hat{\tau}_d - \hat{Q}^{1/2}\tilde{Z}(\tilde{Z}^\top\hat{Q}\tilde{Z})^{-1}\tilde{Z}^\top\hat{Q}^{1/2}\hat{Q}^{1/2}\hat{\tau}_d, \\ &= \hat{Q}^{1/2}\hat{Q}^{-1}(y_d - \hat{\Delta}_d) \\ &= \hat{Q}^{-1/2}(y_d - \hat{\Delta}_d). \end{aligned} \quad (\text{A.10.2})$$

Note que se as quantidades são avaliadas nos parâmetros verdadeiros, obtemos que $E(e_d^*) = 0$ e

$$\text{Var}(e_d^*) = \hat{Q}^{-1/2}\text{Var}(y_d)\hat{Q}^{-1/2}.$$

Por outro lado,

$$\begin{aligned} \text{Var}(e_d^*) &= (I_{2n} - \hat{H}_d^*)\text{Var}(\tau_d^*)(I_{2n} - \hat{H}_d^*) \\ &= (I_{2n} - \hat{H}_d^*)\hat{Q}^{-1/2}\text{Var}(\hat{\tau}_d)\hat{Q}^{-1/2}(I_{2n} - \hat{H}_d^*). \end{aligned}$$

Assintoticamente, temos $\text{Var}(\hat{Y}) = (\tilde{Z}^\top\hat{Q}\tilde{Z})^{-1}$, que implica $\text{Var}(\hat{\tau}_d) = \hat{Q}^{-1}$. Desta forma,

$$\begin{aligned} \text{Var}(e_d^*) &= (I_{2n} - \hat{H}_d^*)\hat{Q}^{-1/2}\hat{Q}^{-1}\hat{Q}^{-1/2}(I_{2n} - \hat{H}_d^*) \\ &= (I_{2n} - \hat{H}_d^*) \end{aligned}$$

A.11. Apêndice 11

e

$$\text{Var}(y_d) = \widehat{Q}^{-1/2}(I_{2n} - \widehat{H}_d^*)\widehat{Q}^{-1/2}.$$

Desta forma, podemos construir o resíduo padronizado para o componente discreto do modelo de regressão beta inflacionado em zero e um na forma:

$$r_{p_i}^d = \frac{(y_{d_i} - \widehat{\Delta}_{d_i})}{\sqrt{\widehat{q}_{ii}(1 - \widehat{h}_{d_{ii}})}}, \quad i = 1, \dots, 2n, \quad (\text{A.10.3})$$

em que $y_d = (y_{\{0\}}^\top, y_{\{1\}}^\top)^\top$, $\Delta_d = (\delta_0^\top, \delta_1^\top)^\top$ são vetores de dimensão $2n \times 1$. Aqui, $\delta_0 = (\delta_{01}, \dots, \delta_{0n})^\top$, $\delta_1 = (\delta_{11}, \dots, \delta_{1n})^\top$, $y_{\{0\}} = (\mathbb{1}_{\{0\}}(y_1), \dots, \mathbb{1}_{\{0\}}(y_n))^\top$ e $y_{\{1\}} = (\mathbb{1}_{\{1\}}(y_1), \dots, \mathbb{1}_{\{1\}}(y_n))^\top$ são vetores de dimensão $n \times 1$. Adicionalmente, \widehat{q}_{ii} é o i -ésimo elemento da diagonal principal da matriz \widehat{Q} definida em (A.10) e $\widehat{h}_{d_{ii}}$ é o i -ésimo elemento da diagonal principal da matriz de projeção.

Note que, para $i = 1, \dots, n$, o resíduo $r_{p_i}^d$ corresponde ao resíduo padronizado do sub-modelo do componente discreto que modela a probabilidade de ocorrência de zeros e para $i = n + 1, \dots, 2n$, $r_{p_i}^d$ corresponde ao resíduo padronizado do sub-modelo do componente discreto que modela a probabilidade de ocorrência de uns. Desta forma, parece razoável sugerir dois conjuntos de gráficos de resíduos para o componente discreto deste modelo. O primeiro é de $r_{p_i}^d$ contra $\widehat{\Delta}_{d_i}$ para $i = 1, \dots, n$ e o segundo de $r_{p_i}^d$ contra $\widehat{\Delta}_{d_i}$ para $i = n + 1, \dots, 2n$. Neste caso, é de se esperar que os gráficos revelem marginalmente pontos aberrantes para cada sub-modelo.

Definimos as matrizes de dimensão $(2n \times n)$, $\Psi_0 = (I_n, 0_n)^\top$ e $\Psi_1 = (0_n, I_n)^\top$, em que I_n é a matriz identidade de dimensão $n \times n$ e 0_n é uma matriz de zeros de dimensão $n \times n$. Desta forma, podemos definir as matrizes de projeção

$$\begin{aligned} H^{\{0\}} &= \Psi_0^\top H_d^* \Psi_0, \\ H^{\{1\}} &= \Psi_1^\top H_d^* \Psi_1, \end{aligned} \quad (\text{A.10.4})$$

correspondentes aos sub-modelos do componente discreto que modelam a probabilidade de ocorrência de zeros e uns respectivamente.

A.11 Apêndice 11

Cálculo de curvaturas para o modelo de regressão beta inflacionado em zero ou um

Matriz Hessiana

Para obter a matriz hessiana utilizamos as quantidades (A.1.1)–(A.1.4) obtidas no Apêndice A.1. Definimos as matrizes diagonais $\mathfrak{W}_1 = \text{diag}\{\mathfrak{w}_{11}, \dots, \mathfrak{w}_{1n}\}$, $\mathfrak{W}_2 = \text{diag}\{\mathfrak{w}_{21}, \dots, \mathfrak{w}_{2n}\}$, $\mathfrak{W}_3 = \text{diag}\{\mathfrak{w}_{31}, \dots, \mathfrak{w}_{3n}\}$, $\mathfrak{W}_4 = \text{diag}\{\mathfrak{w}_{41}, \dots, \mathfrak{w}_{4n}\}$ de dimensão $n \times n$. Para $t = 1, \dots, n$, temos

$$\mathfrak{w}_{1t} = \left(\frac{-\mathbb{1}_{\{0,1\}}(y_t)}{(1 - \alpha_t)^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^2} \right) \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 + \left(\frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t} - \frac{\mathbb{1}_{\{0,1\}}(y_t)}{(1 - \alpha_t)} \right) \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t},$$

