

**Análise de dados com suporte limitado:
modelos *power logit* e
contribuições à inferência robusta**

Francisco Felipe de Queiroz

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

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES e do
CNPq

São Paulo, setembro de 2022

**Análise de dados com suporte limitado:
modelos *power logit* e
contribuições à inferência robusta**

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 11/08/2022. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof^a. Dr^a. Silvia Lopes de Paula Ferrari (orientadora) - IME-USP
- Prof. Dr. Gilberto Alvarenga Paula - IME-USP
- Prof. Dr. Artur José Lemonte - UFRN
- Prof. Dr. Francisco Cribari Neto - UFPE
- Prof. Dr. Mario de Castro Andrade Filho - ICMC-USP

*Dedico este trabalho à minha avó, Antônia,
e à minha mãe, Lúcia, pelos
ensinamentos e valores que me foram passados.*

Agradecimentos

A minha avó Antônia, minha mãe Lúcia, minha madrinha Nenzinha e meu tio João, pelo apoio, amor e por sempre acreditarem em mim. Essa conquista é nossa!

À minha orientadora, Silvia Ferrari, pela orientação impecável e segura, confiança, amizade, disponibilidade e dedicação. Obrigado pelos conselhos e ensinamentos compartilhados durante o doutorado. Tenho muito orgulho de ter sido seu orientando!

Aos professores do Departamento de Estatística da UFRN, em especial a Dione Valença, Artur Lemonte, André Pinho, Pledson Guedes, Carla Vivacqua, Damião Nóbrega, Iloneide Ramos, Luz Milena, Bruno Monte e Moisés Medeiros. Agradeço pela amizade, disponibilidade em ajudar e pelos conhecimentos compartilhados durante minha graduação e mestrado em Estatística. Tenho orgulho de ter sido aluno de vocês.

Aos professores do Departamento de Estatística do IME-USP, em especial aos professores Alexandre Patriota, Silvia Ferrari, Gilberto Paula e Júlia Pavan pela excelência no ensino e pelos conhecimentos compartilhados durante os cursos do Doutorado.

À Regiane, pela simpatia e eficiência com assuntos da pós-graduação do IME-USP.

Um agradecimento especial à Terezinha Ribeiro, pela grandiosa ajuda durante a pesquisa referente à segunda parte desta tese.

A segunda família que encontrei em São Paulo, Giovanni, Rosária e Gabriella. Vocês tornaram essa caminhada menos difícil. Em especial ao Giovanni, pelo companheirismo e amor que me tem dedicado.

À minha querida amiga Joelma, pelo amor, carinho e irmandade. Ao meu pequeno afilhado, Nicolas, pelos momentos de alegria que me proporciona. À Maisa, pelo carinho e por se fazer sempre presente. A Flávia, Inara, Joyce e Vanessa, pela amizade sincera. Aos amigos Carla, Lucas, Erika e Daniel pelo companheirismo.

Aos amigos que fiz durante o Doutorado em São Paulo, Luisa Borsato, Rodrigo Garcia, Cátia Tondolo, Gabriela Vasconcelos, Magno Tairone e Luiza Tuller. Em especial, à Luisa Borsato pela amizade sincera, disponibilidade em ajudar e suporte. Essa caminhada seria bem mais difícil sem vocês.

À minha professora de conversação em inglês do Cambly, Leah, pelas aulas durante quase três anos do doutorado, por me ajudar a melhorar o inglês e me sentir mais confiante.

Aos membros da banca examinadora pelo cuidado na leitura do texto e pelas sugestões e comentários.

À Capes e ao CNPq pelo auxílio financeiro.

Resumo

Queiroz, F. F. **Análise de dados com suporte limitado: modelos *power logit* e contribuições à inferência robusta**. 2022. 191f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2022.

Dados contínuos limitados, particularmente no intervalo unitário, aparecem em diferentes áreas, incluindo ecologia, biologia, economia e saúde pública. Alguns exemplos são a fração da cobertura vegetal, a proporção da renda familiar gasta em planos de saúde e a prevalência de doenças crônicas. Estes dados geralmente são altamente assimétricos, possuem dispersão dependendo da média e muitas vezes apresentam valores nas fronteiras. Modelos de regressão que utilizam a distribuição beta são amplamente empregados em aplicações. A regressão beta permite a interpretação direta dos parâmetros, acomoda assimetria e heterocedasticidade, sendo razoavelmente flexível. A inferência em modelos de regressão beta geralmente é baseada em métodos de máxima verossimilhança ou bayesianos, para os quais a informação dos dados vem da função de verossimilhança. Em ambos os casos, a inferência pode ser altamente influenciada por observações atípicas. O procedimento de inferência pode então ser substituído por um método robusto ou pode-se empregar modelos baseados em distribuições mais flexíveis do que a distribuição beta. Nesta tese, contribuimos para a modelagem estatística de dados limitados em duas direções. Primeiramente, definimos e estudamos os modelos *power logit*, uma classe altamente flexível de modelos de regressão com parâmetros interpretáveis adequados para modelagem de dados limitados com diferentes características. São apresentadas medidas de diagnóstico e de influência, e um novo pacote computacional é desenvolvido. Apresentamos também os modelos de regressão *power logit* inflacionados, que podem ser empregados quando os dados incluem observações em um dos extremos do suporte. A segunda parte desta tese é dedicada ao desenvolvimento de métodos inferenciais robustos em regressão beta inflacionada. Os estimadores propostos possuem boas propriedades e apresentaram bom desempenho em experimentos de simulação. Rotinas computacionais para uso dos estimadores propostos são fornecidas.

Palavras-chave: Dados fracionários, Inferência robusta, Proporções contínuas, Regressão beta, Regressão beta inflacionada.

Abstract

Queiroz, F. F. **Bounded continuous data: power logit models and contributions to robust inference**. 2022. 191f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2022.

Bounded continuous data, particularly on the unit interval, appear in different areas, including ecology, biology, economics, and public health. Some examples are the fraction of vegetation cover, the proportion of family income spent on health plans, and the prevalence of chronic illness. The data are usually highly skewed, have dispersion depending on the mean, and often present values at the boundaries. Regression models that use the beta distribution are widely employed in applications. Beta regression allows direct parameter interpretation, asymmetry and heteroscedasticity while reasonably flexible. Inference in beta regression models is usually based on maximum likelihood or Bayesian methods, for which the information from the data comes from the likelihood function. In either case, the inference can be highly influenced by atypical observations. The inference procedure may then be replaced by a robust method, or one may employ models based on more flexible distributions than the beta distribution. In this dissertation, we contribute to the statistical modeling of bounded data in two directions. First, we define and study the power logit models, a highly flexible class of regression models with interpretable parameters suitable for modeling bounded data with different characteristics. Diagnostic and influence measures are presented, and a new computational package is developed. We also present the inflated power logit regression models, which may be employed when the data include observations at one of the extremes of the support set. The second part of this dissertation is devoted to developing robust inference methods in inflated beta regression. The proposed estimators have good properties and performed well in simulation experiments. Computational routines for using the proposed estimators are provided.

Keywords: Beta regression, Continuous proportion, Fractional data, Inflated beta regression, Robust inference.

Sumário

Lista de Figuras	xiii
Lista de Tabelas	xix
1 Introdução	1
2 A classe de distribuições <i>power logit</i>	5
2.1 Preliminares	5
2.1.1 A classe simétrica de distribuições	5
2.1.2 A modelagem de dados no intervalo unitário	6
2.2 Modelos <i>power logit</i>	9
2.3 Inferência sobre os parâmetros da classe <i>power logit</i>	18
2.3.1 Inferência via máxima verossimilhança penalizada	22
2.3.2 Qualidade do ajuste e estimação de ζ	26
2.4 Resultados numéricos	27
2.4.1 Desempenho dos estimadores de máxima verossimilhança penalizada	28
2.4.2 Desempenho do estimador de ζ via minimização de Υ_ζ	30
2.4.3 Estudo de robustez	30
2.5 Aplicações	38
3 O modelo de regressão <i>power logit</i>	53
3.1 Definição	53
3.2 Inferência sobre os parâmetros do modelo	54
3.2.1 Intervalos de confiança e testes de hipóteses	56
3.2.2 Resultados numéricos	57
3.3 Diagnóstico no modelo de regressão <i>power logit</i>	59
3.3.1 Aplicações em dados simulados	62
3.3.2 Ponto de alavanca generalizado	66
3.4 Influência local no modelo de regressão <i>power logit</i>	67
3.5 Caso limite: o modelo de regressão log-log	70
3.6 Aplicação	72
4 O pacote PLreg	77
4.1 PLreg: apresentação e instalação	77
4.2 A distribuição <i>power logit</i> no pacote PLreg	77
4.3 Implementação do modelo de regressão <i>power logit</i> no pacote PLreg	79

4.4	Aplicações utilizando o pacote PLreg	83
4.4.1	Dados <code>bodyfat_Aeolus</code> : Ajustando a distribuição <i>power logit</i>	83
4.4.2	Dados <code>Firm</code> : Ajustando o modelo de regressão <i>power logit</i> com dispersão constante	86
4.4.3	Dados <code>bodyfat_Aeolus</code> : Ajustando o modelo limite	87
4.4.4	Dados <code>PeruVotes</code> : Ajustando modelos com dispersão variável e λ fixo	89
5	O modelo de regressão <i>power logit</i> inflacionado em zero ou um	93
5.1	A classe <i>power logit</i> inflacionada em zero ou um	93
5.2	O modelo de regressão <i>power logit</i> inflacionado em zero ou um	94
5.3	Diagnóstico e influência no modelo de regressão <i>power logit</i> inflacionado em zero ou um	98
5.4	Implementação computacional	100
5.5	Aplicação	100
6	Inferência robusta no modelo de regressão beta inflacionado	105
6.1	Preliminares	105
6.1.1	Conceitos em inferência robusta	105
6.1.2	Regressão para dados binários na presença de contaminação	107
6.1.3	Regressão beta na presença de contaminação	108
6.2	Inferência robusta no modelo de regressão beta inflacionado	109
6.2.1	O modelo de regressão beta inflacionado	110
6.2.2	Estimação via MDPD para κ	112
6.2.3	Estimação via maximização da L_q -verossimilhança reparametrizada para (β, γ)	114
6.2.4	Estimação via MDPD para (β, γ)	116
6.2.5	Estimadores robustos para o modelo de regressão beta inflacionado em zero ou um e suas propriedades	118
6.2.6	Testes de hipóteses robustos	123
6.2.7	Seleção da constante de afinação α	124
6.3	Aplicações em dados simulados	126
6.4	Resultados numéricos	132
7	Discussões e pesquisas futuras	145
A	A transformação <i>power logit</i>	147
B	Demonstrações	149
B.1	Demonstrações das proposições do Capítulo 2	149
B.2	Derivadas da função de log-verossimilhança no modelo de regressão <i>power logit</i>	152
C	Cálculo da matriz de perturbação Δ	155
C.1	Modelo de regressão <i>power logit</i>	155
C.1.1	Ponderação de casos	155
C.1.2	Perturbação individual nas covariáveis contínuas	156
C.2	Modelo de regressão <i>power logit</i> inflacionado em zero ou um	159

C.2.1	Ponderação de casos	159
C.2.2	Perturbação individual nas covariáveis contínuas	159
D	Gráficos	163
D.1	Gráficos de assimetria para as distribuições <i>power logit</i>	163
E	Resultados de simulações	167
E.1	Simulações caso iid - Capítulo 2	167
E.2	Simulações Capítulo 6	169
F	O modelo de regressão log-log	173
F.1	Estimação	174
F.2	Diagnóstico	175
G	Cálculo das matrizes de variâncias e covariâncias dos estimadores robustos	177
G.1	Cálculo da matriz de variâncias e covariâncias para o MDPDE no modelo de regressão para dados binários	177
G.2	Cálculo da matriz de variâncias e covariâncias para o LSMLE	178
	Referências Bibliográficas	185

Lista de Figuras

2.1	Histogramas (primeira linha) e boxplots (segunda linha) de y (dados originais), $\text{logit}(y)$ e $\text{logit}(y^\lambda)$; $\lambda = 0.11$.	10
2.2	Densidades PL-N.	11
2.3	Densidades PL- $t_{(2)}$.	12
2.4	Densidades PL- $t_{(\zeta)}$ para diferentes valores de ζ .	12
2.5	Densidades PL- $\text{PE}_{(1.5)}$.	13
2.6	Densidades PL- $\text{PE}_{(\zeta)}$ para diferentes valores de ζ .	13
2.7	Densidades PL- LO_1 .	13
2.8	Densidades PL-slash $_{(1.4)}$.	14
2.9	Densidades PL-slash $_{(\zeta)}$ para diferentes valores de ζ .	14
2.10	Densidades PL- $\text{SN}_{(1.5)}$.	15
2.11	Densidades PL- $\text{SN}_{(\zeta)}$ para diferentes valores de ζ .	15
2.12	Densidades PL- $\text{Hyp}_{(1.2)}$.	16
2.13	Densidades PL- $\text{Hyp}_{(\zeta)}$ para diferentes valores de ζ .	16
2.14	Medida de assimetria para a distribuição PL-N considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .	18
2.15	Medida de assimetria para a distribuição PL- $t_{(2)}$ considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .	18
2.16	Medida de assimetria para a distribuição PL- $t_{(\zeta)}$ considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ .	19
2.17	Função de log-verossimilhança perfilada relativa para λ (figura da esquerda) e curvas de contorno da função de verossimilhança relativa (figura da direita).	22
2.18	Função de log-verossimilhança perfilada relativa (linha sólida) e log-verossimilhança perfilada penalizada relativa (linha tracejada) para λ .	24
2.19	Função de log-verossimilhança perfilada relativa (linha sólida) e log-verossimilhança perfilada penalizada relativa (linha tracejada) para λ considerando amostras de tamanho $n = 50$ geradas dos modelos (a) PL-LO, (b) PL- $\text{PE}_{(1.2)}$, (c) PL- $\text{SN}_{(3)}$ e (d) PL- $\text{Hyp}_{(3)}$, considerando $\mu = 0.7$, $\sigma = 0.5$ e $\lambda = 1$.	25
2.20	Boxplots das estimativas de máxima verossimilhança (usual e penalizada) para λ , considerando o modelo PL- $\text{Hyp}_{(1.2)}$. A linha tracejada nos gráficos representa o valor verdadeiro $\lambda = 1$.	29
2.21	Histograma para os dados sem contaminação (y) e com contaminação (y contaminado) com densidades estimadas considerando os modelos beta, Kumaraswamy, GJS-N e PL-N – Cenário 1.	32

2.22 Gráfico de $r_i, i = 1, \dots, 200$, com envelopes simulados referentes aos ajustes dos modelos beta, Kumaraswamy, GJS-N e PL-N, considerando os dados sem contaminação (gráficos da esquerda) e com contaminação (gráficos da direita) – Cenário 1. 33

2.23 Valores de Υ_ζ para os ajustes das distribuições beta, Kumaraswamy, GJS e PL-N para cada uma das 100 amostras geradas de acordo com o Cenário 1. 34

2.24 Valores de Υ_ζ para os ajustes das distribuição PL-N, beta, Kumaraswamy e GJS para cada uma das 8 amostras geradas de acordo com o Cenário 1, com a porcentagem de observações contaminadas igual a 0%, 1%, 2%, 3%, 4%, 5%, 7% e 10%. 35

2.25 Histograma para os dados sem contaminação (y) e com contaminação (y contaminado) com densidades estimadas considerando os modelos beta, Kumaraswamy, GJS-N e PL-N – Cenário 2. 36

2.26 Gráfico de $r_i, i = 1, \dots, 200$, com envelopes simulados referentes aos ajustes dos modelos beta, Kumaraswamy, GJS-N e PL-N, considerando os dados sem contaminação (gráficos da esquerda) e com contaminação (gráficos da direita) – Cenário 2. 37

2.27 Valores de Υ_ζ para os ajustes das distribuições beta, Kumaraswamy, GJS-N e PL-N para cada uma das 100 amostras geradas de acordo com o Cenário 2. 38

2.28 Valores de Υ_ζ para os ajustes das distribuição PL-N, beta, Kumaraswamy e GJS para cada uma das 8 amostras geradas de acordo com o Cenário 2, com a porcentagem de observações contaminadas igual a 0%, 1%, 2%, 3%, 4%, 5%, 7% e 10%. 39

2.29 Boxplot para a variável y 39

2.30 Gráficos de resíduos com envelopes simulados para os ajustes apresentados na Tabela 2.10. 41

2.31 Gráfico de Υ_ζ e as versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo PL-PE $_{(\zeta)}$ 42

2.32 Histograma com densidades estimadas sobrepostas (a) e funções densidade de probabilidade empíricas e estimadas (b) para os dados de porcentagem de gordura corporal de morcegos. 42

2.33 Gráfico da distância quantílica relativa para os dados de porcentagem de gordura corporal de morcegos. 43

2.34 Boxplot para a variável y 43

2.35 Gráficos de resíduo com envelopes simulados para os ajustes apresentados na Tabela 2.12. 45

2.36 Funções densidade de probabilidade ajustas sobrepostas ao histograma de y , funções de distribuição empírica e ajustadas, e gráfico da distância quantílica relativa os modelos considerando os dados da proporção de indivíduos ocupados fora do setor agropecuário. 46

2.37 Gráfico das versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo PL-N. 47

2.38 Boxplot para a variável y 48

2.39 Funções densidade de probabilidade ajustas sobrepostas ao histograma de y , funções de distribuição empírica e ajustadas, e gráfico da distância quantílica relativa para alguns modelos considerando os dados da proporção de indivíduos ocupados com ensino médio completo. 49