A.11. Apêndice 11

$$\mathbf{w}_{2t} = \begin{cases} -\phi^2 w_t \left(\frac{d\mu_t}{d\eta_t} \right)^2 + \phi [y_t^* - \mu_t^*] \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t}, & \text{se } y_t \in (0, 1), \\ 0, & \text{se } y_t = c, \end{cases}$$

$$\mathbf{w}_{3t} = \begin{cases} -\{c_t - (y_t^* - \mu_t^*)\}, & \text{se } y_t \in (0, 1), \\ 0, & \text{se } y_t = c, \end{cases}$$

e

$$\mathbf{w}_{4t} = \begin{cases} -d_t, & \text{se } y_t \in (0, 1), \\ 0, & \text{se } y_t = c. \end{cases}$$

Em notação matricial temos que as segundas derivadas da função de log-verossimilhança $\ell(\theta)$ são: $U_{\gamma\gamma} = Z^\top \mathfrak{W}_1 Z$, $U_{\beta\beta} = X^\top \mathfrak{W}_2 X$, $U_{\phi\beta} = U_{\beta\phi}^\top = X^\top \mathfrak{W}_3$ e $U_{\phi\phi} = \text{tr}(\mathfrak{W}_4)$. Desta forma, a matriz hessiana do modelo de regressão beta inflacionado em zero ou um é

$$\ddot{\ell} = \begin{pmatrix} U_{\gamma\gamma} & 0 & 0 \\ 0 & U_{\beta\beta} & U_{\beta\phi} \\ 0 & U_{\phi\beta} & U_{\phi\phi} \end{pmatrix}. \quad (\text{A.11.1})$$

No modelo de regressão beta inflacionado em zero ou um temos que $\theta = (\gamma^\top, \beta^\top, \phi)^\top$, então Δ é uma matriz de dimensão $(M + m + 1) \times n$ dada por

$$\Delta = \begin{pmatrix} \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_1 \partial \gamma_1} & \cdots & \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_n \partial \gamma_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_1 \partial \gamma_M} & \cdots & \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_n \partial \gamma_M} \\ \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_1 \partial \beta_1} & \cdots & \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_n \partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_1 \partial \beta_m} & \cdots & \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_n \partial \beta_m} \\ \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_1 \partial \phi} & \cdots & \frac{\partial^2 \ell(\theta|\omega)}{\partial \omega_n \partial \phi} \end{pmatrix},$$

avaliada em $\hat{\theta} = (\hat{\gamma}^\top, \hat{\beta}^\top, \hat{\phi})^\top$ e δ_0 . Note que é possível particionar Δ na forma $\Delta = (\Delta_\gamma^\top, \Delta_\beta^\top, \Delta_\phi)^\top$, em que

$$\begin{aligned} \Delta_\gamma &= \partial^2 \ell(\theta|\omega) / \partial \gamma \partial \omega^\top \Big|_{\theta=\hat{\theta}, \delta=\omega_0}, \\ \Delta_\beta &= \partial^2 \ell(\theta|\omega) / \partial \beta \partial \omega^\top \Big|_{\theta=\hat{\theta}, \omega=\delta_0}, \\ \Delta_\phi &= \partial^2 \ell(\theta|\omega) / \partial \phi \partial \omega^\top \Big|_{\theta=\hat{\theta}, \omega=\delta_0}, \end{aligned} \quad (\text{A.11.2})$$

sendo $\hat{\theta}$ o estimador de máxima verossimilhança de θ e ω um vetor de perturbações.

Ponderação de casos

No esquema de ponderação de casos temos que $\ell(\theta|\omega) = \ell_1(\gamma|\omega) + \ell_2(\beta, \phi|\omega)$. Desta forma, a t -ésima linha de Δ , $t = 1, \dots, n$, é da forma

$$\Delta_t = \left(\frac{\partial \ell_t(\alpha_t)}{\partial \gamma_1}, \dots, \frac{\partial \ell_t(\alpha_t)}{\partial \gamma_M}, \frac{\partial \ell_t(\mu_t, \phi)}{\partial \beta_1}, \dots, \frac{\partial \ell_t(\mu_t, \phi)}{\partial \beta_m}, \frac{\partial \ell_t(\mu_t, \phi)}{\partial \phi} \right)^\top \Big|_{\theta=\hat{\theta}}$$

Com base em (3.2.6), (3.2.9) e (3.2.13) temos para $R = 1, \dots, M$ e $r = 1, \dots, m$,

$$\begin{aligned} \frac{\partial \ell_t(\alpha_t)}{\partial \gamma_R} &= \frac{\mathbb{1}_{\{c\}}(y_t) - \alpha_t}{\alpha_t(1 - \alpha_t)} \frac{d\alpha_t}{d\zeta_t} z_{tR}, \\ \frac{\partial \ell_t(\mu_t, \phi)}{\partial \beta_r} &= \phi(1 - \mathbb{1}_{\{c\}}(y_t))(y_t^* - \mu_t^*) \frac{d\mu_t}{d\eta_t} x_{tr} \end{aligned}$$

e

$$\frac{\partial \ell_t(\mu_t, \phi)}{\partial \phi} = (1 - \mathbb{1}_{\{c\}}(y_t))u_t,$$

em que $u_t = \{\mu_t(y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi)\}$. Se definimos as matrizes diagonais $\mathcal{E}_1 = \text{diag}\{(\mathbb{1}_{\{c\}}(y_1) - \alpha_1), \dots, (\mathbb{1}_{\{c\}}(y_n) - \alpha_n)\}$ e $\mathcal{E}_2 = \text{diag}\{(y_1^* - \mu_1^*), \dots, (y_n^* - \mu_n^*)\}$ e o vetor $u = (u_1, \dots, u_n)^\top$, podemos escrever matricialmente

$$\begin{aligned} \Delta_\gamma &= \partial \ell_1(\gamma) / \partial \gamma \Big|_{\theta=\hat{\theta}} = Z^\top \widehat{P} \widehat{G} \widehat{\mathcal{E}}_1, \\ \Delta_\beta &= \partial \ell_2(\beta, \phi) / \partial \beta \Big|_{\theta=\hat{\theta}} = \widehat{\phi} X^\top H \widehat{T} \widehat{\mathcal{E}}_2, \\ \Delta_\phi &= \partial \ell_2(\beta, \phi) / \partial \phi \Big|_{\theta=\hat{\theta}} = \widehat{u}. \end{aligned} \tag{A.11.3}$$

O chapéu ‘ $\widehat{\cdot}$ ’ indica que as quantidades estão sendo avaliadas nos respectivos estimadores de máxima verossimilhança.

Perturbação nas variáveis explanatórias

Sejam Z e X matrizes de covariáveis do componente discreto e contínuo do modelo de regressão beta inflacionado em zero ou um respectivamente. Assumindo que $Z \neq X$ e a p -ésima covariável, z_p , $p = 1, \dots, M$ é perturbada aditivamente, i.e., da forma definida em (5.3.8), temos que a função de log-verossimilhança do modelo perturbado é da forma $\ell(\theta|\omega) = \ell_1(\gamma|\omega) + \ell_2(\beta, \phi)$, em que $\ell_1(\gamma|\omega) = \sum_{t=1}^n \ell_t(\alpha_t(\omega))$. Assim,

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \gamma_R \partial \omega_t} &= \frac{\partial^2 \ell_1(\gamma|\omega)}{\partial \gamma_R \partial \omega_t} = \frac{\partial}{\partial \alpha_t(\omega)} \left(\frac{\partial \ell_t(\alpha_t(\omega))}{\partial \alpha_t(\omega)} \frac{\partial \alpha_t(\omega)}{\partial \zeta_t(\omega)} \frac{\partial \zeta_t(\omega)}{\partial \omega_t} \right) \frac{\partial \alpha_t(\omega)}{\partial \zeta_t(\omega)} \frac{\partial \zeta_t(\omega)}{\partial \gamma_R} \\ &= \left\{ \frac{\partial^2 \ell_t(\alpha_t(\omega))}{\partial \alpha_t(\omega)^2} \left(\frac{\partial \alpha_t(\omega)}{\partial \zeta_t(\omega)} \right)^2 + \frac{\partial \ell_t(\alpha_t(\omega))}{\partial \alpha_t(\omega)} \left(\frac{\partial}{\partial \alpha_t(\omega)} \frac{\partial \alpha_t(\omega)}{\partial \zeta_t(\omega)} \right) \frac{\partial \alpha_t(\omega)}{\partial \zeta_t} \right\} \frac{\partial \zeta_t(\omega)}{\partial \gamma_R} \gamma_p^S z_p. \end{aligned}$$

Definindo a matriz diagonal $\mathfrak{W}_1(\omega) = \text{diag}\{\mathfrak{w}_{11}(\omega), \dots, \mathfrak{w}_{1n}(\omega)\}$, em que

$$\mathfrak{w}_{1t}(\omega) = \left(\frac{-\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t(\omega))^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t(\omega)^2} \right) \left(\frac{\partial \alpha_t(\omega)}{\partial \zeta_t(\omega)} \right)^2 + \left(\frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t(\omega)} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1 - \alpha_t(\omega))} \right) \left(\frac{\partial}{\partial \alpha_t(\omega)} \frac{\partial \alpha_t(\omega)}{\partial \zeta_t(\omega)} \right) \frac{\partial \alpha_t(\omega)}{\partial \zeta_t(\omega)},$$

A.11. Apêndice 11

temos que

$$\frac{\partial^2 \ell(\theta)}{\partial \gamma_R \partial \omega_t} = \begin{cases} \mathbf{w}_{1t}(\omega) \gamma_p \mathbf{s}_{z_p} z_{tR}, & \text{se } R \neq p, \\ \mathbf{w}_{1t}(\omega) \gamma_p \mathbf{s}_{z_p} z_{tp} + \omega_t \mathbf{s}_{z_p}, & \text{se } R = p. \end{cases} \quad (\text{A.11.4})$$

Agora, assumindo que $Z \neq X$ e a p -ésima covariável, x_p , $p = 1, \dots, m$ do componente contínuo é perturbada aditivamente, i.e., da forma definida em (5.3.9), temos que a função de log-verossimilhança do modelo perturbado é da forma $\ell(\theta|\omega) = \ell_1(\gamma) + \ell_2(\beta, \phi|\omega)$, em que $\ell_2(\beta, \phi|\omega) = \sum_{t=1}^n \ell_t(\mu_t(\omega), \phi)$. Assim,

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \beta_r \partial \omega_t} &= \frac{\partial^2 \ell_2(\beta, \phi|\omega)}{\partial \beta_r \partial \omega_t} = \frac{\partial}{\partial \mu_t(\omega)} \left(\frac{\partial \ell_t(\mu_t(\omega))}{\partial \mu_t(\omega)} \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \frac{\partial \eta_t(\omega)}{\partial \omega_t} \right) \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \frac{\partial \eta_t(\omega)}{\partial \beta_r} \\ &= \left\{ \frac{\partial^2 \ell_t(\mu_t(\omega))}{\partial \mu_t(\omega)^2} \left(\frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \right)^2 + \frac{\partial \ell_t(\mu_t(\omega))}{\partial \mu_t(\omega)} \left(\frac{\partial}{\partial \mu_t(\omega)} \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \right) \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \right\} \frac{\partial \eta_t(\omega)}{\partial \beta_r} \beta_p \mathbf{s}_{x_p}. \end{aligned}$$