2.40	Gráfico de resíduo com envelopes simulados para os ajustes apresentados na Tabela 2.14.	50
2.41	Gráfico de Υ_ζ e as versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo PL-t(ζ).	51
2.42	Gráfico de resíduo com envelopes simulados para os ajustes apresentados na Tabela 2.15, feitos após a exclusão dos casos #162 e #164.	52
3.1	Diagrama de dispersão com as curvas ajustadas (via máxima verossimilhança) para os dados completos (linha sólida) e para os dados contaminados (linha tracejada) (a), e gráficos dos resíduos quantílicos (b) e padronizados (c) com envelopes simulados para os dados contaminados; modelo de regressão PL-N.	62
3.2	Diagrama de dispersão com as curvas ajustadas (via máxima verossimilhança) para os dados completos (linha sólida) e para os dados contaminados (linha tracejada) (a), gráficos dos resíduos quantílico (b), deviance (c) e padronizados (d) com envelopes simulados para os dados contaminados, e gráfico da função $v(z)$ versus o resíduo padronizado (e) para os dados contaminados; o modelo de regressão PL-t(ζ).	63
3.3	Gráficos de resíduo <i>versus</i> o índice das observações para os ajustes dos modelos de regressão PL-PE _(1.5) (primeira linha) e PL-slash _(2.3) (segunda linha).	64
3.4	Gráficos de resíduo <i>versus</i> a covariável x para os ajustes dos modelos de regressão PL-SN _(1.6) (primeira linha) e PL-Hyp _(1.5) (segunda linha).	65
3.5	Gráfico da estatística da razão de verossimilhanças penalizadas perfilada para λ , $W_p^*(\lambda)$	73
3.6	Gráficos dos resíduos quantílico (a), deviance (b) e padronizados (c) com envelopes simulados para os modelos de regressão PL-slash (primeira linha) e GJS-slash (segunda linha) com dispersão constantes.	74
3.7	Gráficos de índices de $ h_{\max} $ para o esquema de ponderação de casos (a) e de GL_{ii} (b), e gráfico da função $v(z)$ versus o resíduo padronizado (c) para o modelo de regressão PL-slash com dispersão constante.	75
3.8	Gráficos de dispersão do <code>ind_risco</code> versus y , juntamente com as linhas ajustadas com base na estimativa de máxima verossimilhança (EMV) para os dados completos e os dados sem <i>outliers</i> , considerando o modelo de regressão beta com precisão variável (a) e o modelo de regressão PL-slash com dispersão constante (b).	75
4.1	Histograma de uma amostra aleatória de uma distribuição PL-SN _(2.5)	78
4.2	Função densidade de probabilidade de uma distribuição PL-Hyp ₍₂₎	79
4.3	Função de distribuição acumulada de uma distribuição PL-PE _(1.3)	79
4.4	Gráficos fornecidos pela função <code>extra.parameter()</code> para o ajuste <code>fitPL_PE.aux</code>	84
4.5	Função densidade de probabilidade estimada com base no ajuste <code>fitPL_SN</code>	85
4.6	Gráficos fornecidos pela função <code>extra.parameter()</code> para o ajuste <code>fitPL_slash.aux</code>	86
4.7	Gráfico de probabilidade normal com envelopes simulados (lado esquerdo) e gráfico de $v(z)$ <i>versus</i> o resíduo padronizado para o ajuste <code>fitPL_slash</code>	88
4.8	Gráficos fornecidos pela função <code>plot()</code> para o ajuste <code>fitloglog_PE</code>	90
4.9	Gráficos fornecidos pela função <code>influence()</code> para o ajuste <code>fitloglog_PE</code>	91

4.10	Gráficos de probabilidade normal com envelopes simulados fornecidos pela função <code>envelope()</code> para os ajustes <code>fitPL1</code> , <code>fitPL05</code> e <code>fitPL0</code> , respectivamente.	92
5.1	Densidades $\text{ZIPL-t}_{(2)}$ (primeira linha) e $\text{OIPL-slash}_{(2.5)}$ (segunda linha) para diferentes valores dos parâmetros.	95
5.2	Gráficos de resíduos com envelopes simulados para a parte discreta (a) e para a parte contínua considerando os modelos beta inflacionado (b) e PL normal inflacionado (c).102	
5.3	Gráficos de índice do resíduo deviance para a parte discreta (a), padronizado para a parte contínua (b) e quantílico para o modelo completo (c), respectivos gráficos de resíduo com envelopes simulados (d), (e) e (f), gráfico de índices de $ h_{\max} $ para o esquema de ponderação de casos para a parte contínua (g), diagrama de dispersão dos valores ajustados <i>versus</i> GL_{ii} para a parte contínua (h) e gráfico de índices da distância de Cook para a parte discreta (i).	104
6.1	Diagramas de dispersão da amostra (parte contínua) sem contaminação (esquerda) e da amostra contaminada segundo o primeiro cenário (direita).	127
6.2	Gráfico dos SQVs referentes aos estimadores LSMLE e LMDPDE (parte contínua) para os dados sem contaminação (esquerda) e contaminados segundo o Cenário 1 (direita).	128
6.3	Diagramas de dispersão da amostra sem contaminação (esquerda) e da amostra contaminada segundo o Cenário 2 (direita) para a parte discreta (primeira linha) e contínua (segunda linha).	129
6.4	Gráfico dos SQVs referentes ao estimador MDPDE (parte discreta) para os dados sem contaminação (esquerda) e contaminados segundo o Cenário 2 (direita).	129
6.5	Diagramas de dispersão da amostra sem contaminação (esquerda) e da amostra contaminada segundo o Cenário 3 (direita) para a parte discreta (primeira linha) e contínua (segunda linha).	130
6.6	Gráfico dos SQVs referentes ao estimador MDPDE (parte discreta, primeira linha) e aos estimadores LSMLE e LMDPDE (parte contínua, segunda linha) para os dados sem contaminação (esquerda) e contaminados segundo o Cenário 3 (direita).	131
6.7	Boxplots das estimativas dos parâmetros κ_1 , κ_2 e κ_3 sob o cenário 1: MLE (esquerda) e MDPDE (direita).	133
6.8	Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 1: MLE (esquerda), SMLE (centro) e MDPDE (direita).	134
6.9	Boxplots das estimativas dos parâmetros κ_1 , κ_2 e κ_3 sob o cenário 2: MLE (esquerda) e MDPDE (direita).	135
6.10	Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 2: MLE (esquerda), SMLE (centro) e MDPDE (direita).	136
6.11	Boxplots das estimativas dos parâmetros κ_1 , κ_2 e κ_3 sob o cenário 3: MLE (esquerda) e MDPDE (direita).	137
6.12	Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 3: MLE (esquerda), SMLE (centro) e MDPDE (direita).	138

6.13 Boxplots dos valores ótimos para a constante de afinação α para a parte discreta considerando o MDPDE e para a parte contínua considerando o LSMLE e o LMDPDE, sob os cenários 1 (primeira linha), 2 (segunda linha) e 3 (terceira linha) 139

6.14 Discrepância quantílica relativa para o teste tipo-Wald usual e robusto; Cenário 3 e $n = 100$ 142

6.15 Discrepância quantílica relativa para o teste tipo-Wald usual e robusto; Cenário 3 e $n = 200$ 143

A.1 Gráficos da média dos coeficientes de assimetria de \mathbf{y} , $\text{logit}(\mathbf{y})$ e $\text{logit}(\mathbf{y}^\lambda)$ para dados gerados de diferentes distribuições com suporte em $(0, 1)$ 148

D.1 Medida de assimetria para a distribuição PL-LO considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ 163

D.2 Medida de assimetria para a distribuição PL-PE_(1.5) considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ 163

D.3 Medida de assimetria para a distribuição PL-PE_(\zeta) considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ 164

D.4 Medida de assimetria para a distribuição PL-SN_(1.5) considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ 164

D.5 Medida de assimetria para a distribuição PL-SN_(\zeta) considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ 164

D.6 Medida de assimetria para a distribuição PL-Hyp_(1.2) considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ 165

D.7 Medida de assimetria para a distribuição PL-Hyp_(\zeta) considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ 165

E.1 Boxplots dos valores ótimos para a constante de afinação α para a parte contínua considerando o LSMLE e o LMDPDE, sob o cenário 1 (primeira linha), cenário 2 (segunda linha) e cenário 3 (terceira linha). 169

E.2 Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 1: MLE (esquerda), LSMLE (centro) e LMDPDE (direita). 170

E.3 Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 2: MLE (esquerda), LSMLE (centro) e LMDPDE (direita). 171

E.4 Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 3: MLE (esquerda), LSMLE (centro) e LMDPDE (direita). 172

Lista de Tabelas

2.1	Função peso para alguns modelos na classe <i>power logit</i>	20
2.2	Estimativas de máxima verossimilhança usual e penalizada para $\theta = (\mu, \sigma, \lambda)^\top$, considerando as amostras da Figura 2.19.	26
2.3	Estimativa do viés e erro-padrão (e.p.) das 3000 estimativas de máxima verossimilhança (usual e penalizada) para os modelos PL- $t_{(2)}$, PL-PE $_{(1.5)}$ e PL-Hyp $_{(1.2)}$	28
2.4	Mediana (M) e distância interquartilica (IR) das M réplicas para a escolha de ζ , considerando diferentes modelos na classe <i>power logit</i> e diferentes tamanhos de amostra.	30
2.5	Medidas descritivas para os dados (contaminados e sem contaminação) gerados de acordo com o Cenário 1.	31
2.6	Estimativas dos parâmetros (com respectivos erros-padrão), valores da medida Υ_ζ e AIC para cada modelo ajustado aos dados sem e com contaminação – Cenário 1.	31
2.7	Medidas descritivas para os dados (contaminados e sem contaminação) gerados de acordo com o Cenário 2.	35
2.8	Estimativas dos parâmetros (com respectivos erros-padrão), valores da medida Υ_ζ e AIC para cada modelo ajustado aos dados sem e com contaminação – Cenário 2.	36
2.9	Medidas resumo para a porcentagem de gordura corporal de morcegos (y).	39
2.10	Estimativas de máxima verossimilhança, intervalos de 95% confiança e valores de Υ_ζ para cada modelo ajustado aos dados de porcentagem de gordura corporal de morcegos.	40
2.11	Medidas resumo para a proporção de indivíduos ocupados fora do setor agropecuário (y).	43
2.12	Estimativas de máxima verossimilhança, intervalos de 95% de confiança e valores de Υ_ζ para cada modelo ajustado aos dados de proporção de indivíduos ocupados fora do setor agropecuário.	44
2.13	Medidas resumo para a proporção de indivíduos ocupados com ensino médio completo (y).	48
2.14	Estimativas de máxima verossimilhança, intervalos de 95% de confiança e valores de Υ_ζ para cada modelo ajustado aos dados de proporção de indivíduos ocupados com ensino médio completo.	48
2.15	Estimativas de máxima verossimilhança, intervalos de 95% de confiança e valores de Υ_ζ para cada modelo ajustado aos dados de proporção de indivíduos ocupados com ensino médio completo, após a exclusão dos casos #162 e #164.	51

3.1	Estimativas do viés e da raiz do erro quadrático médio ($\sqrt{\text{EQM}}$) das 3000 estimativas de máxima verossimilhança (usual e penalizada) para cada tamanho de amostra e para os modelos de regressão PL- $t_{(5)}$, PL-PE $_{(1.5)}$, PL-Hyp $_{(1.2)}$ e PL-slash $_{(1.4)}$	58
3.2	Valores de ξ_r e d_r para alguns modelos na classe <i>power logit</i>	61
3.3	Medidas resumo para as variáveis y , ind_risco e log_at	72
3.4	Estimativas (Est.), erros padrão assintóticos (e.p.) e valores p para o modelo de regressão PL-slash com dispersão variável.	72
3.5	Estimativas (Est.), erros padrão assintóticos (e.p.) e valores p para o modelo de regressão PL-slash com dispersão constante.	73
4.1	Componentes retornados pela função PLreg().	82
4.2	Funções e métodos para objetos da classe “PLreg”.	83
4.3	Estimativas e erros-padrão dos ajustes fitPL1, fitPL05 e fitPL0, respectivamente.	92
5.1	Estimativas (Est.), erros padrão assintóticos (e.p.) e valores p para o modelo de regressão IPL-NO com dispersão variável.	101
5.2	Estimativas (Est.), erros padrão assintóticos (e.p.) e valores p para o modelo de regressão beta inflacionado com precisão variável.	102
5.3	Estimativas (Est.), erros padrão assintóticos (e.p.) e valores p para o modelo de regressão reduzido PL-NO com precisão variável.	103
6.1	Estimativas dos parâmetros, com respectivos erros padrão, considerando diferentes métodos de estimação para os dados sem contaminação e contaminados de acordo com o Cenário 1.	128
6.2	Estimativas dos parâmetros, com respectivos erros padrão, considerando diferentes métodos de estimação para os dados sem contaminação e contaminados de acordo com o Cenário 2.	130
6.3	Estimativas dos parâmetros, com respectivos erros padrão, considerando diferentes métodos de estimação para os dados sem contaminação e contaminados de acordo com o Cenário 3.	131
6.4	Razão entre os erros quadráticos médios totais dos estimadores sob os Cenários 1, 2 e 3.	140
6.5	Níveis empíricos sob \mathcal{H}_0 do teste tipo-Wald robusto sob os Cenários 1, 2 e 3 ao nível nominal de 5%.	141
E.1	Média, estimativa do viés e erro-padrão das 3000 estimativas de máxima verossimilhança (usual e penalizada) para os modelos PL-N, PL-LA e PL-SN $_{(1.5)}$	167
E.2	Mediana (M) e distância interquartílica (IR) das M réplicas para a escolha de ζ	168

Capítulo 1

Introdução

Dados sobre frações (proporções) de natureza contínua aparecem frequentemente em estudos de diversas áreas como medicina, biologia, sociologia, psicologia, economia, entre muitas outras. Alguns exemplos são: fração da renda familiar gasta com planos de saúde, taxa de mortalidade, porcentagem de gordura no corpo, fração do território coberta por copas de árvores, etc.. É comum que o interesse resida em prever ou explicar o comportamento de proporções a partir de um conjunto de outras variáveis. Nestes casos, a variável resposta pode assumir valores no intervalo unitário e o modelo de regressão linear usual pode não ser apropriado, pois pode fornecer valores ajustados para a variável resposta de interesse que excedem os limites do intervalo (inferior ou superior). Uma solução alternativa e bastante utilizada é aplicar uma transformação na variável resposta, de modo que ela assuma valores na reta real e, então, modelar a média da variável resposta transformada. A transformação logito, definida por $h(y) = \log[y/(1 - y)]$, $y \in (0, 1)$, é uma das mais comuns. No entanto, essa abordagem também pode trazer alguns inconvenientes, entre eles: a interpretação dos parâmetros do modelo é dada em termos da variável transformada; os valores ajustados podem exceder os limites do intervalo; essas transformações não garantem que a variância será constante (homocedasticidade); entre outros.

A modelagem de dados contínuos que tomam valores no intervalo unitário, que chamaremos de dados fracionários, deve levar em conta alguns aspectos inerentes a dados dessa natureza. Por exemplo, tipicamente exibem assimetria e dispersão dependente da média. Assim, um enfoque mais natural para lidar com essas situações é utilizar algum modelo de regressão adequado para variável resposta contínua que assumam valores no intervalo unitário $(0, 1)$. Uma das principais distribuições de probabilidade que assume valores no intervalo unitário é a distribuição beta.

Desde a década de 2000, um número considerável de trabalhos tem focado em modelos de regressão baseados na distribuição beta, usualmente chamados de “modelos de regressão beta”. Entre eles, o artigo de [Ferrari e Cribari-Neto \(2004\)](#) é um dos mais conhecidos. Este trabalho apresenta 1361 citações no Web of Science (consulta feita em 08/07/2022). Os autores propuseram um modelo de regressão beta similar aos modelos lineares generalizados, em que se modela diretamente a média, ao invés dos dois parâmetros de forma que indexam a distribuição beta na parametrização usual; os autores desenvolveram estimação pontual e intervalar, testes de hipóteses e medidas de diagnóstico; é possível utilizar pacotes implementados no *software* R. Além disso, o modelo de regressão beta com precisão variável também foi desenvolvido ([Ferrari et al., 2011](#); [Smithson e Verkuilen, 2006](#)).

Existem diferentes especificações do modelo de regressão beta na literatura, por exemplo, modelos com erros nas variáveis ([Carrasco et al., 2014](#)), modelos para dados longitudinais ([Wang et al., 2014](#)), modelos para dados de alta dimensão ([Schmid et al., 2009](#)), entre outros. Adicionalmente, modelos de regressão para dados fracionários que usam outras distribuições aparecem em diversos trabalhos. Por exemplo, [Bayes et al. \(2012\)](#) apresentam o modelo de regressão beta retangular; a distribuição log-Lindley é estudada em [Gómez-Déniz et al. \(2014\)](#); [Lemonte e Bazán \(2016\)](#) propõem a classe de distribuições GJS; [Smithson e Shou \(2017\)](#) estudam a classe de distribuições *CDF-quantile*; entre muitos outros.

Em uma ampla variedade de aplicações envolvendo proporções, a variável de interesse pode

assumir não apenas valores no intervalo $(0, 1)$ como, também, os valores zero ou um ou ambos. Por exemplo, taxa de mortalidade ou infecção para uma determinada doença, proporção da renda familiar gasta com educação dos filhos e proporção de pessoas que apresentam uma determinada característica. Nessas situações, o modelo de regressão beta não é adequado. Uma solução mais apropriada é considerar um modelo estatístico que permita adicionar à distribuição contínua da variável resposta um ponto de massa em zero, em um, ou em ambos os extremos. Modelos de regressão beta que acomodam observações nos extremos são encontrados em [Ospina e Ferrari \(2012\)](#); ver também [Cook et al. \(2008\)](#), [Ospina e Ferrari \(2010\)](#) e [Pereira et al. \(2012\)](#). Outros modelos de regressão para dados de proporção que apresentam observações nos extremos também foram propostos considerando outras distribuições; veja, por exemplo, os trabalhos de [Liu et al. \(2020\)](#), [Queiroz e Lemonte \(2021\)](#), [Bayer et al. \(2021\)](#).

Inferência no modelo de regressão beta e vários de seus competidores é baseada em estimadores de máxima verossimilhança ou em métodos bayesianos, para os quais a informação proveniente dos dados é baseada na função de verossimilhança. Em ambos os casos, a inferência pode ser muito influenciada por observações atípicas e pela presença de pontos influentes. Existem diferentes formas de lidar com isso, uma delas é considerar modificações na forma de estimação de modo a obter métodos inferenciais robustos. Outra alternativa é considerar modelos flexíveis que sejam menos sensíveis a esse tipo de observações. Neste tese, estudamos essas duas perspectivas.

No contexto de modelos mais flexíveis, [Lemonte e Bazán \(2016\)](#) propõem uma nova classe de modelos de regressão denominada “classe de distribuições GJS” (*Generalized Johnson S_B*), que é uma generalização da distribuição Johnson SB ([Johnson, 1949](#)). A classe de distribuições GJS é indexada pela mediana e um parâmetro de dispersão. Essa classe pode modelar de forma razoavelmente simples dados que apresentem caudas com diferentes pesos a depender do gerador de densidades utilizado. Os autores também propuseram o modelo de regressão GJS. Como uma extensão, [Queiroz e Lemonte \(2021\)](#) propuseram o modelo de regressão GJS inflacionado em zero ou um para o caso em que a variável resposta assume valores em um dos extremos. Essencialmente, a classe de distribuições GJS é construída a partir de distribuições simétricas atribuídas ao logito da variável que tem suporte em $(0, 1)$. Entretanto, é sabido que a transformação logito não é capaz de levar distribuições comuns de frações contínuas à simetria.

A primeira parte desta tese dedica-se no desenvolvimento de uma classe flexível de modelos de regressão para dados contínuos que assumem valores no intervalo limitado, denominada classe *power logit*. Essa nova classe de distribuições tem três parâmetros, que representam mediana, dispersão e assimetria, e é obtida através de uma transformação conveniente aplicada a uma variável aleatória com distribuição simétrica. A classe *power logit* tem a classe GJS como um caso particular, com a vantagem de que o parâmetro de assimetria fornece flexibilidade extra e melhor acomoda dados altamente assimétricos. O modelo de regressão *power logit* é proposto e os parâmetros de regressão são interpretados em termos da mediana e da dispersão da variável resposta. Como o interesse está em dados assimétricos, a mediana pode ser uma medida de tendência central mais apropriada do que a média. A distribuição *power logit* inflacionada em zero ou um e o respectivo modelo de regressão também são desenvolvidos. Além disso, definimos medidas diagnósticas e de influência e apresentamos aplicações que mostram a utilidade dos modelos de regressão propostos na prática. Adicionalmente, desenvolvemos um novo pacote R, chamado *PLreg*, disponível no repositório CRAN, para ajustar os modelos *power logit* e GJS.

A segunda parte dessa tese concentra-se no desenvolvimento de métodos inferenciais robustos para a estimação no modelo de regressão beta inflacionado em zero ou um ([Ospina e Ferrari, 2012](#)). A distribuição beta inflacionada pode ser vista como uma mistura entre uma variável aleatória degenerada no valor c (com $c = 0$ ou $c = 1$) e uma variável com distribuição beta. A estimação pontual via máxima verossimilhança no modelo de regressão beta inflacionado é feita de forma separada para a parte discreta e contínua. Dessa forma, a inferência é fortemente influenciada por observações atípicas (assim como nos modelos de regressão para dados binários e em regressão beta). A influência de observações atípicas na inferência via máxima verossimilhança em regressão para dados binários tem sido estudada em diversos trabalhos nas últimas décadas; veja [Copas \(1988\)](#),

Pregibon (1982), Bianco e Yohai (1996), Croux *et al.* (2002) e Ghosh e Basu (2016). A falta de robustez na inferência via máxima verossimilhança no modelo de regressão beta foi estudada nos trabalhos Ghosh (2019), Ribeiro e Ferrari (2022) e Maluf *et al.* (2022). Nesses trabalhos, os autores propõem métodos de estimação robustos para inferência em modelos de regressão beta para o caso em que o suporte da variável resposta é $(0, 1)$, não englobando o caso em que há observações nos extremos do intervalo. Com base nesses trabalhos, desenvolvemos estimadores robustos para o modelo de regressão beta inflacionado.

Organização do trabalho

Esta tese está dividida em sete capítulos e, conforme mencionado, está dividida em duas partes. Na primeira parte, definimos novas classes de distribuições e os respectivos modelos de regressão para modelagem de dados no intervalo limitado (podendo a variável resposta assumir valores nos extremos do intervalo) e propomos medidas de diagnóstico e implementação computacional. Essa parte é apresentada do segundo ao quinto capítulo. A segunda parte é referente à inferência robusta no modelo de regressão beta inflacionado e é apresentada no sexto capítulo. As notações utilizadas nessas duas partes são feitas de forma independente.

Os capítulos desta tese são organizados como segue. Neste primeiro capítulo, uma breve revisão sobre modelagem de dados fracionários é apresentada. No segundo capítulo, propomos a classe de distribuições *power logit* e exploramos algumas de suas propriedades. Aspectos inferenciais baseando-se na teoria da máxima verossimilhança são discutidos, bem como medidas de qualidade de ajuste. Além disso, simulações Monte Carlo são conduzidas a fim de avaliar o desempenho dos métodos propostos. Ainda nesse capítulo, são apresentadas aplicações em conjuntos de dados reais. O modelo de regressão *power logit* é apresentado no Capítulo 3. Nesse capítulo também são discutidas medidas de diagnóstico e influência e inferência sobre os parâmetros do modelo. O modelo de regressão log-log, que é um caso limite do modelo de regressão *power logit*, também é desenvolvido. No Capítulo 4 apresentamos o novo pacote do R, chamado `PLreg`, que permite o ajuste dos modelos de regressão *power logit*, log-log e GJS. Nesse capítulo discutimos as principais funções disponíveis no pacote e diferentes aplicações em conjuntos de dados reais. No quinto capítulo apresentamos o modelo de regressão *power logit* inflacionado em zero ou um e desenvolvemos medidas de diagnóstico e influência. Aplicações em dados reais são feitas e uma implementação computacional é sugerida. No Capítulo 6 discutimos a influência de observações atípicas em inferência via máxima verossimilhança no modelo de regressão beta inflacionado em zero ou um e propomos métodos inferenciais robustos. Discutimos procedimentos robustos para testar hipóteses e apresentamos estudos de simulação que evidenciam o bom desempenho dos métodos propostos. Finalmente, no Capítulo 7 discutimos os resultados obtidos e apresentamos sugestões de pesquisas futuras.

Contribuições desta tese

As principais contribuições desta tese são:

1. desenvolvimento de uma nova classe de distribuições, denominada classe *power logit*, para dados de suporte limitado e respectivo modelo de regressão;
2. discussão da classe de distribuições log-log e do modelo de regressão log-log;
3. derivação das classes de distribuições *power logit* inflacionadas e log-log inflacionadas e dos respectivos modelos de regressão para a modelagem de dados de suporte limitado que apresentem valores em um dos extremos do intervalo;
4. desenvolvimento de métodos inferenciais, resíduos, ferramentas de diagnóstico e influência para todos os modelos propostos, bem como aplicações em conjuntos de dados reais;

5. elaboração de um novo pacote computacional no *software* R que possibilida o ajuste de todos os modelos desenvolvidos e de alguns modelos já existentes na literatura;
6. estudo da influência de observações atípicas na inferência via máxima verossimilhança no modelo de regressão beta inflacionado em zero ou um;
7. desenvolvimento de estimadores robustos para os parâmetros do modelo de regressão beta inflacionado em zero ou um;
8. adaptação de um algoritmo orientado pelos dados para seleção da constante de afinação dos estimadores robustos propostos;
9. ilustração da eficácia dos métodos robustos propostos em aplicações em dados simulados e em estudos de simulação.

Capítulo 2

A classe de distribuições *power logit*

Neste capítulo, introduzimos uma nova classe de distribuições para a modelagem de dados fracionários de natureza contínua: a classe *power logit* (PL). Essa nova classe é bastante flexível, tem parâmetros interpretáveis e acomoda bem observações próximas aos extremos do suporte. Este capítulo encontra-se organizado da seguinte forma. Na Seção 2.1, apresentamos alguns conceitos necessários para o desenvolvimento deste trabalho, assim como algumas distribuições utilizadas para a modelagem de dados de suporte limitado. A classe PL é definida na Seção 2.2 e algumas propriedades dessa classe são exploradas bem como a motivação para a definição desta nova classe de distribuições. Na Seção 2.3 são estudados aspectos inferenciais em modelos *power logit* através do método da máxima verossimilhança usual e penalizada. Nesta seção também é discutida a estimação de parâmetros extras e uma medida de qualidade de ajuste é proposta. Estudos de simulação e um breve estudo de robustez são apresentados na Seção 2.4. Finalmente, aplicações a dados reais são conduzidas na Seção 2.5.

2.1 Preliminares

Nesta seção, apresentamos alguns conceitos e/ou tópicos que serão úteis para o desenvolvimento deste trabalho.

2.1.1 A classe simétrica de distribuições

Definimos, a seguir, a classe simétrica univariada de distribuições.

Definição 2.1 (A classe simétrica.) *Seja a variável aleatória W com suporte em \mathbb{R} da classe simétrica, então sua função densidade de probabilidade é dada por*

$$h(w) = h(w; \mu, \sigma) = \frac{1}{\sigma} r \left(\frac{(w - \mu)^2}{\sigma^2} \right), \quad w \in \mathbb{R},$$

em que μ é o parâmetro de posição, $\sigma > 0$ é o parâmetro de escala e $r : \mathbb{R} \rightarrow [0, +\infty)$ é denominada função geradora de densidades, tal que $\int_0^\infty r(u) du < \infty$ e $\int_0^\infty u^{-1/2} r(u) du = 1$. Esta condição é necessária para que $h(w)$ seja uma função densidade de probabilidade. Usamos a notação $W \sim S(\mu, \sigma^2; r)$.

Essa família de distribuições, também conhecida como família de distribuições elípticas univariadas, é uma extensão da distribuição normal para modelar dados estatísticos envolvendo distribuições com caudas mais leves ou mais pesadas do que as da distribuição normal. Essa classe de distribuições aparece frequentemente na literatura na modelagem de dados que contêm mais *outliers* do que o esperado para a distribuição normal.

Algumas propriedades da distribuição normal podem ser estendidas para a classe simétrica. Por exemplo, se $W \sim S(\mu, \sigma^2; r)$ então $a + bW \sim S(a + b\mu, b^2\sigma^2; r)$, em que $a, b \in \mathbb{R}$ com $b \neq 0$. Isto é,

a distribuição de qualquer combinação linear de uma variável aleatória com distribuição simétrica é também simétrica. Da mesma forma, se $W \sim S(\mu, \sigma^2; r)$ então $Z = (W - \mu)/\sigma \sim S(0, 1; r)$, com função densidade $h(z) = h(z; 0, 1) = r(z^2)$, $z \in \mathbb{R}$, e chamaremos Z de uma variável aleatória com distribuição simétrica padrão. A função característica $\psi(t) = \mathbb{E}(e^{itW})$ é dada por $\psi(t) = e^{it\mu} \varphi(t^2 \sigma^2)$, $t \in \mathbb{R}$ e $i = \sqrt{-1}$, para alguma função $\varphi(\cdot)$, com $\varphi(v) \in \mathbb{R}$, para $v > 0$. Se existirem, temos $\mathbb{E}(W) = \mu$ e $\text{Var}(W) = \xi_r \sigma^2$, em que $\xi_r > 0$ é uma constante dada por $\xi_r = -2\varphi'(0)$, em que $\varphi'(0) = d\varphi(v)/dv|_{v=0}$.

Existem diferentes modelos que fazem parte dessa classe de distribuições: normal, t de Student, logístico, exponencial potência, hiperbólico, slash, entre muitos outros. Esses modelos serão estudados nas próximas seções e outras propriedades dessa classe de distribuições podem ser encontradas em [Fang et al. \(1990\)](#).

2.1.2 A modelagem de dados no intervalo unitário

No que segue, apresentamos brevemente algumas importantes distribuições para a modelagem de dados fracionários; mais detalhes podem ser encontrados em [Lima \(2018\)](#) e em [Smithson e Merkle \(2013\)](#).

Distribuição beta

A distribuição beta é uma das principais distribuições utilizadas na modelagem de dados de suporte limitado. Essa distribuição pode apresentar formas unimodais, simétricas, assimétricas, de U, J, J invertido, uniforme. Existem diferentes parametrizações para a distribuição beta na literatura, mas a proposta por [Ferrari e Cribari-Neto \(2004\)](#) tem ganhado destaque. Nessa parametrização, a função densidade de probabilidade de uma variável aleatória Y com distribuição beta é dada por

$$f(y; \xi, \phi) = \frac{\Gamma(\phi)}{\Gamma(\xi\phi)\Gamma((1-\xi)\phi)} y^{\xi\phi-1} (1-y)^{(1-\xi)\phi-1}, \quad (2.1)$$

com $y \in (0, 1)$ e $\xi \in (0, 1)$ e $\phi > 0$, e escrevemos $Y \sim \text{beta}(\xi, \phi)$. Na parametrização proposta por [Ferrari e Cribari-Neto \(2004\)](#),

$$\mathbb{E}(Y) = \xi \quad \text{e} \quad \text{Var}(Y) = \frac{\xi(1-\xi)}{1+\phi}.$$

Assim, os parâmetros ξ e ϕ representam a média e a precisão de Y , respectivamente.

Distribuição Kumaraswamy

A distribuição Kumaraswamy ([Kumaraswamy, 1980](#)) é similar à distribuição beta, mas tem a vantagem da função de distribuição acumulada apresentar forma analítica fechada. Neste trabalho, consideramos a parametrização proposta por [Mitnik e Baek \(2013\)](#) cuja função densidade de probabilidade é dada por

$$f(y; \mu, \sigma) = \frac{\log(0.5)}{\sigma \log(1 - \mu^{\sigma^{-1}})} y^{\sigma^{-1}-1} (1 - y^{\sigma^{-1}})^{\frac{\log(0.5)}{\log(1 - \mu^{\sigma^{-1}})} - 1},$$

com $y \in (0, 1)$ e $\mu \in (0, 1)$ e $\sigma > 0$ são a mediana e a dispersão de Y , respectivamente. Escrevemos $Y \sim \text{Kumaraswamy}(\mu, \sigma)$.

Distribuição L-Logística

A distribuição L-logística ([da Paz et al., 2019](#); [Tadikamalla e Johnson, 1982](#)), assim como a distribuição Kumaraswamy, também tem função de distribuição em forma fechada. A função den-

sidade de uma variável aleatória Y com distribuição L-Logística é dada por

$$f(y; \mu, \sigma) = \frac{\sigma(1-\mu)^\sigma \mu^\sigma y^{\sigma-1} (1-y)^{\sigma-1}}{[(1-\mu)^\sigma y^\sigma + \mu^\sigma (1-y)^\sigma]^2},$$

com $y \in (0, 1)$ e $\mu \in (0, 1)$ e $\sigma > 0$ são a mediana e a dispersão de Y , respectivamente. Escrevemos $Y \sim \text{LL}(\mu, \sigma)$.

Distribuição simplex

A distribuição simplex foi introduzida por [Barndorff-Nielsen e Jørgensen \(1991\)](#) e tem função densidade de probabilidade da forma

$$f(y; \mu, \sigma) = \left\{ \frac{1}{2\pi\sigma^2 \{y(1-y)\}^3} \right\}^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \frac{(y-\xi)^2}{y(1-y)\xi^2(1-\xi)^2} \right\},$$

em que $y \in (0, 1)$ e $\xi \in (0, 1)$ e $\sigma > 0$ são a média e a dispersão de Y , respectivamente. A função densidade da distribuição simplex pode apresentar forma de J, U, J invertido e bimodal. Escrevemos $Y \sim S^{-1}(\xi, \sigma)$.

Distribuição logito slash

A distribuição logito slash foi introduzida por [Korkmaz \(2020\)](#) e pode ser obtida através da transformação logito aplicada à distribuição slash ([Rogers e Tukey, 1972](#)). Se Y tem distribuição logito slash, sua função densidade de probabilidade é dada por

$$f(y; \xi, \sigma, q) = \frac{q}{\sigma y(1-y)} \int_0^1 t^q \phi \left(t \left[\frac{\log \left(\frac{y}{1-y} \right) - \xi}{\sigma} \right] \right) dt,$$

para $y \in (0, 1)$, $q > 0$, $\sigma > 0$ e $-\infty < \xi < \infty$. Os parâmetros ξ e σ podem ser interpretados como parâmetros de posição e escala, respectivamente. A distribuição slash tem caudas mais pesadas do que a distribuição normal.

Distribuições geradas por transformações

Além das distribuições mencionadas, diferentes métodos para gerar distribuições no intervalo unitário através de transformações têm sido explorados na literatura. Por exemplo, a transformação logarítmica do tipo $Y = e^{-X}$, em que X é uma variável aleatória definida no intervalo $(0, \infty)$; transformações do tipo $Y = X/(1+X)$, em que X é, também, uma variável aleatória definida no intervalo $(0, \infty)$. Um procedimento para construir distribuições no intervalo unitário é gerar uma função de distribuição acumulada, digamos $G(x)$, de uma variável aleatória X da seguinte forma

$$G(x) = \int_0^{S(x)} r(t) dt,$$

sendo $r(t)$ a função densidade de probabilidade de uma variável aleatória T com suporte $(0, 1)$ e $S(x)$ a função de distribuição acumulada de uma variável aleatória qualquer. Existem diferentes especificações de variáveis aleatórias definidas através dessa transformação. Por exemplo, [Eugene et al. \(2002\)](#) consideram $r(t)$ a função densidade de uma distribuição beta e definem a distribuição beta-normal; [Jones \(2009\)](#) e [Cordeiro e de Castro \(2011\)](#) consideram a função densidade de uma distribuição Kumaraswamy para $r(t)$, originando a classe de distribuições Kumaraswamy generalizada; entre outras.

Outras extensões para essa família foram propostas, por exemplo, a família T-X ([Alzaatreh et al.](#),

2013), definida por

$$G(x) = \int_0^{W(S(x))} r(t) dt,$$

em que $r(t)$ é a função densidade de probabilidade de uma variável aleatória T com suporte em $[a, b]$, para $-\infty \leq a < b \leq \infty$. A função $W(S(x))$ deve satisfazer

1. $W(S(x)) \in [a, b]$;
2. $W(S(x))$ é diferenciável e monotonicamente não decrescente;
3. $W(S(x)) \rightarrow a$ quando $x \rightarrow -\infty$ e $W(S(x)) \rightarrow b$ quando $x \rightarrow \infty$.