Definindo a matriz diagonal $\mathfrak{M}_2(\omega) = \text{diag}\{\mathbf{w}_{21}(\omega), \dots, \mathbf{w}_{2n}(\omega)\}$, sendo

$$\mathbf{w}_{2t}(\omega) = \begin{cases} -\phi^2 w_t(\omega) \left(\frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \right)^2 + \phi [y_t^* - \mu_t^*(\omega)] \left(\frac{\partial}{\partial \mu_t(\omega)} \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \right) \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)}, & \text{se } y_t \in (0, 1), \\ 0, & \text{se } y_t = c, \end{cases}$$

em que $w_t(\omega) = \psi'(\mu_t(\omega)\phi) + \psi'(1 - \mu_t(\omega)\phi)$ e

$$\mu_t^*(\omega) = \begin{cases} \psi(\mu_t(\omega)\phi) - \psi((1 - \mu_t(\omega)\phi)), & \text{se } y_t \in (0, 1), \\ 0, & \text{caso contrário,} \end{cases}$$

temos que

$$\frac{\partial^2 \ell(\theta)}{\partial \beta_r \partial \omega_t} = \begin{cases} \mathbf{w}_{2t}(\omega) \beta_p \mathbf{s}_{x_p} x_{tr}, & \text{se } r \neq p, \\ \mathbf{w}_{2t}(\omega) \beta_p \mathbf{s}_{x_p} x_{tp} + \omega_t \mathbf{s}_{x_p}, & \text{se } r = p. \end{cases} \quad (\text{A.11.5})$$

Agora,

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \phi \partial \omega_t} &= \frac{\partial^2 \ell_2(\beta, \phi|\omega)}{\partial \phi \partial \omega_t} = \frac{\partial}{\partial \phi} \left(\frac{\partial \ell_t(\mu_t(\omega))}{\partial \mu_t(\omega)} \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \frac{\partial \eta_t(\omega)}{\partial \omega_t} \right) \\ &= \frac{\partial^2 \ell_t(\mu_t(\omega))}{\partial \phi \partial \mu_t(\omega)} \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \beta_p \mathbf{s}_{x_p}. \end{aligned}$$

Definindo o vetor $\mathfrak{M}_3(\omega) = (\mathbf{w}_{31}(\omega), \dots, \mathbf{w}_{3n}(\omega))^\top$, sendo

$$\mathbf{w}_{3t}(\omega) = \begin{cases} -\{\mathbf{c}_t(\omega) - (y_t^* - \mu_t^*(\omega))\}, & \text{se } y_t \in (0, 1), \\ 0, & \text{se } y_t = c, \end{cases}$$

em que $\mathbf{c}_t(\omega) = \phi[\mu_t(\omega)\psi'(\mu_t(\omega)\phi) - (1 - \mu_t(\omega))\psi'((1 - \mu_t(\omega)\phi))]$, temos que

$$\frac{\partial^2 \ell(\theta)}{\partial \phi \partial \omega_t} = \mathbf{w}_{3t}(\omega) \frac{\partial \mu_t(\omega)}{\partial \eta_t(\omega)} \beta_p \mathbf{s}_{x_p}. \quad (\text{A.11.6})$$

A.12 Apêndice 12

Cálculo de curvaturas para o modelo de regressão beta inflacionado em zero e um

Matriz Hessiana

Para o cálculo da matriz hessiana utilizamos as quantidades (A.1.2) –(A.1.4) obtidas no Apêndice A.1 e as quantidades (A.8.4) – (A.8.6) do Apêndice A.8. Desta forma, a matriz hessiana do modelo de regressão beta inflacionado em zero e um é

$$\ddot{\ell} = \begin{pmatrix} U_{\rho\rho} & U_{\rho\gamma} & 0 & 0 \\ U_{\gamma\rho} & U_{\gamma\gamma} & 0 & 0 \\ 0 & 0 & U_{\beta\beta} & U_{\beta\phi} \\ 0 & 0 & U_{\phi\beta} & U_{\phi\phi} \end{pmatrix}, \quad (\text{A.12.1})$$

em que

$$\begin{aligned} U_{\rho\rho} &= V^\top \text{diag}\{-\delta_{01}(1 - \delta_{01}), \dots, -\delta_{0n}(1 - \delta_{0n})\}V, \\ U_{\rho\gamma} &= U_{\gamma\rho}^\top = Z^\top \text{diag}\{\delta_{01}\delta_{11}, \dots, \delta_{0n}\delta_{1n}\}V, \\ U_{\gamma\gamma} &= Z^\top \text{diag}\{-\delta_{11}(1 - \delta_{11}), \dots, -\delta_{1n}(1 - \delta_{1n})\}Z, \\ U_{\beta\beta} &= X^\top \mathfrak{M}_2 X, \\ U_{\beta\phi} &= U_{\phi\beta}^\top = X^\top \mathfrak{M}_3, \\ U_{\phi\phi} &= \text{tr}(\mathfrak{M}_4), \end{aligned}$$

onde V , Z e X são matrizes de regressores conhecidos e \mathfrak{M}_2 , \mathfrak{M}_3 e \mathfrak{M}_4 são as matrizes definidas na subseção A.11.

Para o modelo RBIZU temos que $\theta = (\rho^\top, \gamma^\top, \beta^\top, \phi)^\top$, então Δ é uma matriz de dimensão $(k_0 + k_1 + k) \times n$ da forma $\Delta = (\Delta_\rho^\top, \Delta_\gamma^\top, \Delta_\beta^\top, \Delta_\phi)^\top$ em que

$$\begin{aligned} \Delta_\rho &= \partial^2 \ell(\theta|\omega) / \partial \rho \partial \omega^\top \Big|_{\theta=\hat{\theta}, \delta=\omega_0}, \\ \Delta_\gamma &= \partial^2 \ell(\theta|\omega) / \partial \gamma \partial \omega^\top \Big|_{\theta=\hat{\theta}, \delta=\omega_0}, \\ \Delta_\beta &= \partial^2 \ell(\theta|\omega) / \partial \beta \partial \omega^\top \Big|_{\theta=\hat{\theta}, \omega=\delta_0}, \\ \Delta_\phi &= \partial^2 \ell(\theta|\omega) / \partial \phi \partial \omega^\top \Big|_{\theta=\hat{\theta}, \omega=\delta_0}, \end{aligned} \quad (\text{A.12.2})$$

sendo $\hat{\theta}$ o estimador de máxima verossimilhança de θ e ω um vetor de perturbações.

Ponderação de casos

No esquema de ponderação de casos temos que $\ell(\theta|\omega) = \ell_1(\rho, \gamma|\omega) + \ell_2(\beta, \phi|\omega)$. Neste caso, Δ_t , a t -ésima linha da matriz Δ , $t = 1, \dots, n$, é da forma

$$\left(\frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \rho_1}, \dots, \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \rho_{k_0}}, \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \gamma_1}, \dots, \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \gamma_{k_1}}, \frac{\partial \ell_t(\mu_t, \phi)}{\partial \beta_1}, \dots, \frac{\partial \ell_t(\mu_t, \phi)}{\partial \beta_k}, \frac{\partial \ell_t(\mu_t, \phi)}{\partial \phi} \right)^\top \Big|_{\theta=\hat{\theta}}$$

A.12. Apêndice 12

Com base em (A.8.2), (A.8.3), (3.2.6), (3.2.9) e (3.2.13), respectivamente, temos para $r' = 1, \dots, k_0$, $r'' = 1, \dots, k_1$ e $r = 1, \dots, k$ que

$$\begin{aligned}\frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \rho_{r'}} &= (\mathbb{1}_{\{0\}}(y_t) - \delta_{0t})v_{tr'}, \\ \frac{\partial \ell_t(\delta_{0t}, \delta_{1t})}{\partial \gamma_{r''}} &= (\mathbb{1}_{\{1\}}(y_t) - \delta_{1t})z_{tr''}, \\ \frac{\partial \ell_t(\mu_t, \phi)}{\partial \beta_r} &= \phi \mathbb{1}_{(0,1)}(y_t)(y_t^* - \mu_t^*) \frac{d\mu_t}{d\eta_t} x_{tr}, \\ \frac{\partial \ell_t(\mu_t, \phi)}{\partial \phi} &= \mathbb{1}_{(0,1)}(y_t)u_t,\end{aligned}$$

em que $u_t = \{\mu_t(y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi)\}$. Se definimos as matrizes diagonais $\mathcal{E}_0 = \text{diag}\{(\mathbb{1}_{\{0\}}(y_1) - \delta_{01}), \dots, (\mathbb{1}_{\{0\}}(y_n) - \delta_{0n})\}$, $\mathcal{E}_1 = \text{diag}\{(\mathbb{1}_{\{1\}}(y_1) - \delta_{11}), \dots, (\mathbb{1}_{\{1\}}(y_n) - \delta_{1n})\}$, $\mathcal{E}_2 = \text{diag}\{(y_1^* - \mu_1^*), \dots, (y_n^* - \mu_n^*)\}$ e o vetor $u = (u_1, \dots, u_n)^\top$, podemos escrever matricialmente

$$\begin{aligned}\Delta_\rho &= V^\top \widehat{\mathcal{E}}_0, \\ \Delta_\gamma &= Z^\top \widehat{\mathcal{E}}_1, \\ \Delta_\beta &= \widehat{\phi} X^\top \text{mddiag}(y_{(0,1)}) \widehat{T} \widehat{\mathcal{E}}_2, \\ \Delta_\phi &= \widehat{u},\end{aligned}\tag{A.12.3}$$

lembrando que y^* , μ^* foram definidos em (3.2.7), (3.2.8) respectivamente. Aqui, $y_{(0,1)} = (\mathbb{1}_{(0,1)}(y_1), \dots, \mathbb{1}_{(0,1)}(y_n))^\top$ e $\text{mddiag}(\cdot)$ é o operador que transforma um vetor numa matriz diagonal. Novamente, o chapéu ‘ $\widehat{\cdot}$ ’ indica que as quantidades estão sendo avaliadas nos respectivos estimadores de máxima verossimilhança.

Referências Bibliográficas

- [1] Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, **50**, 901–908.
- [2] Aitchison, J. (1969). *The Lognormal Distribution*. Cambridge: Cambridge University Press.
- [3] Akaike, H. (1974). A new look at the statistical model identification. *IEEE. Transactions on Automatic Control*, **19**, 716–723.
- [4] Akaike, H. (1983). Information measures and model selection. *Bulletin of International Statistic Institute*, **50**, 277–290.
- [5] Anderson, J. A. & Richardson, S. C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, **22**, 621–627.
- [6] Atkinson, A.C. (1981). Two graphical display for outlying and influential observations in regression. *Biometrika*, **68**, 13–20.
- [7] Atkinson, A. C. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. New York: Oxford University Press.
- [8] Aubin, E. C. Q. & Cordeiro, G. M. (1997). Bias in linear regression models with unknown covariance matrix. *Communications in Statistics, Simulation and Computation*, **26**, 813–828.
- [9] Bartlett, M. S. (1953). Aproximate confidence intervals. *Biometrika*, **40**, 12–19.
- [10] Bickel, P. J. & Doksum, K. A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics*. 2nd ed. New York: Prentice Hall.
- [11] Bowman, K. & Shenton, L. R. (1965). Biases and covariances of maximum likelihood estimators. *Report K-1633, Union Carbide Corporation, Oak Ridge*.
- [12] Botter, D. A. & Cordeiro, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Communications in Statistics, Simulation and Computation*, **59**, 1–14.
- [13] Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of Royal Statistical Society B*, **26**, 211–246.
- [14] Box, M. (1971). Bias in nonlinear estimation (with discussion). *Journal of the Royal Statistical Society B*, **33**, 171–201.
- [15] Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics*, **53**, 603–618.