Note que as condições 2. e 3. são necessárias para que $G(x)$ seja uma função de distribuição legítima. A função $G(x)$ pode, também, ser escrita como

$$G(x) = R[W(S(x))],$$

sendo $R(\cdot)$ a função de distribuição acumulada da variável aleatória T . Note que esse método pode gerar distribuições com suporte limitado e ilimitado. Os parâmetros da distribuição $G(x)$ são derivados dos parâmetros das distribuições $R(\cdot)$ e $S(\cdot)$. Uma outra extensão da família T-X é feita em Aljarrah *et al.* (2014). Os autores propõem que W seja a função quantílica de uma terceira variável aleatória, digamos Y , que tem o mesmo suporte da variável T . Essa classe é chamada de família T-X{Y}. Para a escolha de W , seja (a, b) o suporte de T e seja P a função de distribuição da variável aleatória Y , também com suporte (a, b) . A função quantílica de P é definida por

$$Q_Y(p) = \inf\{y : P(y) \geq p\},$$

para $p \in (0, 1)$. Dessa forma, a função de distribuição de uma variável aleatória pertencente à família T-X{Y} é dada por

$$G(x) = \int_0^{Q_Y(S(x))} r(t) dt = R[Q_Y(S(x))].$$

Smithson e Merkle (2013) descrevem uma classe de distribuições, que é um caso especial da família T-X{Y}, mas que tem ganhado bastante destaque. Essa classe também é caracterizada em termos de sua função de distribuição. Seja $G(x; \mu, \sigma)$ a função de distribuição de X com suporte $(0, 1)$, parâmetro de localização $\mu \in \mathbb{R}$ e de escala $\sigma > 0$. A função $G(x; \mu, \sigma)$ é definida por

$$G(x; \mu, \sigma) = F[U(H^{-1}(x); \mu, \sigma)],$$

em que F é a função de distribuição com suporte D_1 , H é uma função de distribuição invertível com suporte D_2 e $U : D_2 \rightarrow D_1$ é uma transformação apropriada que introduz os parâmetros μ e σ . Os domínios D_1 e D_2 podem ser $(-\infty, \infty)$ e/ou $(0, \infty)$ e a escolha da função U pode variar. Por exemplo, se $D_1 = D_2 = (-\infty, \infty)$, então uma escolha para U é

$$U(x; \mu, \sigma) = \frac{x - \mu}{\sigma}.$$

Assumindo que os parâmetros das distribuições F e H são conhecidos, essa classe biparamétrica é chamada de *CDF-quantile* (Smithson e Shou, 2017). Além disso, se $F = R$ e $U(H^{-1}(x); \mu, \sigma) = W(S(x))$, com H diferenciável e $x \in (0, 1)$, essa classe se reduz à família T-X. Uma outra subfamília desta classe é a apresentada em Lemonte e Bazán (2016), denominada classe GJS. Essa classe foi definida a partir da generalização da distribuição Johnson S_B (Johnson, 1949) e sua construção é apresentada a seguir.

A distribuição GJS

A distribuição GJS (Lemonte e Bazán, 2016) é uma generalização da distribuição Johnson S_B (Johnson, 1949). A distribuição Johnson S_B (ou logito normal) é construída a partir da transformação logito em uma variável aleatória com distribuição normal. Isto é, seja X uma variável aleatória normal padrão, então

$$Y = t^{-1} \left(\frac{X - \gamma}{\delta} \right)$$

tem distribuição Johnson S_B , em que $t(w) = \log(w/(1-w))$, e $\gamma \in \mathbb{R}$ e $\delta > 0$ são os parâmetros.

A distribuição GJS (*generalized Johnson S_B*) assume que X tem distribuição na classe simétrica, isto é, $X \sim S(0, 1; r)$. A motivação principal para essa suposição é baseada na busca por distribuições com suporte limitado que apresentem maior (ou menor) curtose do que a distribuição Johnson S_B , de modo a obter uma distribuição Johnson S_B que é mais platicúrtica (ou leptocúrtica), além de outras propriedades.

Uma parametrização alternativa, apresentada em Lemonte e Bazán (2016), é fornecida. A construção dessa parametrização segue do fato de que a mediana da distribuição GJS é dada por $\mu = (1 + e^{\gamma/\delta})^{-1}$. Dessa forma, pode-se expressar

$$\gamma = \delta \log \left(\frac{1 - \mu}{\mu} \right) = -\delta t(\mu).$$

Neste trabalho, vamos considerar a parametrização proposta por Lemonte e Bazán (2016) com $\sigma = \delta^{-1}$. Assim, a função densidade de probabilidade de Y é dada por

$$f(y; \mu, \sigma) = \frac{r(\{\sigma^{-1}[t(y) - t(\mu)]\}^2)}{\sigma y(1-y)}, \quad y \in (0, 1), \quad (2.2)$$

em que $0 < \mu < 1$ é a mediana e $\sigma > 0$ é a dispersão de Y .

A distribuição GJS é um caso particular da *CDF-quantile*. A distribuição GJS é obtida quando H é a função de distribuição de uma variável aleatória com distribuição logística padrão e F é a função de distribuição de uma variável aleatória W pertencente à classe simétrica.

2.2 Modelos *power logit*

Definimos, a seguir, a construção de uma nova classe de distribuições univariadas com suporte no intervalo unitário.

Definição 2.2 (A classe *power logit*.) *Seja Y uma variável aleatória contínua com suporte $(0, 1)$ e defina $Y^{(\lambda)} = Y^\lambda/(1 - Y^\lambda)$. A distribuição de Y é definida a partir da transformação*

$$Z \equiv h(Y; \mu, \sigma, \lambda) = \frac{1}{\sigma} \log \left(\frac{Y^{(\lambda)}}{\mu^{(\lambda)}} \right), \quad (2.3)$$

em que $\mu^{(\lambda)} = \mu^\lambda/(1 - \mu^\lambda)$, com $0 < \mu < 1$, $\sigma > 0$ e $\lambda > 0$. Esta nova classe de distribuições, denominada aqui de classe de distribuições *power logit*, é definida ao atribuir a Z uma distribuição simétrica padrão.

Seja Y como definido acima e tome $Z \equiv h(Y; \mu, \sigma, \lambda)$ como em (2.3). Se Z tem distribuição simétrica padrão, dizemos que Y tem distribuição *power logit* (PL) com parâmetros $0 < \mu < 1$, $\sigma > 0$ e $\lambda > 0$, e função geradora de densidades r ; escrevemos $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$. Em outras palavras, $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$ se a variável transformada Z em (2.3) tem distribuição $S \sim S(0, 1; r)$. Observe que o suporte de Y é o intervalo unitário $(0, 1)$. Adicionalmente, o gerador de densidades $r(\cdot)$ pode depender de um parâmetro extra (ou de um vetor de parâmetros), que será denotado aqui por ζ .

A motivação para a classe *power logit* é dada a seguir. A transformação logito mapeia o intervalo $(0, 1)$ em $(-\infty, \infty)$ e, portanto, é uma candidata para definir distribuições com suporte no intervalo unitário a partir de uma distribuição com suporte na reta. No lugar de uma distribuição particular com suporte em $(-\infty, \infty)$, consideramos distribuições na classe simétrica. Além disso, introduzimos um parâmetro de forma λ na transformação logito que possibilita maior flexibilidade para acomodar uma gama mais variada de formas assimétricas para dados fracionários. De fato, a função *power logit*, definida por $\log[y^\lambda/(1-y^\lambda)]$, para $\lambda > 0$, é bem sucedida em atingir simetria em situações em que a transformação logito ($\lambda = 1$) falha. A fim de ilustrar, a Figura 2.1 apresenta o histograma e o boxplot de uma amostra de 1000 observações com distribuição assimétrica no intervalo $(0, 1)$ e os histogramas e boxplots das observações transformadas pelas funções logito e *power logit* com $\lambda = 0.11$.

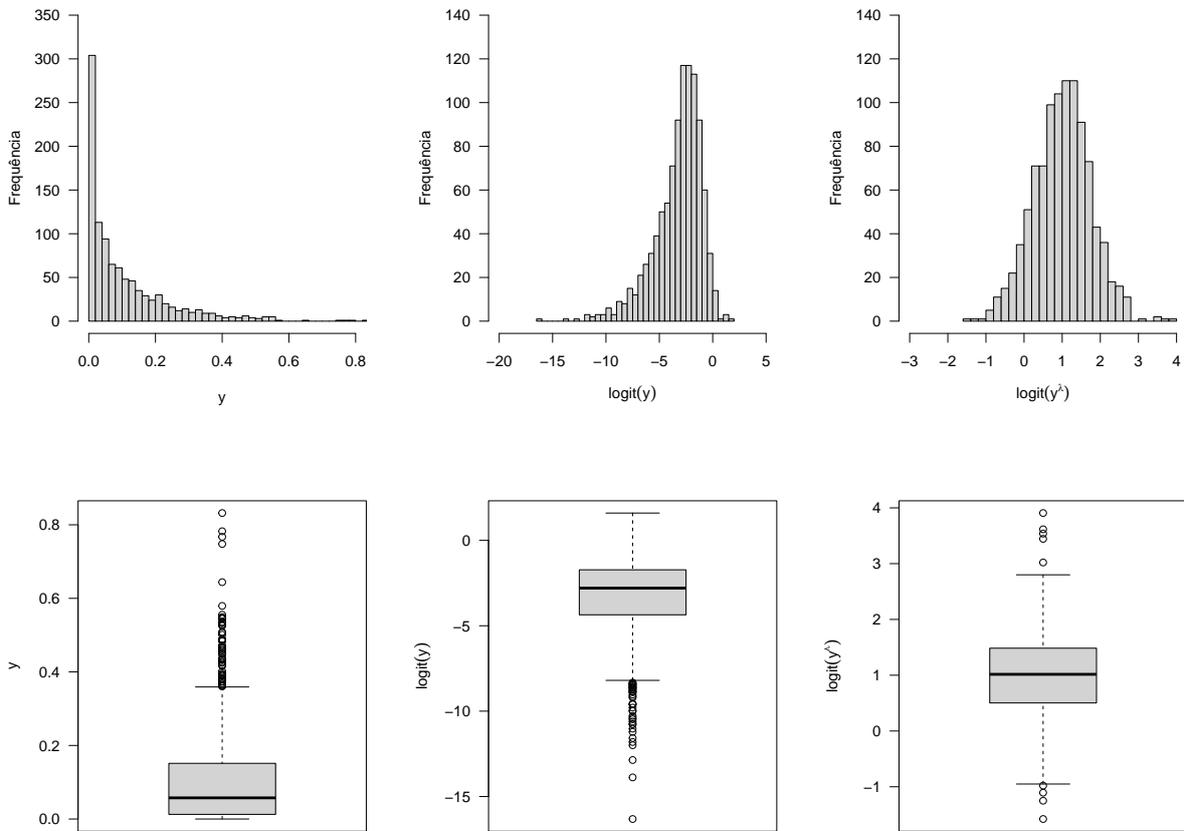


Figura 2.1: Histogramas (primeira linha) e boxplots (segunda linha) de y (dados originais), $\text{logit}(y)$ e $\text{logit}(y^\lambda)$; $\lambda = 0.11$.

Conforme observamos, a distribuição dos dados transformados através da função logito é assimétrica, enquanto a dos dados transformados pela transformação *power logit* é aproximadamente simétrica. De fato, o coeficiente de assimetria amostral dos dados originais é 1.92, enquanto que os dos dados transformados são -1.30 (função logito) e 0.00 (função *power logit*). O potencial da função *power logit* para transformar distribuições assimétricas em simétricas é ilustrado no Apêndice A. Finalmente, a Definição 2.2 implica que $\log[Y^\lambda/(1-Y^\lambda)]$ tem distribuição simétrica com média $\log[\mu^\lambda/(1-\mu^\lambda)]$ e parâmetro de escala σ . Essa parametrização é conveniente pois μ representa a mediana de Y , enquanto σ e λ são parâmetros de dispersão e de assimetria, respectivamente, conforme será visto.

A classe de distribuições *power logit* generaliza algumas distribuições existentes na literatura,

a saber: a distribuição GJS (Lemonte e Bazán, 2016) é obtida tomando $\lambda = 1$; se Z em (2.3) tem distribuição logística padrão e $\lambda = 1$ temos a distribuição L-Logística (da Paz *et al.*, 2019; Tadikamalla e Johnson, 1982); a distribuição logito slash (Korkmaz, 2020) é obtida quando Z tem distribuição slash padrão e fazemos $\lambda = 1$. As distribuições *power logit* e *CDF-quantile* são equivalentes apenas quando $\lambda = 1$ na classe *power logit* e quando H , na distribuição *CDF-quantile*, é a função de distribuição de uma variável aleatória logística padrão e W pertence à classe simétrica. Embora a classe *CDF-quantile* seja bastante ampla, os desenvolvimentos teóricos (e também implementados) apresentados em Smithson e Shou (2017) são para modelos mais específicos. Além disso, os parâmetros dessa classe não são facilmente interpretáveis em generalidade (por exemplo, μ não representa a média ou a mediana de Y).

Algumas propriedades da classe de distribuições *power logit*

Sejam $f_S(\cdot)$ e $F_S(\cdot)$ a função densidade de probabilidade e a função de distribuição de $S \sim S(0, 1; r)$. Considere $Y \sim PL(\mu, \sigma, \lambda; r)$ e seja $z = h(y; \mu, \sigma, \lambda)$, com $h(\cdot)$ dado em (2.3). O Jacobiano da transformação de y para z é $|\partial z / \partial y| = \lambda / [\sigma y(1 - y^\lambda)]$ e, então, a função densidade de probabilidade de Y é dada por

$$f_Y(y; \mu, \sigma, \lambda) = \frac{\lambda}{\sigma y(1 - y^\lambda)} f_S(z), \quad y \in (0, 1).$$

Sendo $f_S(s) = r(s^2)$, podemos reescrever $f_Y(y)$ como

$$f_Y(y; \mu, \sigma, \lambda) = \frac{\lambda}{\sigma y(1 - y^\lambda)} r(z^2), \quad y \in (0, 1). \quad (2.4)$$

A função distribuição acumulada de Y é dada por

$$F_Y(y; \mu, \sigma, \lambda) = \int_{-\infty}^{\frac{1}{\sigma} \log(y^\lambda / \mu^\lambda)} r(u^2) du, \quad y \in (0, 1). \quad (2.5)$$

Note que $F_Y(y; \mu, \sigma, \lambda) = R(z)$, sendo $R(s) = \int_{-\infty}^s r(u^2) du$, para $s \in \mathbb{R}$.

A depender da distribuição simétrica assumida para a variável aleatória Z e, conseqüentemente, do gerador de densidades $r(\cdot)$, temos diferentes distribuições na classe *power logit*. A seguir, apresentamos algumas distribuições na classe *power logit*, seus respectivos geradores de densidade e gráficos da função densidade de probabilidade (Figuras de 2.2 a 2.13).

1. A distribuição *power logit* normal, $PL-N(\mu, \sigma, \lambda)$, é obtida quando $Z \sim N(0, 1)$, com gerador de densidades

$$r(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}, \quad z \geq 0.$$

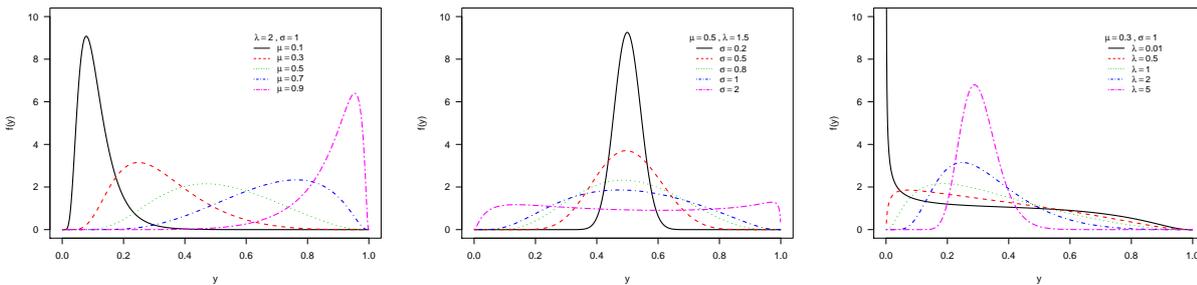


Figura 2.2: Densidades PL-N.

2. Distribuição *power logit t*, $PL-t_{(\zeta)}(\mu, \sigma, \lambda)$, é obtida quando a distribuição de Z é t-Student padrão com ζ graus de liberdade. A distribuição $PL-t_{(\zeta)}$ tem um parâmetro extra (ζ), que representa o parâmetros de graus de liberdade da variável Z . Além disso, ela tem caudas mais pesadas do que a distribuição normal. O gerador de densidades é

$$r(z) = \frac{\zeta^{\zeta/2}}{B(1/2, \zeta/2)} (\zeta + z)^{-(\zeta+1)/2}, \quad z \geq 0,$$

com $\zeta > 0$ e $B(\cdot, \cdot)$ denotando a função beta. Se $\zeta = 1$, obtemos a distribuição *power logit Cauchy*, isto é, PL-Cauchy.

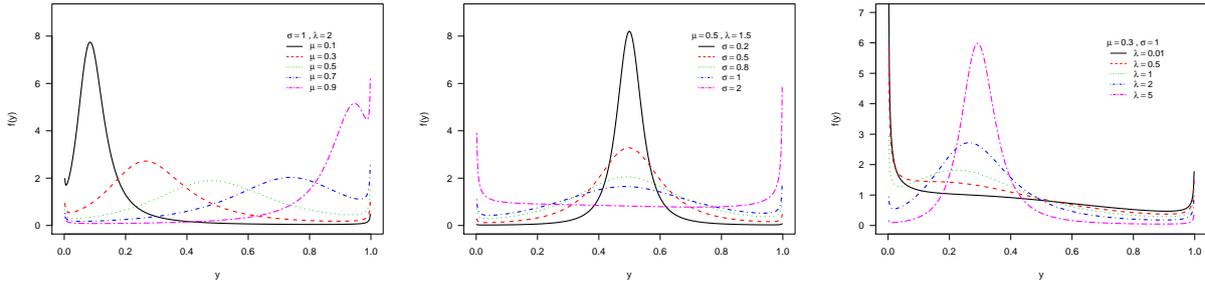


Figura 2.3: Densidades $PL-t_{(2)}$.

A Figura 2.4 apresenta gráficos da distribuição $PL-t_{(\zeta)}(0.8, 0.5, 2)$, para diferentes valores de $\zeta > 0$.

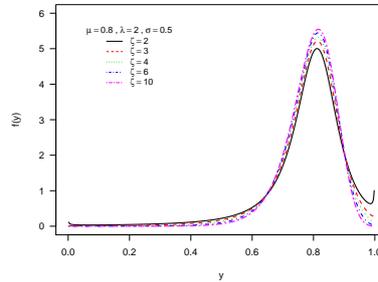


Figura 2.4: Densidades $PL-t_{(\zeta)}$ para diferentes valores de ζ .

3. A distribuição *power logit exponencial potência*, $PL-PE_{(\zeta)}(\mu, \sigma, \lambda)$, é obtida quando Z tem distribuição exponencial potência padrão (Nelson, 1991) com parâmetro $\zeta > 0$. Assim como a distribuição $PL-t_{(\zeta)}$, a distribuição $PL-PE_{(\zeta)}$ depende de um parâmetro extra ζ . Quando $\zeta = 1$ e $\zeta = 2$, temos as distribuições *power logit exponencial dupla* ou Laplace (PL-LA) e PL-N, respectivamente. O gerador de densidades é dado por

$$r(z) = \frac{\zeta}{p(\zeta)2^{1+1/\zeta}\Gamma(1/\zeta)} \exp\left\{-\frac{z^{\zeta/2}}{2p(\zeta)\zeta}\right\}, \quad z \geq 0,$$

em que $p(\zeta)^2 = 2^{-2/\zeta}\Gamma(1/\zeta)/\Gamma(3/\zeta)$. A distribuição exponencial potência pode ter caudas mais pesadas (quando $\zeta < 2$) ou mais leves (quando $\zeta > 2$) do que a distribuição normal. Além disso, a distribuição uniforme no intervalo $(0, 1)$ é um caso limite quando $\zeta \rightarrow \infty$. A distribuição exponencial potência tem outra parametrização bastante conhecida na literatura; ver, por exemplo, Box e Tiao (2011). Assuma que Z tem distribuição exponencial potência com gerador de densidades $r(z)$ dado acima e $\zeta = 2/(1+k)$, então a parametrização apresentada em Box e Tiao (2011) é dada pela distribuição de $W = Z/p(2/(1+k))$. Dessa forma,

teremos uma nova parametrização e $k \in (-1, 1]$. Utilizamos a parametrização apresentada em Nelson (1991) pois ela já está implementada no pacote `gamlss`.