REFERÊNCIAS BIBLIOGRÁFICAS

- [16] Burnham, K. P., Anderson, D. R. (2002). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. 2nd ed. New York: Springer.
- [17] Bury, K. (1999). *Statistical Distributions in Engineering*. New York: Cambridge University Press.
- [18] Cole, T. & Green, P. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, **11**, 1305–1319.
- [19] Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, **19**, 15–18.
- [20] Cook, R.D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B*, **48**, 133–169.
- [21] Cook, R. D., Tsai, C. & Wei, B. (1986). Bias in nonlinear regression. *Biometrika*, **73**, 615–623.
- [22] Cook, D. O., Kieschnick, R. & McCullough, B. D. (2006). On the heterogeneity of corporate capital structures and its implications. Disponível em <http://ssrn.com/abstract=671061>.
- [23] Copas, J. B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society B*, **50**, 225–265.
- [24] Cordeiro, G. M. (1993). Barlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, Theory and Methods*, **22**, 169–188.
- [25] Cordeiro, G. M. (1999). *Introdução à Teoria Assintótica*. Notas do 22^o Colóquio Brasileiro de Matemática. Rio de Janeiro, Brasil.
- [26] Cordeiro, G. M. & Cribari-Neto, F. (1993). On Bartlett corrections, bias reduction and a new class of transformations. *Brazilian Journal of Probability and Statistics*, **7**, 179–200.
- [27] Cordeiro, G. M., & Cribari-Neto, F. (1998). On bias reduction in exponential and non-exponential family regression models. *Communications in Statistics, Simulation and Computation*, **27**, 485–500.
- [28] Cordeiro, G. M. & Klein, R. (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, **19**, 169–176.
- [29] Cordeiro, G. M. & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B*, **53**, 629–643.
- [30] Cordeiro, G. M., Rocha, E. C., Rocha, J. G. C. & Cribari-Neto, F. (1997). Bias-corrected maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, **58**, 21–35.
- [31] Cordeiro, G. M. & Vasconcellos, K. L. P. (1997). Bias correction for a class of multivariate nonlinear regression models. *Statistics and Probability Letters*, **35**, 155–164.
- [32] Cordeiro, G. M., Vasconcellos, K. L. P. & Santos, M. L. (1998). On the second order bias of parameter estimates in nonlinear regressions models with Student t erros. *Journal of Statistical Computation and Simulation*, **60**, 363–378.
- [33] Cordeiro, G. M. & Vasconcellos, K. L. P. (1999). Second-Order Biases of the Maximum Likelihood Estimates in von Mises Regression Models. *Australian & New Zealand Journal of Statistics*, **41**, 189–198.
- [34] Cox, D. & Snell, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B*, **30**, 248–275.

REFERÊNCIAS BIBLIOGRÁFICAS

- [35] Cox, C. (1996). Nonlinear quasi-likelihood models: applications to continuous proportions. *Computational Statistics and Data Analysis*, **21**, 449–461.
- [36] Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall.
- [37] Cox, D. R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B*, **49**, 1–39.
- [38] Cribari-Neto, F., Botter, D. A., Cordeiro, G. M. & Ferrari, S. L. P. (1998). Bias reduction in one-parameter exponential family models. *Communications in Statistics, Simulation and Computation*, **27**, 761–782.
- [39] Cribari-Neto, F. & Vasconcellos, K. L. P. (2002). Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, **72**, 107–118.
- [40] Cribari-Neto, F. & Zarkos, S.G. (1999). R: yet another econometric programming environment. *Journal of Applied Econometrics*, **14**, 319–329.
- [41] Cribari-Neto, F. & Zarkos, S. G. (2003). Econometric and statistical computing using Ox. *Computational Economics*, **21**, 277–295.
- [42] Davison, A. C. & Gigli, A. (1989). Deviance residuals and normal scores plots. *Biometrika*, **76**, 211–221.
- [43] de Souza, F.A.M. & Paula, G.A. (2002). Deviance residuals for an angular response. *Australian and New Zealand Journal of Statistics*, **44**, 345–356.
- [44] Doornik, J.A. (2001). *Ox: An Object-Oriented Matrix Language*, 4th ed. London: Timberlake Consultants Press.
- [45] Dunn, P.K. & Smyth, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 1–10.
- [46] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26.
- [47] Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- [48] Espinheira, P. L., Ferrari, S. L. P. & Cribari-Neto, F. (2008a). On beta regression residuals. *Journal of Applied Statistics*, a aparecer.
- [49] Espinheira, P. L., Ferrari, S. L. P. & Cribari-Neto, F. (2008b). Influence diagnostics in beta regression *Computational Statistics & Data Analysis*, a aparecer.
- [50] Farhrmeir, L. & Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer: New York.
- [51] Federal Communications Commission (1993). FCC 93-177, Report and order and further notice of proposed rule making, *MM Docket 92-266 (3 May 1993)*, 6134.
- [52] Federal Communications Commission (1994). FCC 94-38, Second order on reconsideration, fourth report and order and fifth notice of proposed rulemaking, *MM Docket 92-266 (30 March 1994)*, 4277.
- [53] Ferrari, S. L. P., Botter, D. A., Cordeiro, G. M. & Cribari-Neto, F. (1996). Second and third order bias reduction for one-parameter family models. *Statistics and Probability Letters*, **30**, 339–345.
- [54] Ferrari, S. L. P. & Cribari-Neto, F. (1998). On bootstrap and analytical bias corrections. *Economics Letters*, **58**, 7–15.

REFERÊNCIAS BIBLIOGRÁFICAS

- [55] Ferrari, S. L. P. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions *Journal of Applied Statistics*, **7**, 799–815.
- [56] Feuerverger, A. (1979). On some methods of analysis for weather experiments. *Biometrika*, **66**, 665–668.
- [57] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.
- [58] Fryer, J. G. & Robertson, C. A. (1972). A comparison of some methods for estimating mixed normal distributions. *Biometrika*, **59**, 639–648.
- [59] Galea, M., Paula, G.A. & Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, **46**, 71–79.
- [60] Galea, M., Paula, G.A. & Uribe-Opazo, M. (2003). On influence diagnostics in univariate elliptical linear regression models. *Statistical Papers*, **44**, 23–45.
- [61] Graham, V. A. & Hollands, K. G. T. (1990). Method to generate synthetic hourly solar radiation globally. *Solar Energy*, **44**, 333–341.
- [62] Haldane, J. B. S. (1953). The estimation of two parameters from a sample. *Sankhyā*, **12**, 313–320.
- [63] Haldane, J. B. S. & Smith, S. M. (1956). The sampling distribution of a maximum likelihood estimate. *Biometrika*, **43**, 96–103.
- [64] Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, **56**, 1030–1039.
- [65] Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. London: Chapman & Hall.
- [66] Heller, G., Stasinopoulos, M. & Rigby, B. (2006). The zero-adjusted Inverse Gaussian distribution as a model for insurance claims. *Proceedings of the 21th International Workshop on Statistical Modelling*, eds J. Hinde, J. Einbeck and J. Newell, 226–233, Galway: Ireland.
- [67] Hoff, A. (2007). Second stage DEA: Comparison of approaches for modelling the DEA escore. *European Journal of Operational Research*, **181**, 425–435.
- [68] Horowitz, J. L. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, **61**, 395–411.
- [69] Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.
- [70] Janardan, K. G. & Padmanabhan, G. (1986). Double bounded beta distribution for hydrologic variables. *Proceedings of 17th Annual Pittsburgh Conference*, **17**, 1107–1111.
- [71] Jansen, I., Molenberghs, G., Aerts, G., Thijs, H. & Steen, K. V. (2003). A local influence approach applied to binary data from a psychiatric study. *Biometrics*, **59**, 410–419.
- [72] Johnson, N., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions*. 2nd ed. New York: John Wiley and Sons.
- [73] Kieschnick R. & McCullough, B. D. (2003). Regression analysis of variates observed on (0,1): percentages, proportions, and fractions. *Statistical Modelling*, **3**, 1–21.