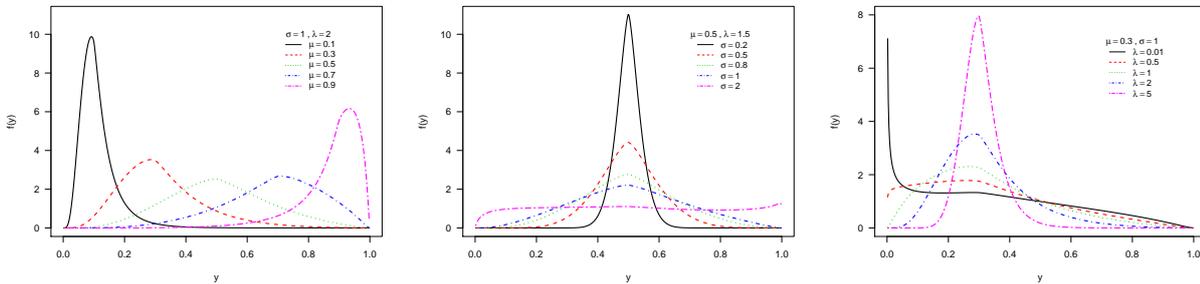


Figura 2.5: Densidades $PL-PE_{(1.5)}$.

A Figura 2.6 apresenta gráficos da distribuição $PL-PE_{(\zeta)}(0.7, 0.5, 2)$, para diferentes valores de $\zeta > 0$.

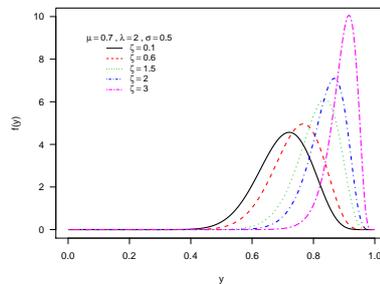


Figura 2.6: Densidades $PL-PE_{(\zeta)}$ para diferentes valores de ζ .

4. A distribuição *power logit* logística, $PL-LO(\mu, \sigma, \lambda)$, é obtida quando Z tem distribuição logística do tipo 1 padrão. O gerador de densidades é da forma

$$r(z) = c \exp\{-z\}(1 + \exp\{-z\})^2, \quad z \geq 0,$$

em que $c \approx 1.484300029$ é a constante de normalização, obtida de modo que $\int_0^\infty z^{1/2}r(z)dz = 1$.

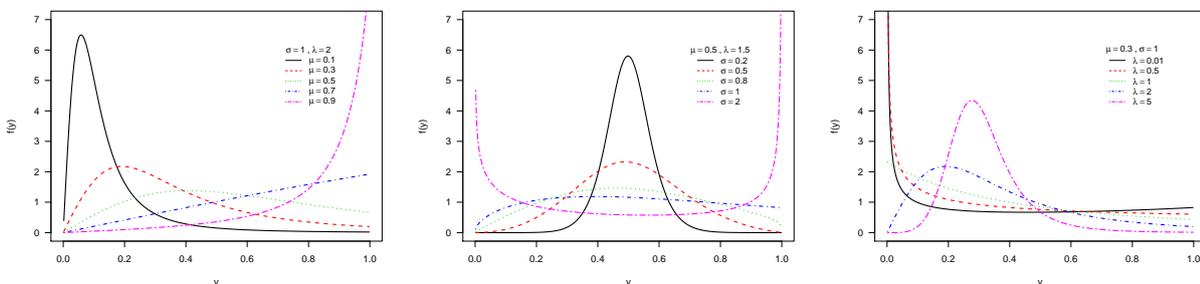


Figura 2.7: Densidades $PL-LO_1$.

5. A distribuição *power logit slash*, $PL-slash_{(\zeta)}(\mu, \sigma, \lambda)$, é obtida quando Z tem distribuição slash padrão com parâmetro de forma $\zeta > 0$. Uma variável aleatória Z com distribuição $slash_{(\zeta)}$ pode ser obtida através da relação $Y = V^{-1/2}Z$, em que V é uma variável aleatória

com densidade $f(v) = \zeta v^{\zeta-1}$, para $0 \leq v \leq 1$ e Z é uma variável aleatória com distribuição normal padrão, com V e Z independentes. O gerador de densidades é dado por

$$r(z) = \begin{cases} \frac{\zeta}{\sqrt{2\pi}} \left(\frac{z}{2}\right)^{-(\zeta+1/2)} G\left(\zeta + \frac{1}{2}, \frac{z}{2}\right), & \text{se } z > 0 \\ \frac{2\zeta}{(2\zeta + 1)\sqrt{2\pi}}, & \text{se } z = 0, \end{cases}$$

em que $G(a, x) = \int_0^x t^{a-1} e^{-t} dt$ é a função gamma incompleta (*lower*). Se $\zeta = 2$, obtemos a distribuição Slash canônica (Rogers e Tukey, 1972) e a distribuição normal é obtida como caso limite quando $\zeta \rightarrow \infty$. Essa distribuição pode ter caudas mais pesadas do que a distribuição normal.

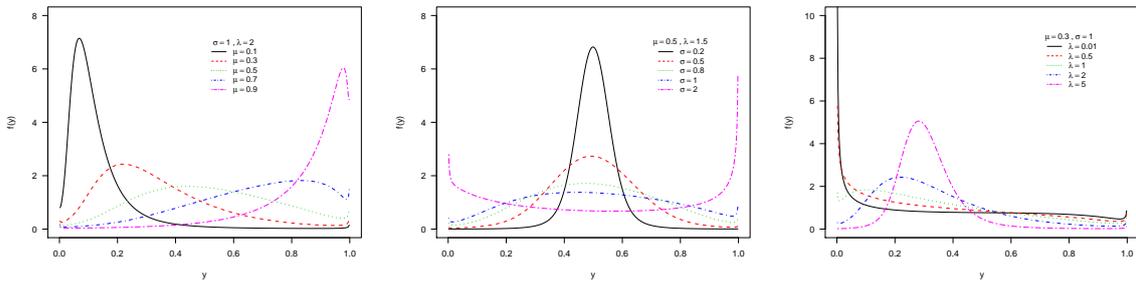


Figura 2.8: Densidades $PL\text{-slash}_{(1.4)}$.

A Figura 2.9 apresenta gráficos da distribuição $PL\text{-slash}_{(\zeta)}(0.7, 0.5, 2)$, para diferentes valores de $\zeta > 0$.

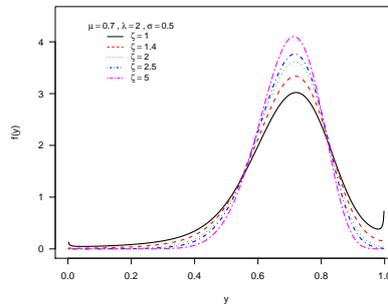


Figura 2.9: Densidades $PL\text{-slash}_{(\zeta)}$ para diferentes valores de ζ .

6. A distribuição *power logit* seno hiperbólico normal, $PL\text{-SN}_{(\zeta)}(\mu, \sigma, \lambda)$, é obtida quando Z tem distribuição seno hiperbólico normal padrão (Rieck e Nedelman, 1991). Essa distribuição depende de um parâmetro extra, $\zeta > 0$, que é um parâmetro de forma. Se $\zeta \leq 2$, a distribuição seno hiperbólico normal é unimodal, e é bimodal para $\zeta > 2$. Além disso, a distribuição normal é um caso limite quando $\zeta \rightarrow 0$. O gerador de densidades é dado por

$$r(z) = \frac{1}{\zeta\sqrt{2\pi}} \cosh(z^{1/2}) \exp\left[-\frac{2}{\zeta^2} \sinh^2(z^{1/2})\right], \quad z \geq 0,$$

em que $\cosh(\cdot)$ e $\sinh(\cdot)$ são as funções cosseno e seno hiperbólico, respectivamente.

A Figura 2.11 apresenta gráficos da distribuição $PL\text{-PE}_{(\zeta)}(0.7, 0.5, 2)$, para diferentes valores de $\zeta > 0$.

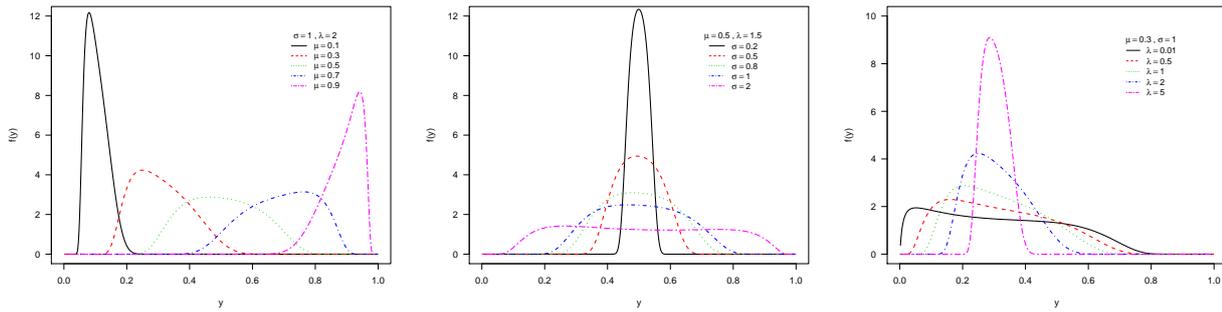


Figura 2.10: Densidades $PL-SN_{(1.5)}$.

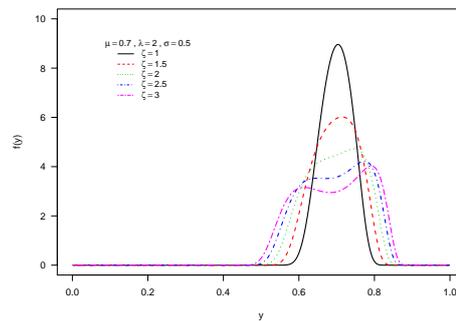


Figura 2.11: Densidades $PL-SN_{(\zeta)}$ para diferentes valores de ζ .

7. A distribuição *power logit* hiperbólica, denotada por $PL-Hyp_{(\zeta)}(\mu, \sigma, \lambda)$. A distribuição hiperbólica foi proposta por [Barndorff-Nielsen \(1977\)](#), com o objetivo de modelar o movimento de massas de areia impulsionadas por ventos com velocidade contínua. O gráfico do logaritmo da função densidade de probabilidade de uma variável aleatória com distribuição hiperbólica descreve uma hipérbole. Essa distribuição depende de um parâmetro extra, $\zeta > 0$, que é um parâmetro de forma e está associado com o peso das caudas dessa distribuição. Valores pequenos de ζ resultam em distribuições com caudas mais pesadas do que a normal. A distribuição normal é um caso limite da distribuição hiperbólica quando $\zeta \rightarrow \infty$. O gerador de densidades é dado por

$$r(z) = \frac{1}{2\zeta K_1(\zeta)} \exp\{-\zeta\sqrt{1+z}\},$$

em que $K_r(\cdot) = \frac{1}{2} \int_0^\infty x^{r-1} \exp\{-\frac{\zeta}{2}(x + \frac{1}{x})\} \partial x$ é a função Bessel modificada de terceira ordem e índice r . Se Z tem distribuição hiperbólica padrão com parâmetro de forma ζ , então podemos escrever $X = \sqrt{V}Z$, em que V tem distribuição inversa Gaussiana generalizada com parâmetros $(1, 1, \zeta^2)$ ([Jørgensen, 2012](#)), $X \sim N(0, 1)$ e X e V são variáveis aleatórias independentes.

A Figura 2.13 apresenta gráficos da distribuição $PL-Hyp_{(\zeta)}(0.7, 0.5, 2)$, para diferentes valores de $\zeta > 0$.

Com base nos gráficos apresentados, percebe-se que a classe de distribuições *power logit* é bastante flexível e pode ser uma alternativa interessante, quando comparada a outras distribuições que têm o suporte limitado, para a análise de dados univariados. Evidentemente, outras formas podem ser obtidas com a mudança dos valores dos parâmetros. Além disso, como esperado, os gráficos sugerem que o parâmetro λ controla a assimetria, enquanto o comportamento das caudas é controlado, no caso da distribuição $PL-t_{(\zeta)}$, por exemplo, pelo parâmetro extra.

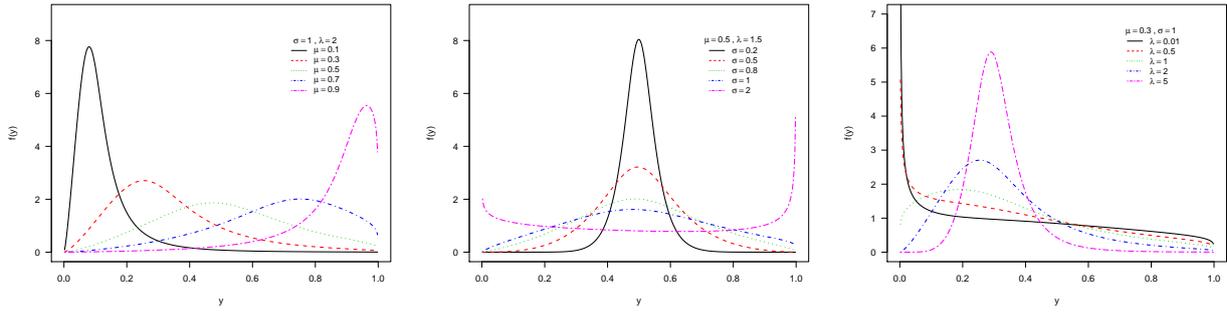


Figura 2.12: Densidades $PL-Hyp_{(1,2)}$.

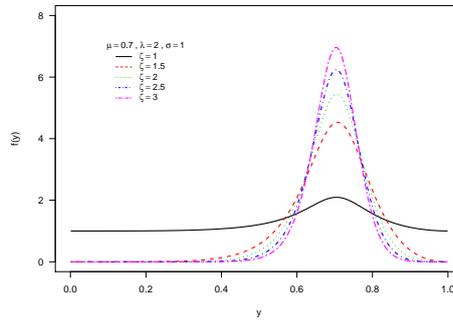


Figura 2.13: Densidades $PL-Hyp_{(\zeta)}$ para diferentes valores de ζ .

Além das distribuições *power logit* apresentadas aqui, outros modelos nessa classe podem ser obtidos quando assumimos para Z outras distribuições na classe simétrica. Entre elas, podemos considerar, para Z , a distribuição t-Student generalizada (Dickey, 1967), obtendo a distribuição $PL-tG_{(\zeta_1, \zeta_2)}$ para Y ; a distribuição logística generalizada (Johnson *et al.*, 1995), obtendo a distribuição $PL-LG_{(\zeta)}$ para Y ; a distribuição Kotz (Kotz, 1975), resultando em Y com distribuição $PL-K_{(\zeta_1, \zeta_2)}$; a distribuição normal contaminada (Berkane e Bentler, 1986), obtendo a distribuição $PL-NC_{(\zeta_1, \zeta_2)}$ para Y ; entre muitas outras.

A classe de distribuições *power logit* apresenta propriedades bastante interessantes. Algumas dessas propriedades são enunciadas através das seguintes proposições. As demonstrações das proposições podem ser encontradas no Apêndice B, Seção B.1.

Proposição 2.1 *Seja $Y \sim PL(\mu, \sigma, \lambda; r)$. O quantil de ordem u da distribuição de Y é dado por*

$$y_u = \left[\frac{\mu^\lambda e^{\sigma z_u}}{1 - \mu^\lambda (1 - e^{\sigma z_u})} \right]^{1/\lambda}$$

e z_u é o quantil de ordem u de $S \sim S(0, 1; r)$. Consequentemente, $y_{0.5} = \mu$, isto é, μ é a mediana da distribuição de Y .

Proposição 2.2 *Seja $Y \sim PL(\mu, \sigma, \lambda; r)$, então σ é parâmetro de dispersão.*

Proposição 2.3 *Se $Y \sim PL(\mu, \sigma, \lambda = 1; r)$, então $1 - Y \sim PL(1 - \mu, \sigma, \lambda = 1; r)$.*

Proposição 2.4 *Relações entre as classes power logit e GJS.*

1. *Se $Y \sim PL(\mu, \sigma, \lambda = 1; r)$, então a classe de distribuições PL reduz-se à classe de distribuições GJS, ou seja, $Y \sim GJS(\mu, \sigma; r)$, o que equivale a*

$$\log \left(\frac{Y}{1 - Y} \right) \sim S(\log(\mu/(1 - \mu)), \sigma^2; r).$$

2. $Y \sim PL(\mu, \sigma, \lambda; r)$ se, e somente se, $Y^\lambda \sim GJS(\mu^\lambda, \sigma; r)$.
3. $Y \sim PL(\mu = 0.5, \sigma, \lambda = 1; r)$ então Y é simétrica em torno de $\mu = 0.5$.

Proposição 2.5 *Seja $Y \sim PL(\mu, \sigma, \lambda; r)$, então*

$$-\log(-\log Y) \xrightarrow{\mathcal{D}} W, \quad \text{quando } \lambda \rightarrow 0^+$$

em que $\xrightarrow{\mathcal{D}}$ denota convergência em distribuição e $W \sim S(-\log(-\log \mu), \sigma^2; r)$.

Ou seja, o caso limite das distribuições *power logit* quando λ tende a zero equivale a uma distribuição simétrica de parâmetro de localização $-\log(-\log \mu)$ e dispersão σ^2 para a variável transformada pela transformação log-log. Dessa forma, as distribuições *power logit* têm como caso limite, quando $\lambda \rightarrow 0^+$, a (nomeada) família de distribuições log-log, conforme definida a seguir.