REFERÊNCIAS BIBLIOGRÁFICAS

- [74] Klein, J.P. & Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*, New York: Springer.
- [75] Knuth, D. E. (1986). *The T_EX book*. New York: Addison–Wesley.
- [76] Kwan, C. W. & Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: Applications to factor analysis. *Psychometrika*, **63**, 35–46.
- [77] L^AT_EX L^AT_EX (1985). *L^AT_EX Document Preparation System User's Guide and Reference Manual*, New York: Addison–Wesley.
- [78] Landsman, Z. M. & Makov, U. E. (2003). Contaminated exponential dispersion loss models. *North American Actuarial Journal*, **7**, 116–127.
- [79] Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295–303.
- [80] Lawrence, A. J. (1998). Regression transformation diagnostics using local influence. *Journal of the American Statistical Association*, **84**, 125–141.
- [81] Lehmann, E. L. & Casella, G. (2002). *Theory of Point Estimation*, 2nd ed. New York: Springer.
- [82] Lesaffre, E., & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.
- [83] Lesaffre, E., Rizoupoulos, D. & Tsonaka, S. (2007). The logistic-transform for bounded outcome scores. *Biostatistics*, **8**, 72–85.
- [84] Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications. NSF-CMBS Regional Conference Series in Probability and Statistics*. Vol. 5. Institute of Mathematical Statistics, Hayward, California.
- [85] Linhart, H. & Zucchini, W., (1986). *Model Selection*. John Wiley & Sons: New York.
- [86] Liu, S.Z. (2000). On local influence for elliptical linear models. *Statistical Papers*, **41**, 211–224.
- [87] Liu, S. Z. (2002). Local influence in multivariate elliptical linear regression models. *Linear Algebra and its Applications*, **354**, 159–174.
- [88] MacKinnon, J. G. & Smith, Jr. A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, **85**, 205–230.
- [89] Maffet, A. L. & Wackerman, C. C. (1991). The modified beta density function as a model for synthetic aperture radar clutter statistics. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 277-283.
- [90] Manly, B. F. J. (1976). Exponential data transformation. *The Statistician*, **25**, 37–42.
- [91] Marsaglia, G. (1997). A random number generator for C. Mensagem circulada no grupo de discussão do sci.stat.math em 29 de setembro de 1997.
- [92] McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.
- [93] McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.
- [94] McLachlan, G. J. (1980). A note on bias correction in maximum likelihood estimation with logistic discrimination. *Technometrics*, **21**, 71–78.

REFERÊNCIAS BIBLIOGRÁFICAS

- [95] McNally, R. J. (1990). Maximum likelihood estimation of the parameters of the prior distributions of three variables that strongly influence reproductive performance in cows. *Biometrics*, **446**, 501–514.
- [96] Milyutin, E. R. & Yaromenko, Y. I. (1991). Statistical characteristics of atmospheric transparency index over tilted routes. *Meteorologiya i Gidrologiya*, **12**, 72–76.
- [97] Mittelhammer, R.C., Judge, G.G. & Miller, D.J. (2000). *Econometric Foundations*. Cambridge University Press: NewYork.
- [98] Nocedal, J. & Wright, S. J. (1999). *Numerical Optimization*. New York: Springer-Verlag.
- [99] Ortega, E. M. M., Bolfarine, H. & Paula, G. A. (2003). Influence diagnostic in generalized loggamma regression models. *Computational Statistics & Data Analysis*, **42**, 165–186.
- [100] Ospina, R. (2006). The zero-inflated beta distribution for fitting a GAMLSS. Available at `gamlss.dist`: Extra distributions to be used for GAMLSS modelling. <http://cran.r-project.org/src/contrib/>.
- [101] Ospina, R., Cribari-Neto, F. & Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis*, **51**, 960–981.
- [102] Pace L. & Salvan, A. (1997). *Principles of Statistical Inference*. Singapore: World Scientific in Advanced Series on Statistical Science & Applied Probability, Vol. 4, 1997.
- [103] Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis*, **9**, 325–346.
- [104] Papke L. & Wooldridge J. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, **11**, 619–632.
- [105] Paula, G. A. (1992). Bias correction for exponential family nonlinear models. *Journal of Statistical Computation and Simulation*, **40**, 43–54.
- [106] Paula, G. A. (1993). Assessing local influence in restricted regression models. *Computational Statistics & Data Analysis*, **16**, 63–79.
- [107] Paula, G.A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation*, **51**, 315–352.
- [108] Paula, G.A. (1996). Influence diagnostics in proper dispersion models. *Australian Journal of Statistics*, **38**, 307–316.
- [109] Paula, G. A. & Cordeiro, G. M. (1995). Bias correction and improved residuals for non-exponential family nonlinear models. *Communications in Statistics, Theory and Methods*, **24**, 1193–1210.
- [110] Piepho, H. P. (2003). The folded exponential transformation for proportions. *The Statistician*, **52**, 575–589.
- [111] Pierce, D. A. & Schafer D. W. (1986). Residuals in Generalized Linear Models. *Journal of the American Statistical Association*, **81**, 977–986
- [112] Poon, W. Y. & Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B*, **61**, 51–61.
- [113] Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, **9**, 705–724.