Definição 2.3 (A classe log-log.) *Dizemos que uma variável aleatória Y com suporte no intervalo $(0, 1)$ tem distribuição log-log se*

$$\frac{1}{\sigma} \{-\log(-\log Y) - [-\log(-\log \mu)]\} = Z,$$

em que $Z \sim S(0, 1; r)$, $\mu \in (0, 1)$ e $\sigma > 0$. Escrevemos $Y \sim \text{log-log}(\mu, \sigma; r)$. A função densidade de probabilidade de Y é dada por

$$f(y; \mu, \sigma) = \frac{1}{\sigma y(-\log y)} r(z^2), \quad y \in (0, 1). \quad (2.6)$$

A classe de distribuições log-log é uma generalização da distribuição Johnson S_B (Johnson, 1949), em que assumimos que $t(\cdot)$ é a transformação log-log, isto é, $t(w) = -\log(-\log w)$. Pode-se verificar que μ e σ são a mediana e a dispersão de Y , respectivamente.

Proposição 2.6 *Se $Y \sim PL(\mu, \sigma, \lambda; r)$, então $Y^c \sim PL(\mu^c, \sigma, \lambda/c; r)$ para $c > 0$.*

Adicionalmente, sendo $Y \sim PL(\mu, \sigma, \lambda; r)$ e fazendo uma transformação do tipo $W = (b-a)Y + a$, obtém-se uma distribuição PL com suporte em (a, b) , $a < b$.

O grau de assimetria das distribuições *power logit* pode ser quantificado através de algumas medidas de assimetria. Neste trabalho, consideramos a medida proposta por Hinkley (1975), denominada *quantile skewness* e denotada aqui por γ_p . Esta medida de assimetria é dada por

$$\gamma_p = \frac{(y_{1-p} - \mu) - (\mu - y_p)}{y_{1-p} - y_p},$$

em que $p \in (0, 0.5)$. Observe que o valor máximo dessa medida é 1, representando forte assimetria à direita; enquanto o valor mínimo é -1 , representando forte assimetria à esquerda. Se a distribuição é simétrica, então $\gamma_p = 0$. Além disso, se $y_{1-p} - \mu > \mu - y_p$, a distribuição é assimétrica à direita (assimetria positiva); por outro lado, se $y_{1-p} - \mu < \mu - y_p$, a distribuição é assimétrica à esquerda (assimetria negativa). Essa medida depende do valor de p e, seguindo as ideias de da Paz et al. (2019), consideramos $p = 0.125$ e, assim, temos

$$\gamma_{0.125} = \frac{(y_{0.875} - \mu) - (\mu - y_{0.125})}{y_{0.875} - y_{0.125}}.$$

As Figuras 2.14, 2.15 e 2.16 apresentam gráficos de $\gamma_{0.125}$ para os modelos PL-N e PL- $t_{(\zeta)}$, considerando diferentes valores para os parâmetros. Nas Figuras 2.14, 2.15 consideramos diferentes valores de σ , a mediana próxima de zero ($\mu = 0.1$), igual a meio ($\mu = 0.5$) e próxima de um ($\mu = 0.9$), variamos o λ . Note que, em todos os casos, quando λ cresce, a assimetria tende para zero; isto é,

a distribuição tende à simetria. Além disso, considerando σ fixo, a medida de assimetria decresce quando λ cresce, para $\mu < 0.5$. Por outro lado, para $\mu > 0.5$, a medida de assimetria cresce quando λ cresce. Quando $\mu = 0.5$, independente do valor de σ , a distribuição é simétrica quando $\lambda = 1$ (Proposição 2.4), tem fraca assimetria negativa quando $\lambda < 1$ (distribuição assimétrica à esquerda) e positiva quando $\lambda > 1$ (distribuição assimétrica à direita). Note que, embora as distribuições *power logit* não sejam simétricas para todo valor de λ quando $\mu = 0.5$, elas são aproximadamente simétricas ($\gamma_{0.125} \approx 0$). Adicionalmente, observe que o aumento de σ , mantendo λ e μ fixos, faz a assimetria aumentar (se $\mu < 0.5$) ou diminuir (se $\mu > 0.5$)

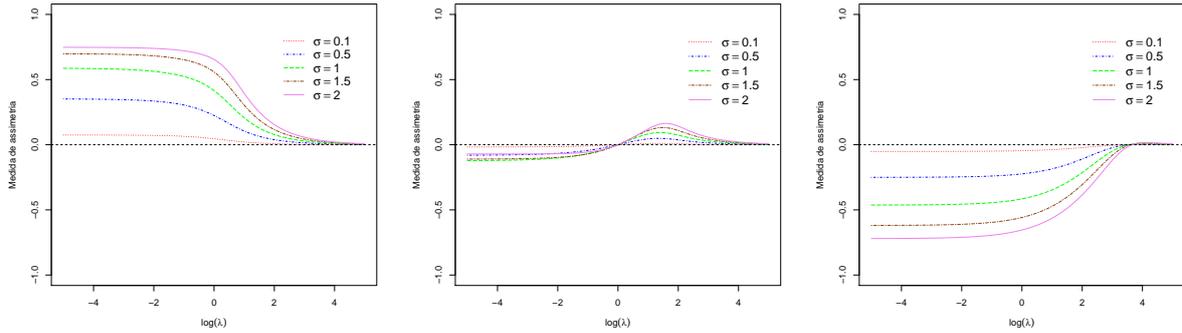


Figura 2.14: Medida de assimetria para a distribuição *PL-N* considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .

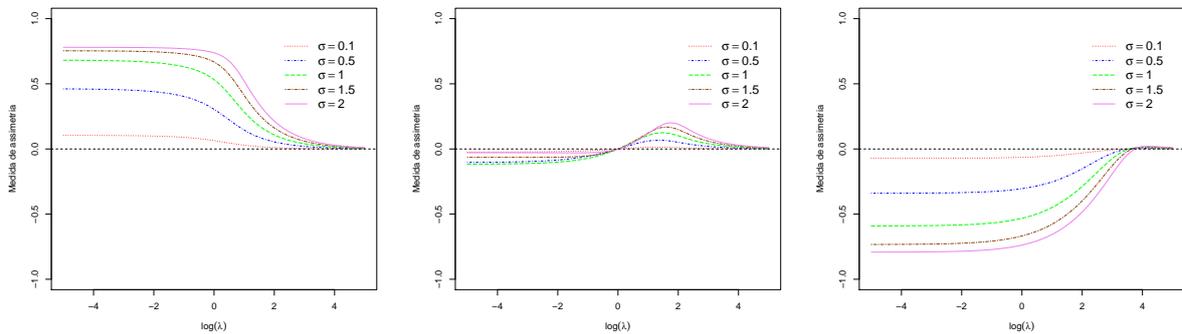


Figura 2.15: Medida de assimetria para a distribuição *PL-t*₍₂₎ considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .

A Figura 2.16 mostra a relação do parâmetro extra (ζ) e do λ na assimetria da distribuição *PL-t*_(ζ). Os parâmetros μ e σ estão fixos. Observe, através da figura da esquerda, que a assimetria diminui quando λ aumenta, para todos os valores de ζ . No gráfico da direita notamos que o parâmetro ζ não influencia substancialmente na assimetria da distribuição. Note que, para λ fixo, a assimetria da distribuição não varia muito quando ζ cresce.

Gráficos para outros modelos na classe *power logit* foram feitos e foram observados comportamentos similares aos apresentados aqui; esses gráficos estão disponíveis na Seção D.1. Assim, uma vez que o parâmetro de forma λ está fortemente relacionado com a assimetria das distribuições *power logit*, ele será tratado como um parâmetro de assimetria.

2.3 Inferência sobre os parâmetros da classe *power logit*

Consideramos o método da máxima verossimilhança usual como foco principal para obter estimativas para os parâmetros do modelo. Assim, considere Y_1, \dots, Y_n uma amostra aleatória de tamanho n de $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$ e $\theta = (\mu, \sigma, \lambda)^\top$ o vetor de parâmetros de interesse. Caso o

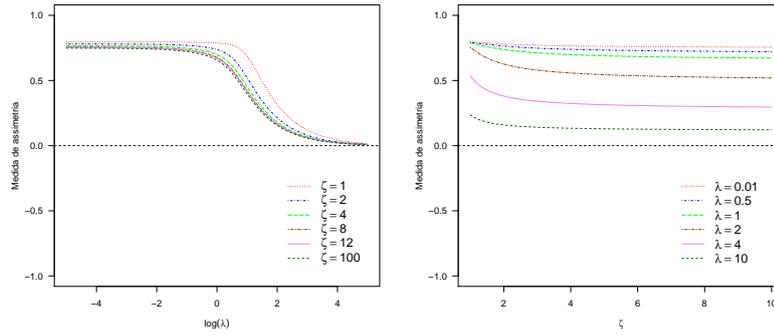


Figura 2.16: Medida de assimetria para a distribuição $PL-t_{(\zeta)}$ considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ .

gerador de densidades $r(\cdot)$ dependa de parâmetros extras, por exemplo o parâmetro de graus de liberdade da distribuição t-Student, este será considerado fixado por enquanto. A função de log-verossimilhança de $\boldsymbol{\theta}$ considerando a amostra observada y_1, \dots, y_n é da forma

$$\ell(\boldsymbol{\theta}) = n \log(\lambda/\sigma) - \sum_{i=1}^n \log(1 - y_i^\lambda) + \sum_{i=1}^n \log\{r(z_i^2)\} + c,$$

em que $z_i = h(y_i; \mu, \sigma, \lambda)$ e c é uma constante. O escore é dado por $\mathbf{U}(\boldsymbol{\theta}) = (U_\mu(\boldsymbol{\theta}), U_\sigma(\boldsymbol{\theta}), U_\lambda(\boldsymbol{\theta}))^\top$, cujos componentes são

$$\begin{aligned} U_\mu(\boldsymbol{\theta}) &= \frac{\lambda}{\sigma\mu(1 - \mu^\lambda)} \sum_{i=1}^n z_i v(z_i), \\ U_\sigma(\boldsymbol{\theta}) &= \frac{1}{\sigma} \sum_{i=1}^n [z_i^2 v(z_i) - 1], \\ U_\lambda(\boldsymbol{\theta}) &= \frac{n}{\lambda} + \sum_{i=1}^n \frac{y_i^\lambda \log y_i}{1 - y_i^\lambda} - \frac{1}{\sigma} \sum_{i=1}^n z_i v(z_i) \left[\frac{\log y_i}{(1 - y_i^\lambda)} - \frac{\log \mu}{(1 - \mu^\lambda)} \right], \end{aligned}$$

em que $v(t) = -2r'(t^2)/r(t^2)$ e $r'(u) = dr(u)/du$. Note que $v(t)$ pode ser vista como uma função peso e $v(t) > 0$ se $r(t)$ for uma função monotonicamente decrescente para $t > 0$. Os estimadores de máxima verossimilhança de μ , σ e λ podem ser escritos, respectivamente, como

$$\hat{\mu} = \left(\frac{\hat{\mu}^{(\hat{\lambda})}}{1 + \hat{\mu}^{(\hat{\lambda})}} \right)^{1/\hat{\lambda}}, \quad \hat{\sigma} \propto \left(\frac{\sum_{i=1}^n v(\hat{z}_i) \log(y_i^{(\hat{\lambda})}/\hat{\mu}^{(\hat{\lambda})})}{\sum_{i=1}^n v(\hat{z}_i)} \right)^{1/2}$$

e

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n \frac{\log y_i}{1 - y_i^{\hat{\lambda}}} \left(\frac{1}{\hat{\sigma}} \hat{z}_i v(\hat{z}_i) - y_i \right)},$$

em que $\hat{\mu}^{(\hat{\lambda})} = \left[\prod_{i=1}^n y_i^{(\hat{\lambda})} v(\hat{z}_i) \right]^{1/\sum_{i=1}^n v(\hat{z}_i)}$. Note que, desde que $v(\hat{z}_i) > 0$, para $i = 1, \dots, n$, o estimador de máxima verossimilhança de $\mu^{(\lambda)}$, pode ser interpretado como uma média geométrica ponderada dos valores $y_1^{(\hat{\lambda})}, \dots, y_n^{(\hat{\lambda})}$, enquanto o estimador de máxima verossimilhança de σ é proporcional à raiz quadrada de uma média aritmética ponderada, em que $v(\hat{z}_1), \dots, v(\hat{z}_n)$ são os

pesos para cada observação.

O estimador de máxima verossimilhança de $\boldsymbol{\theta}$ pode ser obtido através da solução simultânea do sistema de equações $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$ ou, equivalentemente, através da maximização da função de log-verossimilhança $\ell(\boldsymbol{\theta})$. Essa maximização pode ser feita através de procedimentos numéricos, usando algum algoritmo de otimização não linear como, por exemplo, Newton-Raphson, escore de Fisher, BFGS, entre outros. Estes algoritmos necessitam que sejam especificados valores iniciais para os parâmetros que desejamos estimar. Sugerimos utilizar a média e o desvio padrão amostral de $\log(y_i/(1 - y_i))$ como estimativas iniciais para μ e σ , respectivamente. Outra possibilidade para um ponto inicial para μ é a mediana amostral. Para o parâmetro λ , sugerimos o valor 1 como ponto inicial (ou seja, considerar o modelo GJS inicialmente).

A escolha de $r(\cdot)$ induz uma função $v(\cdot)$ que permite uma ponderação na estimação dos parâmetros da distribuição *power logit* via máxima verossimilhança. A escolha dessa função pode ser vista como uma maneira de robustificar o método de estimação para observações extremas ou *outliers*, uma vez que a escolha de $r(\cdot)$ pode induzir uma função $v(z)$ cujos valores decrescem em z à medida que o valor de z se afasta do centro da distribuição de Z . A Tabela 2.1 apresenta a função peso para alguns modelos na classe *power logit*.

Tabela 2.1: Função peso para alguns modelos na classe *power logit*.

Distribuição	$v(z)$
PL-N	1
PL- $t_{(\zeta)}$	$\frac{\zeta + 1}{\zeta + z^2}$
PL- $PE_{(\zeta)}$	$\frac{\zeta(z^2)^{\zeta/2-1}}{2p(\zeta)^\zeta}$
PL-LO	$-2(e^{-z^2} - 1)/(e^{-z^2} + 1)$
PL-slash $_{(\zeta)}$	$\frac{2}{z^2} G\left(\zeta + \frac{3}{2}, \frac{z^2}{2}\right) G\left(\zeta + \frac{1}{2}, \frac{z^2}{2}\right)^{-1}$
PL-Hyp $_{(\zeta)}$	$\zeta/\sqrt{1 + z^2}$
PL-SN $_{(\zeta)}$	$\frac{\sinh(z)}{z} \left[\frac{4 \cosh(z)}{\zeta^2} - \frac{1}{\cosh(z)} \right]$

Os valores dos pesos $v(z)$ são positivos para as distribuições PL-N, PL- $t_{(\zeta)}$, PL-LO, PL- $PE_{(\zeta)}$ (se $\zeta \neq 1$), PL-slash, PL-Hyp $_{(\zeta)}$ e para a distribuição PL-SN $_{(\zeta)}$ (para $\zeta \leq 2$). Adicionalmente, para as distribuições que têm caudas mais pesadas do que a distribuição normal (PL- $t_{(\zeta)}$ e PL- $PE_{(\zeta)}$, por exemplo), os pesos individuais tendem a ser menores à medida que z se distancia do centro da distribuição de Z . Dessa forma, as estimativas de μ e σ tendem a ser menos sensíveis a observações extremas quando consideramos distribuições com caudas mais pesadas do que o modelo PL-N, por exemplo.

Sob condições gerais de regularidade, o estimador de máxima verossimilhança de $\boldsymbol{\theta} = (\mu, \sigma, \lambda)^\top$ tem distribuição assintótica dada por

$$\sqrt{n} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \\ \hat{\lambda} - \lambda \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_3(\mathbf{0}_3, \mathbf{K}(\boldsymbol{\theta})^{-1}), \text{ quando } n \rightarrow \infty,$$

em que $\mathbf{0}_3$ denota um vetor coluna de zeros de dimensão 3 e $\mathbf{K}(\boldsymbol{\theta})$ é a matriz da informação de Fisher para uma única observação, dada por

$$\mathbf{K}(\boldsymbol{\theta}) = n^{-1} \mathbf{K}_n(\boldsymbol{\theta}) = n^{-1} \mathbb{E} \left[\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right] = \begin{bmatrix} \mathbf{K}_{\mu\mu} & 0 & \mathbf{K}_{\mu\lambda} \\ 0 & \mathbf{K}_{\sigma\sigma} & \mathbf{K}_{\sigma\lambda} \\ \mathbf{K}_{\lambda\mu} & \mathbf{K}_{\lambda\sigma} & \mathbf{K}_{\lambda\lambda} \end{bmatrix},$$

em que $\mathbf{K}_n(\boldsymbol{\theta})$ é a matriz da informação de Fisher.

$$\begin{aligned} K_{\mu\mu} &= \frac{\lambda^2}{[\sigma\mu(1-\mu^\lambda)]^2} d_r(\zeta), \\ K_{\sigma\sigma} &= \frac{1}{\sigma^2} [f_r(\zeta) - 1], \\ K_{\lambda\lambda} &= m_r(\zeta, \boldsymbol{\theta}) + \frac{2}{\sigma} \frac{\log \mu}{1-\mu^\lambda} n_r(\zeta, \boldsymbol{\theta}) + \left(\frac{\log \mu}{1-\mu^\lambda} \right)^2 \frac{1}{\sigma^2} d_r(\zeta) - \frac{1}{\lambda}, \\ K_{\mu\lambda} &= K_{\lambda\mu} = \frac{\lambda}{\sigma\mu(1-\mu^\lambda)} \left[n_r(\zeta, \boldsymbol{\theta}) + \frac{\log \mu}{1-\mu^\lambda} d_r(\zeta) \right], \\ K_{\sigma\lambda} &= K_{\lambda\sigma} = \frac{1}{\sigma\lambda} + \frac{1}{\sigma^2} h_r(\zeta, \boldsymbol{\theta}), \end{aligned}$$

sendo $d_r(\zeta) = \mathbb{E}[Z^2 v^2(Z)]$, $f_r(\zeta) = \mathbb{E}[Z^4 v^2(Z)]$, $m_r(\zeta, \boldsymbol{\theta}) = \sigma^{-1} \mathbb{E}\{[(\log Y/(1-Y^\lambda))(\sigma Y^\lambda - v(Z)Z)]^2\}$, $n_r(\zeta, \boldsymbol{\theta}) = \sigma^{-1} \mathbb{E}\{(\log Y/(1-Y^\lambda))[\sigma Y^\lambda - v(Z)Z]v(Z)Z\}$ e $h_r(\zeta, \boldsymbol{\theta}) = \mathbb{E}\{(\log Y/(1-Y^\lambda))[\sigma Y^\lambda - v(Z)Z]v(Z)Z^2\}$, com $Z \sim S(0, 1; r)$. Note que $d_r(\cdot)$ e $f_r(\cdot)$ são funções apenas do parâmetro extra ζ , enquanto que as funções $m_r(\cdot, \cdot)$, $n_r(\cdot, \cdot)$ e $h_r(\cdot, \cdot)$ envolvem ζ e $\boldsymbol{\theta}$. Algumas dessas funções não apresentem forma analítica fechada e podem ser calculadas numericamente através de *softwares* como R, MATLAB, etc.. Uma condição necessária para que a matriz $\mathbf{K}_n(\boldsymbol{\theta})$ esteja bem definida é que as funções $d_r(\zeta)$, $f_r(\zeta)$, $m_r(\zeta, \boldsymbol{\theta})$, $n_r(\zeta, \boldsymbol{\theta})$ e $h_r(\zeta, \boldsymbol{\theta})$ existam e sejam finitas. A depender do gerador de densidades $r(\cdot)$ e dos valores de ζ , essa condição pode não ser satisfeita.