REFERÊNCIAS BIBLIOGRÁFICAS

- [114] Press, W. H., Teulosky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Cambridge: Cambridge University Press.
- [115] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications* 2nd ed. New York: Wiley.
- [116] Rigby, R. A. & Stasinopoulos D. M. (2005). Generalized additive models for location, scale and shape (with discussion), *Applied Statistics*, **54**, 507–554.
- [117] Ridout, M., Demétrio, C. G. B. & Hinde, J. (1998). Models for count data with many zeros. International Biometric Conference – IBC 98, Cape Town. Invited Papers. Cape Town, África do Sul, 179–192.
- [118] Robertson, C. A. & Fryer, J. G. (1970). The bias and accuracy of moment estimators. *Biometrika*, **57**, 57–65.
- [119] Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill: Japan.
- [120] SAS Institute (2005). *The GLIMMIX Procedure*. Cary, NC: SAS Institute.
- [121] Schwarz, G. (1978). Estimating the dimension of a mode. *Annals of Statistics*, **6**, 461–464.
- [122] Seshadri, V. (1991). *Finite Mixtures of Natural Exponential Families*. Chapman & Hall: New York.
- [123] Shenton, L. R. & Wallington, P. A. (1962). The bias of moment estimators with application to the negative binomial distribution. *Biometrika*, **49**, 193–204.
- [124] Shenton, L. R. & Bowman, K. (1963). Higher moments of a maximum likelihood estimate. *Journal of the Royal Statistical Society B*, **25**, 305–317.
- [125] Shenton, L. R. & Bowman, K. (1977). *Maximum Likelihood Estimation in Small Samples*. London: Charles Griffin.
- [126] Shi, L. & Wang, X. (1999). Local influence in ridge regression. *Computational Statistics & Data Analysis*, **31**, 341–353.
- [127] Smithson, M., & Verkuilen, J. (2006). A better lemon-squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods*, **11**, 54–71.
- [128] Song, P. X. K. & Tan, M. (2000). Marginal models for longitudinal continuous proportional data. *Biometrics*, **56**, 496–502.
- [129] Song, P. X. K. Qiu, Z. & Tan, M. (2004). Modelling heterogeneous dispersion in marginal models for longitudinal proportional data. *Biometrical Journal*, **46**, 540–553.
- [130] Sowden, R. R. (1972). On the first-order bias of parameter estimates in a quantal response model under alternative estimation procedures. *Biometrika*, **59**, 573–579.
- [131] Stasinopoulos D. M., Rigby R. A. & Akantziliotou C. (2006). Instructions on how to use the GAMLSS package in R. Accompanying documentation in the current GAMLSS help files (see also <http://www.londonmet.ac.uk/gamlss/>).
- [132] Sum, S. T. & Oommen B. J. (1994). Mixture decomposition for distributions from the exponential family using a generalized method of moments. *International Symposium on Computer and Information Sciences IX*. Antalya, Turkey.

REFERÊNCIAS BIBLIOGRÁFICAS

- [133] Tsai, C.H., Wu, X. (1992). Assessing local influence in linear regression models with first-order autoregressive or heteroscedastic error structure. *Statistics and Probability Letters*, **14**, 247–252.
- [134] Tu, W. (2002). Zero inflated data. *Encyclopedia of Environmetrics*, **4**, 2387–2391.
- [135] Tu, W. & Zhou, X.H. (1999). A Wald test for the mean equality of log-normal data containing zeros. *Statistics in Medicine*, **55**, 645–651.
- [136] Vasconcellos, K. L. P. & Cordeiro, G. M. (1997a). Approximate bias for multivariate nonlinear heteroscedastic regressions. *Brazilian Journal of Probability and Statistics*, **11**, 141–159.
- [137] Vasconcellos, K. L. P. & Cordeiro, G. M. (1997b). Estimadores corrigidos para modelos SUR não lineares. *Revista de Econometria*, **17**, 45–65.
- [138] Vasconcellos, K. L. P. & Cordeiro, G. M. (1999). Bias corrected estimates in multivariate student t regression models. *Communications in Statistics, Simulation and Computation*, **29**, 797–822.
- [139] Vasconcellos, K. L. P. & Cribari-Neto, F. (2005). Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of Probability and Statistics*, **19**, 13–31.
- [140] Vieira, A.M.C., Hinde, J. P. & Demétrio, C. G. B. (2000). Zero-inflated proportion data models applied to a biological control assay. *Journal of Applied Statistics*, **27**, 373–389.
- [141] Wiley, J. A., Herschokoru, S. J. & Padiou, N. S. (1989). Heterogeneity in the probability of HIV transmission per sexual contact: the case of male-to-female transmission in penile-vaginal intercourse. *Statistics in Medicine*, **8**, 93–102.
- [142] Williams, D.A. (1984). Residuals in generalized linear models. In: *Proceedings of the 12th International Conference, Tokyo*. 59–68.
- [143] Williams, D.A. (1987). Generalized linear models diagnostic using the deviance and single case deletion. *Applied Statistics*, **36**, 181–191.
- [144] Xiang L, Yau K. W., Lee A. H. & Fung W. K. (2005). Influence diagnostics for two-component Poisson mixture regression models: applications in public health. *Statistics in Medicine*, **24**, 3053–3071.
- [145] Yee, T.W. & Wild, C. J. (1996). Vector generalized additive models. *Journal of The Royal Statistical Society B*, **58**, 481–493.
- [146] Yoo, S. (2004). A note on an approximation of the mobile communications expenditures distribution function using a mixture model. *Applied Statistics*, **31**, 747–752.
- [147] Young, D. & Bakir, S. (1987). Bias correction for a generalized log-gamma regression model. *Technometrics*, **29**, 183–191.
- [148] Zhu, H. T. & Lee, S. Y. (2001). Local influence for incomplete data models. *Journal of The Royal Statistical Society B*, **63**, 111–126.
- [149] Zhu, H. & Zhang, H. A. (2004). A diagnostic procedure based on local influence. *Biometrika*, **91**, 597–589.