Embora não exista uma forma analítica fechada para a matriz de informação de Fisher, resultados assintóticos mostram que podemos utilizar a matriz de informação observada, denotada aqui por $\mathbf{J}_n(\boldsymbol{\theta})$, para realizar inferência sob os parâmetros do modelo. A matriz $\mathbf{J}_n(\boldsymbol{\theta})$ é dada por

$$\mathbf{J}_n(\boldsymbol{\theta}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{bmatrix} J_{\mu\mu} & J_{\mu\sigma} & J_{\mu\lambda} \\ J_{\sigma\mu} & J_{\sigma\sigma} & J_{\sigma\lambda} \\ J_{\lambda\mu} & J_{\lambda\sigma} & J_{\lambda\lambda} \end{bmatrix},$$

com

$$\begin{aligned} J_{\mu\mu} &= \frac{-\lambda}{[\sigma\mu(1-\mu^\lambda)]^2} \left\{ \sigma[(\lambda+1)\mu^\lambda - 1] \sum_{i=1}^n z_i v(z_i) - \lambda \sum_{i=1}^n [v(z_i) + z_i v'(z_i)] \right\}, \\ J_{\sigma\sigma} &= -\frac{1}{\sigma^2} \left\{ n - \sum_{i=1}^n z_i^2 [3v(z_i) + z_i v'(z_i)] \right\}, \\ J_{\mu\sigma} &= J_{\sigma\mu} = \frac{\lambda}{\sigma^2 \mu(1-\mu^\lambda)} \sum_{i=1}^n z_i [2v(z_i) + z_i v'(z_i)], \\ J_{\lambda\lambda} &= \frac{n}{\lambda^2} - \sum_{i=1}^n \left\{ \frac{y_i^\lambda \log^2 y_i}{(1-y_i^\lambda)^2} - \frac{1}{\sigma^2} \left(\frac{\log y_i}{1-y_i^\lambda} - \frac{\log \mu}{1-\mu^\lambda} \right)^2 [v(z_i) + z_i v'(z_i)] \right. \\ &\quad \left. + \frac{1}{\sigma} z_i v(z_i) \left[\frac{y_i^\lambda \log^2 y_i}{(1-y_i^\lambda)^2} - \frac{\mu^\lambda \log^2 \mu}{(1-\mu^\lambda)^2} \right] \right\}, \\ J_{\mu\lambda} &= J_{\lambda\mu} = \frac{\lambda}{\sigma^2 \mu(1-\mu^\lambda)} \left[\frac{\sigma\mu^\lambda [1 - \lambda \log \mu] - 1}{\lambda(1-\mu^\lambda)} \sum_{i=1}^n z_i v(z_i) \right. \\ &\quad \left. + \sum_{i=1}^n \left(\frac{\log y_i}{1-y_i^\lambda} - \frac{\log \mu}{1-\mu^\lambda} \right) [v(z_i) + z_i v'(z_i)] \right], \\ J_{\sigma\lambda} &= J_{\lambda\sigma} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left(\frac{\log y_i}{1-y_i^\lambda} - \frac{\log \mu}{1-\mu^\lambda} \right) z_i [2v(z_i) + z_i v'(z_i)]. \end{aligned}$$

Espera-se que o método usual de estimação via máxima verossimilhança apresente bom com-

portamento em grandes amostras. Entretanto, estudos prévios com amostras de tamanho moderado mostraram que, em alguns casos, a função de log-verossimilhança perfilada para λ apresenta fraca concavidade ou mesmo exibe forma monótona.

A Figura 2.17 apresenta o gráfico da função de log-verossimilhança perfilada relativa para λ e as curvas de contorno da função de verossimilhança relativa variando (σ, λ) e mantendo μ fixo no verdadeiro valor. A versão relativa da função de log-verossimilhança é dada pela diferença entre $\ell(\boldsymbol{\theta})$ e $\ell(\hat{\boldsymbol{\theta}})$. Esses gráficos foram obtidos a partir de uma amostra de tamanho 50 da distribuição PL- $t_{(2)}$ com $\mu = 0.7$, $\sigma = 0.5$ e $\lambda = 1$. A função de log-verossimilhança perfilada para λ apresenta fraca concavidade. Adicionalmente, nas curvas de contorno percebemos o quanto as estimativas de máxima verossimilhança de σ e, principalmente, de λ são afetadas. Note que a função de verossimilhança não é maximizada próximo dos valores verdadeiros. As estimativas de máxima verossimilhança, para essa amostra, foram $\hat{\mu} = 0.6991$, $\hat{\sigma} = 1.4450$ e $\hat{\lambda} = 10.2014$.

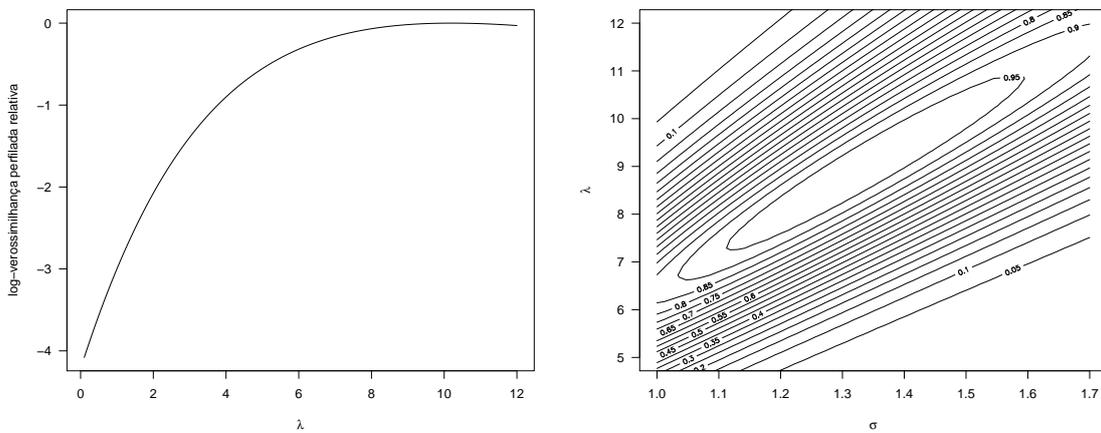


Figura 2.17: Função de log-verossimilhança perfilada relativa para λ (figura da esquerda) e curvas de contorno da função de verossimilhança relativa (figura da direita).

Nestes casos, a forma da função de log-verossimilhança perfilada sugere o uso de alguma penalização adequada. Esta estratégia é utilizada em diversos contextos, entre eles: modelo de regressão de Cox (Bryson e Johnson, 1981); modelo de regressão logística (Albert e Anderson, 1984); distribuições *skew normal* e *skew t* (Azzalini e Capitanio, 1999; Sartori, 2006); distribuição Weibull estendida modificada (Lima e Cribari-Neto, 2019). Além disso, Sartori (2003) comenta que, quando há muitos parâmetros de perturbação, a função de log-verossimilhança penalizada leva a estimadores com melhores propriedades em amostras finitas do que a função de log-verossimilhança usual.

2.3.1 Inferência via máxima verossimilhança penalizada

Existem diferentes formas de penalizar a função de log-verossimilhança, de modo a obter estimadores com melhores propriedades. Uma das principais penalizações presentes na literatura é baseada no método de redução de viés proposto por Firth (1993).

Considere um modelo paramétrico regular com função de log-verossimilhança $\ell(\boldsymbol{\xi})$, em que $\boldsymbol{\xi}$ é o vetor de parâmetros e sejam $\mathbf{U}(\boldsymbol{\xi})$ e $\mathbf{J}_n(\boldsymbol{\xi})$ o vetor escore e a matriz da informação observada, respectivamente. Além disso, seja $\mathbf{b}(\boldsymbol{\xi})$ o viés de primeira ordem de $\hat{\boldsymbol{\xi}}$, isto é, $\mathbb{E}_{\boldsymbol{\xi}}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \mathbf{b}(\boldsymbol{\xi}) + \mathcal{O}(n^{-2})$. A ideia do método de Firth é modificar a função escore através de uma função $\mathbf{M}(\boldsymbol{\xi})$ da seguinte forma

$$\mathbf{U}^*(\boldsymbol{\xi}) = \mathbf{U}(\boldsymbol{\xi}) + \mathbf{M}(\boldsymbol{\xi}),$$

sendo $\mathbf{M}(\boldsymbol{\xi})$ escolhida de modo que $\mathbb{E}_{\boldsymbol{\xi}}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \mathcal{O}(n^{-2})$, em que $\tilde{\boldsymbol{\xi}}$ é a solução de $\mathbf{U}^*(\boldsymbol{\xi}) = 0$. Dessa

forma, pode-se mostrar que

$$\mathbf{M}(\boldsymbol{\xi}) = -\mathbf{K}_n(\boldsymbol{\xi})\mathbf{b}(\boldsymbol{\xi}), \quad (2.7)$$

em que $\mathbf{K}_n(\boldsymbol{\xi}) = \mathbb{E}_{\boldsymbol{\xi}}\{\mathbf{J}_n(\boldsymbol{\xi})\}$ é a matriz de informação de Fisher.

Em modelos exponenciais de posto completo, $\tilde{\boldsymbol{\xi}}$ coincide com a moda a *posteriori* da distribuição obtida usando a distribuição a priori de Jeffreys $|\mathbf{K}(\boldsymbol{\xi})|^{1/2}$, em que $\mathbf{K}(\boldsymbol{\xi})$ é a informação esperada para uma única observação. Para modelos em geral, essa equivalência não é válida. Por outro lado, [Azzalini e Arellano-Valle \(2013\)](#) observam que há uma similaridade, em termos de forma, entre $\mathbf{M}(\boldsymbol{\xi})$ e a derivada do logaritmo da distribuição a priori de Jeffreys.

De maneira geral, uma função de log-verossimilhança $\ell(\boldsymbol{\xi})$ de um parâmetro $\boldsymbol{\xi}$ pode ser penalizada da seguinte forma

$$\ell_p(\boldsymbol{\xi}) = \ell(\boldsymbol{\xi}) - Q,$$

sendo Q uma quantidade não negativa que penaliza a função $\ell(\boldsymbol{\xi})$ e é de ordem $\mathcal{O}_p(1)$ quando n cresce. Se existir, a quantidade $\tilde{\boldsymbol{\xi}}$, que maximiza $\ell_p(\boldsymbol{\xi})$, é chamada estimador de máxima verossimilhança penalizada. A função Q depende dos parâmetros e pode, ou não, depender dos dados. Uma consequência dessa formulação é que a distribuição assintótica de $\tilde{\boldsymbol{\xi}}$ e $\hat{\boldsymbol{\xi}}$ é a mesma até a primeira ordem. Isso ocorre porque $\ell(\boldsymbol{\xi})$ e $\ell_p(\boldsymbol{\xi})$ são ambas de ordem $\mathcal{O}_p(n)$ e diferem apenas através de Q , que é de ordem $\mathcal{O}_p(1)$. Assim, sob algumas condições de regularidade, concluímos que

$$\sqrt{n}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{K}(\boldsymbol{\xi})^{-1}), \text{ quando } n \rightarrow \infty.$$

Na prática, é comum utilizar o negativo da matriz de segundas derivadas de $n^{-1}\ell(\hat{\boldsymbol{\xi}})$ como uma aproximação para $\mathbf{K}(\boldsymbol{\xi})$ e, assim, obter os erros padrão de $\tilde{\boldsymbol{\xi}}$. Analogamente, como $\tilde{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}} = \mathcal{O}_p(n^{-1})$ e assumindo a existência da terceira derivada de Q , então

$$\text{Cov}(\tilde{\boldsymbol{\xi}}) \approx -\ell_p''(\tilde{\boldsymbol{\xi}})^{-1}.$$

A escolha da função Q pode variar. Por exemplo, tomando-se Q tal que $\mathbf{M}(\boldsymbol{\xi}) = -\partial Q(\boldsymbol{\xi})/\partial \boldsymbol{\xi}$, em que $\mathbf{M}(\boldsymbol{\xi})$ é dada em (2.7), elimina-se o viés de primeira ordem de $\tilde{\boldsymbol{\xi}}$. Uma dificuldade em obter Q a partir do critério de redução de viés é trabalhar com as expressões explícitas de $\mathbf{b}(\boldsymbol{\xi})$. [Azzalini e Arellano-Valle \(2013\)](#) estudam diferentes escolhas da função Q para os modelos *skew-normal* e *skew-t* multivariados.

Neste trabalho consideramos uma penalização na função de log-verossimilhança através de um fator que é equivalente a penalizar a função de verossimilhança utilizando a priori de Jeffreys ([Jeffreys, 1946](#)). Chamaremos esse método de penalização via priori de Jeffreys. Existem diversos trabalhos na literatura que utilizam essa estratégia, entre eles [Fonseca et al. \(2012\)](#), [Firth \(1992\)](#) e [Pianto e Cribari-Neto \(2011\)](#). Consideramos uma penalização apenas no escore associado a λ , isto é, $U_\lambda(\boldsymbol{\theta})$; equivalentemente, penalizamos a função de log-verossimilhança perfilada de λ . Embora seja possível penalizar todos os componentes do vetor escore de $\boldsymbol{\theta}$, essa abordagem poderia ser computacionalmente complicada, principalmente se considerarmos o contexto de regressão. Além disso, estudos de simulação previamente realizados sugerem que os estimadores de máxima verossimilhança de μ e σ apresentam boas propriedades.

Dessa forma, a função de log-verossimilhança perfilada penalizada para λ é dada por

$$\ell_p^*(\lambda) = \ell^*(\lambda) + \frac{1}{2} \log \left(\frac{\mathbf{J}_{\lambda\lambda}^*}{n} \right),$$

em que $\ell^*(\lambda) = \ell(\hat{\mu}_\lambda, \hat{\sigma}_\lambda, \lambda)$ é a função de log-verossimilhança perfilada para λ e $\mathbf{J}_{\lambda\lambda}^* = \mathbf{J}_{\lambda\lambda}(\hat{\mu}_\lambda, \hat{\sigma}_\lambda, \lambda)$, sendo $\hat{\mu}_\lambda$ e $\hat{\sigma}_\lambda$ as estimativas de máxima verossimilhança para μ e σ , respectivamente, para λ fixo. Note que estamos utilizando a versão observada da matriz de informação de Fisher. De fato, a penalização deveria ser feita utilizando as quantidades esperadas, isto é, a quantidade $\mathbf{K}_{\lambda\lambda}$. Isso não foi feito aqui por dois motivos: não existe forma analítica fechada para a quantidade $\mathbf{K}_{\lambda\lambda}$ e

teríamos que trabalhar com integração numérica, e estudos de simulação conduzidos com o uso de $J_{\lambda\lambda}$ apresentaram bons resultados.

Novamente, não existem expressões analíticas fechadas para o estimador de máxima verossimilhança penalizada de θ , denotado aqui por $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma}, \tilde{\lambda})^\top$. As estimativas de máxima verossimilhança penalizada podem ser encontradas através dos seguintes passos.

i. Encontre $\tilde{\lambda}$ tal que

$$\tilde{\lambda} = \operatorname{argmax}_{\lambda > 0} \ell_p^*(\lambda).$$

ii. Obtenha $\tilde{\mu}$ e $\tilde{\sigma}$ maximizando $\ell(\mu, \sigma, \tilde{\lambda})$.

Observe que, embora a penalização seja aplicada apenas à log-verossimilhança perfilada de λ , as estimativas de μ e σ também são afetadas por essa modificação. Conforme mencionado anteriormente, a distribuição assintótica de primeira de $\tilde{\theta}$ é igual à de $\hat{\theta}$, isto é, sob condições gerais de regularidade,

$$\sqrt{n} \begin{pmatrix} \tilde{\mu} - \mu \\ \tilde{\sigma} - \sigma \\ \tilde{\lambda} - \lambda \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_3(\mathbf{0}_3, \mathbf{K}(\theta)^{-1}), \text{ quando } n \rightarrow \infty.$$

Podemos utilizar a matriz $\mathbf{J}_n(\theta)$ para obter erros-padrão assintóticos e intervalos de confiança aproximados para μ , σ e λ . Uma alternativa à construção de intervalo através da distribuição assintótica de $\tilde{\lambda}$ é utilizar $W_p^*(\lambda)$, que tem distribuição assintótica χ_1^2 (Barndorff-Nielsen, 1995), definida por

$$W_p^*(\lambda) = 2\{\ell_p^*(\tilde{\lambda}) - \ell_p^*(\lambda)\}.$$

Adicionalmente, estudos de simulação mostram que, em algumas amostras, a estimativa de máxima verossimilhança pode ser muito pequena (tender a zero). Neste caso, o processo de estimação indica que o modelo limite, apresentado na Proposição 2.5, é o adequado. Isto é, o ajuste de uma distribuição simétrica com parâmetros $-\log(-\log \mu)$ e σ^2 à variável $-\log(-\log Y)$. Uma aplicação deste caso é considerada na Seção 2.5.

Para ilustrar o uso dessa abordagem, consideremos novamente os dados da Figura 2.17. O gráfico da função de log-verossimilhança perfilada penalizada para esses dados é apresentado na Figura 2.18.

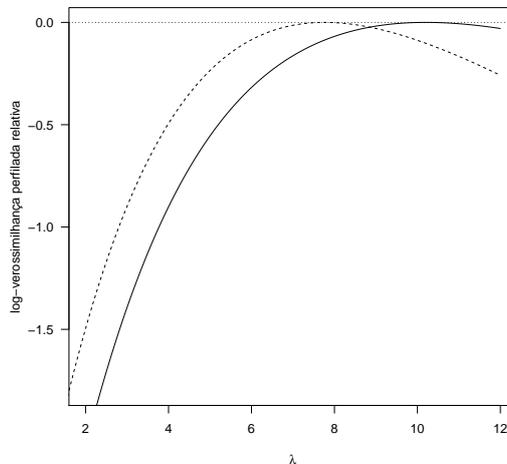


Figura 2.18: Função de log-verossimilhança perfilada relativa (linha sólida) e log-verossimilhança perfilada penalizada relativa (linha tracejada) para λ .

Note que a função de log-verossimilhança penalizada não mais apresenta fraca concavidade (ver linha tracejada).

A Figura 2.19 apresenta os gráficos da função de log-verossimilhança penalizada relativa para diferentes amostras geradas a partir de diferentes distribuições na classe *power logit*. Consideramos os modelos PL-LO, PL-PE_(1,2), PL-SN₍₃₎ e PL-Hyp₍₃₎, com $\mu = 0.7$, $\sigma = 0.5$ e $\lambda = 1$. Esses gráficos sugerem que a penalização é eficiente para corrigir a fraca concavidade da função de log-verossimilhança. As estimativas de máxima verossimilhança penalizada dos parâmetros do modelo, para cada amostra em estudo, são apresentadas na Tabela 2.2.

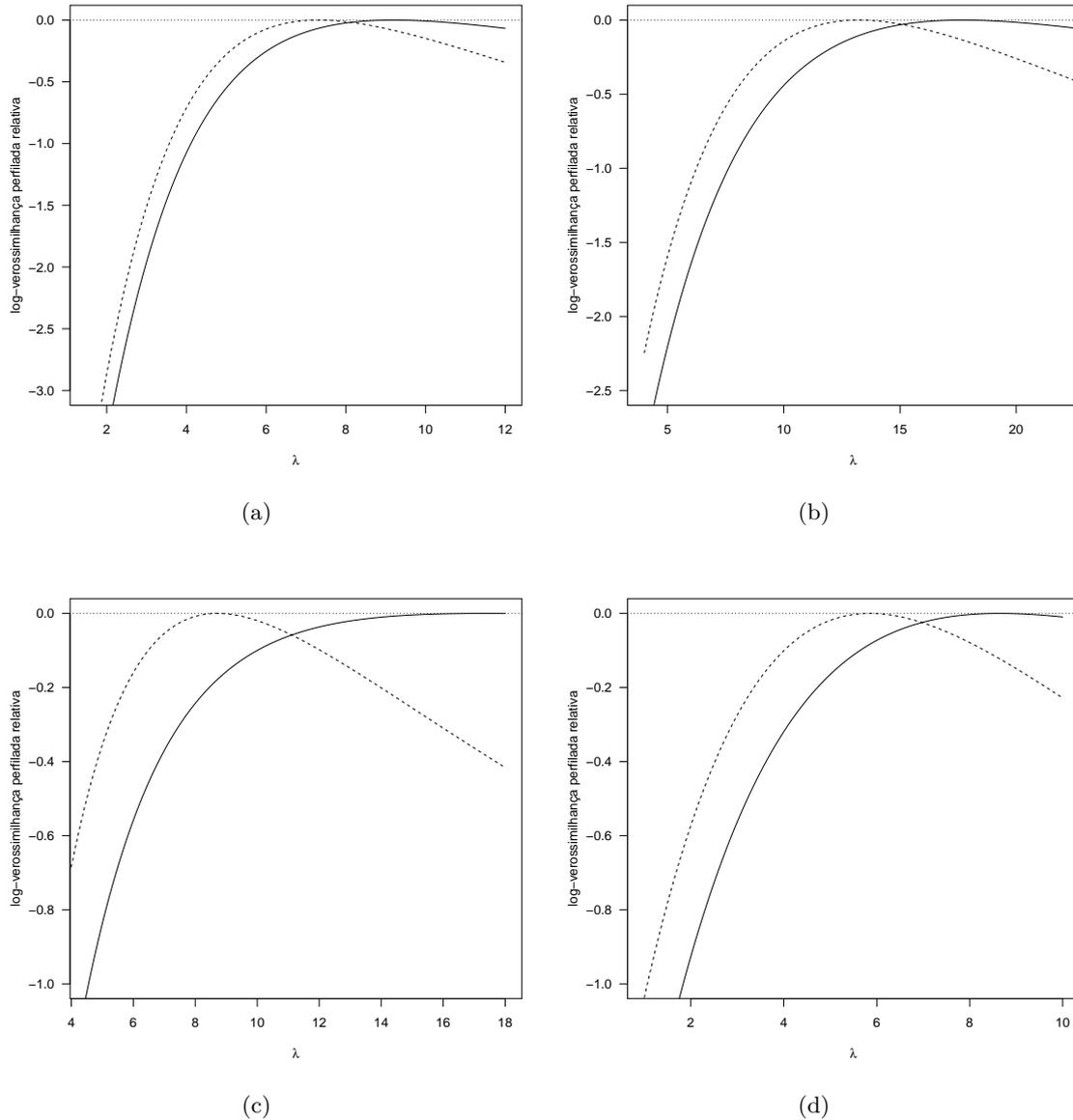


Figura 2.19: Função de log-verossimilhança perfilada relativa (linha sólida) e log-verossimilhança perfilada penalizada relativa (linha tracejada) para λ considerando amostras de tamanho $n = 50$ geradas dos modelos (a) PL-LO, (b) PL-PE_(1,2), (c) PL-SN₍₃₎ e (d) PL-Hyp₍₃₎, considerando $\mu = 0.7$, $\sigma = 0.5$ e $\lambda = 1$.

Tabela 2.2: Estimativas de máxima verossimilhança usual e penalizada para $\theta = (\mu, \sigma, \lambda)^\top$, considerando as amostras da Figura 2.19.

	(a)	(b)	(c)	(d)
$\hat{\mu}$	0.67	0.73	0.66	0.70
$\tilde{\mu}$	0.68	0.73	0.66	0.70
$\hat{\sigma}$	2.69	2.25	2.53	1.42
$\tilde{\sigma}$	2.26	1.74	1.34	1.06
$\hat{\lambda}$	9.20	17.70	17.35	8.64
$\tilde{\lambda}$	7.32	13.22	8.71	5.85

2.3.2 Qualidade do ajuste e estimação de ζ

Motivados por Vanegas e Paula (2016), uma medida que pode ser útil na verificação da qualidade do ajuste de uma distribuição na classe *power logit* é definida por

$$\Upsilon_\zeta = n^{-1} \sum_{i=1}^n |\Phi^{-1}[R(\tilde{z}^{(i)})] - v^{(i)}|,$$

em que $\tilde{z}^{(i)}$ é a i -ésima estatística de ordem de \tilde{z} , $v^{(i)}$ é o valor esperado da i -ésima estatística de ordem para uma amostra aleatória de tamanho n de uma variável aleatória com distribuição normal padrão, $N(0, 1)$, $\Phi^{-1}(\cdot)$ é a função quantílica de uma variável aleatória com distribuição $N(0, 1)$ e $R(\cdot)$ é a função de distribuição acumulada de uma variável aleatória com distribuição $S(0, 1; r)$. Note que $\Phi^{-1}[R(z)]$ tem distribuição $N(0, 1)$ e, como $\tilde{\mu}$, $\tilde{\sigma}$ e $\tilde{\lambda}$ são estimadores consistentes para μ , σ e λ , respectivamente, $\Phi^{-1}[R(\tilde{z})]$ tem distribuição assintótica normal padrão. Dessa forma, se o modelo está bem ajustado, espera-se que os quantis $\Phi^{-1}[R(\tilde{z}^{(i)})]$ e $v^{(i)}$ estejam próximos, para todo $i = 1, \dots, n$. Assim, valores pequenos de Υ_ζ indicam um melhor ajuste.

A quantidade $r_i = \Phi^{-1}[R(\tilde{z}^{(i)})]$, para $i = 1, \dots, n$, pode ser vista como um resíduo quantílico (Dunn e Smyth, 1996). Dessa, pode-se fazer o gráfico de probabilidade normal de r_1, \dots, r_n com envelopes simulados a fim de verificar a adequação do modelo em estudo. Os envelopes simulados foram propostos por Atkinson (1981) e podem ser construídos da seguinte forma:

- i. ajuste o modelo de interesse e obtenha as estimativas dos parâmetros do modelo;
- ii. gere uma amostra de n observações independentes usando os parâmetros estimados em i.;
- iii. ajuste o modelo de interesse na amostra gerada, calcule os valores dos resíduos e ordene-os em ordem crescente;
- iv. repita os passos ii. e iii. K vezes;
- v. calcule, para cada um dos n conjuntos de K resíduos ordenados, os quantis de ordem $p/2$, $1/2$ (mediana) e $1 - p/2$. Os envelopes são dados por esses valores calculados.
- vi. Coloque, em um mesmo gráfico, os valores calculados em v. (usualmente em linha) e os resíduos da amostra original ordenados *versus* os valores esperados das estatísticas de ordem de uma variável aleatória com distribuição normal padrão, isto é, $v^{(i)}$, $i = 1, \dots, n$.

Se o modelo está corretamente especificado, espera-se que mais do que $p \times 100\%$ das observações estejam dentro das bandas do envelope.

O critério da informação de Akaike (Akaike, 1973) pode ser utilizado para seleção de modelos. Esse critério é baseado na divergência de Kullback-Leibler e é definido da seguinte forma

$$\text{AIC} = -2\ell(\tilde{\theta}) + 2k,$$

em que k é o número de parâmetros do modelo. Dado um conjunto de modelos estatísticos, o modelo escolhido é aquele com menor AIC. Note que esse critério, diferente da medida Υ_ζ , leva em consideração o número de parâmetros do modelo.

As distribuições na classe *power logit* dependem de um gerador de densidades $r(\cdot)$, que pode envolver parâmetro extra (ζ). Estamos assumindo, até o momento, que este parâmetro (ou vetor de parâmetros) é fixado, tornando o processo de estimação mais simples. No entanto, em aplicações práticas, é necessário escolher um valor apropriado, ou obter uma estimativa, para esse parâmetro. Neste trabalho, estudamos a obtenção (ou estimação) do parâmetro ζ através da estatística Υ_ζ .

Conforme comentado anteriormente, quanto menor o valor observado de Υ_ζ , melhor o ajuste do modelo. Dessa forma, definimos $\hat{\zeta}$ como sendo uma estimativa para ζ tal que

$$\hat{\zeta} = \operatorname{argmin} \Upsilon_\zeta.$$

Desde que as funções $d_r(\cdot)$, $f_r(\cdot)$, $m_r(\cdot, \cdot)$, $n_r(\cdot, \cdot)$ e $h_r(\cdot, \cdot)$ sejam contínuas e finitas e $\hat{\zeta}$ seja um estimador consistente para ζ , a distribuição assintótica de $\hat{\theta}$ não é alterada quando substituímos ζ por $\hat{\zeta}$. Uma outra forma de obter uma estimativa para ζ é através do método de máxima verossimilhança. Este método, no entanto, não será abordado aqui pois estudos de simulação revelaram um bom desempenho da estimativa obtida através da minimização de Υ_ζ . Além disso, alguns problemas podem surgir na estimação de ζ via máxima verossimilhança em alguns modelos da classe simétrica, por exemplo, as funções de log-verossimilhança perfilada de ζ nos modelos t-Student e hiperbólico podem apresentar forma monótona (Fonseca *et al.*, 2008, 2012) e seria necessário o uso de alguma penalização, tornando o processo de estimação mais complexo.

2.4 Resultados numéricos

Os estimadores obtidos através do método de máxima verossimilhança penalizada apresentam boas propriedades quando o tamanho da amostra aumenta. Por outro lado, é importante estudar as propriedades desses estimadores em amostras de tamanho finito. Nesse sentido, simulações Monte Carlo foram realizadas para estudar o comportamento dos estimadores de θ e de ζ em amostras de tamanho pequeno e moderado. Todas as simulações foram realizadas no *software* R (R Core Team, 2019) e as maximizações foram feitas através do método BFGS com derivadas numéricas. Para a análise dos resultados da estimação pontual, foram calculadas, para cada modelo e para cada tamanho de amostra, algumas medidas, entre elas

$$\begin{aligned} \text{Média} &= \hat{\psi}_{med} = \frac{1}{k} \sum_{i=1}^k \hat{\psi}_i, \\ \text{Mediana} &= M = Q_2, \\ \text{Viés} &= b(\hat{\psi}_{med}) = \frac{1}{k} \sum_{i=1}^k (\hat{\psi}_i - \psi) = \hat{\psi}_{med} - \psi, \\ \text{DP} &= \sqrt{\frac{1}{k-1} \sum_{i=1}^k (\hat{\psi}_i - \hat{\psi}_{med})^2}, \\ \sqrt{\text{EQM}} &= \sqrt{\frac{1}{k} \sum_{i=1}^k (\hat{\psi}_i - \psi)^2}, \\ \text{IQ} &= Q_3 - Q_1, \end{aligned}$$

em que ψ é o valor verdadeiro do parâmetro, $\hat{\psi}_i$ é o estimador de máxima verossimilhança de ψ referente a i -ésima réplica, k é o número de réplicas Monte Carlo, e Q_1 , Q_2 e Q_3 são o primeiro, segundo e terceiro quartil, respectivamente.

Além disso, uma aplicação a dados simulados é apresentada a fim de estudar a sensibilidade

do modelo proposto a observações discrepantes e comparar o desempenho com outras distribuições para análise de dados de suporte limitado.

2.4.1 Desempenho dos estimadores de máxima verossimilhança penalizada

Inicialmente, consideramos um estudo de simulação com o objetivo de avaliar o comportamento dos estimadores de máxima verossimilhança usual e penalizado dos parâmetros da classe de distribuições *power logit* em amostras de tamanho finito. Consideramos diferentes distribuições: PL-N, PL- $t_{(2)}$, PL- $PE_{(1.5)}$, PL-LA, PL- $Hyp_{(1.2)}$ e PL- $SN_{(1.5)}$. Foram geradas amostras de tamanho 30, 60, 120 e 240. Os valores verdadeiros dos parâmetros do modelo são $\mu = 0.7$, $\sigma = 0.5$ e $\lambda = 1$. Finalmente, consideramos $k = 3000$ réplicas Monte Carlo e, para cada réplica, estimamos os parâmetros através do método de máxima verossimilhança usual e penalizado. A seguir, reportamos os resultados de simulação para os modelos PL- $t_{(2)}$, PL- $PE_{(1.5)}$ e PL- $Hyp_{(1.2)}$. A Tabela 2.3 apresenta, para cada tamanho amostra, a estimativa do viés e o erro-padrão das 3000 estimativas.

Tabela 2.3: Estimativa do viés e erro-padrão (e.p.) das 3000 estimativas de máxima verossimilhança (usual e penalizada) para os modelos PL- $t_{(2)}$, PL- $PE_{(1.5)}$ e PL- $Hyp_{(1.2)}$.

	$n = 30$		$n = 60$		$n = 120$		$n = 240$		
	Viés	e.p.	Viés	e.p.	Viés	e.p.	Viés	e.p.	
PL- $t_{(2)}$	$\hat{\mu}$	-0.00	0.03	0.00	0.02	0.00	0.01	-0.00	0.01
	$\tilde{\mu}$	-0.00	0.03	-0.00	0.02	-0.00	0.01	-0.00	0.01
	$\hat{\sigma}$	0.16	0.36	0.09	0.22	0.05	0.13	0.03	0.08
	$\tilde{\sigma}$	0.05	0.21	0.04	0.16	0.02	0.12	0.01	0.08
	$\hat{\lambda}$	1.62	3.25	0.93	2.01	0.51	1.28	0.28	0.81
	$\tilde{\lambda}$	0.46	1.81	0.37	1.52	0.22	1.13	0.14	0.77
PL- $PE_{(1.5)}$	$\hat{\mu}$	-0.00	0.02	-0.00	0.01	-0.00	0.01	-0.00	0.01
	$\tilde{\mu}$	-0.00	0.02	-0.00	0.01	-0.00	0.01	-0.00	0.01
	$\hat{\sigma}$	0.12	0.31	0.06	0.19	0.03	0.12	0.01	0.08
	$\tilde{\sigma}$	0.04	0.19	0.03	0.15	0.02	0.11	0.01	0.08
	$\hat{\lambda}$	1.35	3.09	0.69	1.90	0.34	1.26	0.16	0.84
	$\tilde{\lambda}$	0.55	1.90	0.39	1.52	0.22	1.15	0.10	0.81
PL- $Hyp_{(1.2)}$	$\hat{\mu}$	-0.00	0.03	-0.00	0.02	-0.00	0.01	-0.00	0.01
	$\tilde{\mu}$	-0.00	0.03	-0.00	0.00	-0.00	0.01	-0.00	0.01
	$\hat{\sigma}$	0.09	0.24	0.05	0.16	0.02	0.10	0.01	0.07
	$\tilde{\sigma}$	0.03	0.17	0.03	0.13	0.01	0.09	0.01	0.07
	$\hat{\lambda}$	1.11	2.43	0.58	1.62	0.26	0.96	0.13	0.67
	$\tilde{\lambda}$	0.53	1.68	0.35	1.33	0.16	0.90	0.09	0.65

Em geral, observamos que a estimativa do viés e o erro-padrão diminuem quando o tamanho da amostra aumenta, conforme esperado. Os estimadores de máxima verossimilhança para μ , tanto usual quanto penalizado, são aparentemente não viesados e apresentam baixa variabilidade. Por exemplo, para $n = 30$ e considerando o modelo PL- $PE_{(1.5)}$, os vieses estimados e erros-padrão de $\hat{\mu}$ e $\tilde{\mu}$ são, respectivamente, 0.00 e 0.02. Já o viés do estimador de máxima verossimilhança de σ é ligeiramente acentuado para amostras de tamanho pequeno ou moderado. Por exemplo, para $n = 30$, a estimativa do viés para o modelo PL- $t_{(2)}$ é 0.16. A estimativa do viés para o estimador de máxima verossimilhança do parâmetro λ é consideravelmente alta em amostras de tamanho pequeno e moderado, para todos os modelos em estudo. Note que, para o modelo PL- $t_{(2)}$, a estimativa do viés de $\hat{\lambda}$ é 1.62 e 0.93, quando $n = 30$ e $n = 60$, respectivamente. Além disso, os erros-padrão para estes casos são 3.25 e 2.01, respectivamente. Isso indica que, para amostras de tamanho pequeno ou moderado, a diferença entre a estimativa de máxima verossimilhança e o verdadeiro valor de λ é relativamente alta.

Contudo, observamos que a penalização melhora o desempenho dos estimadores de máxima verossimilhança de σ e, principalmente, de λ . Note que, em geral, as estimativas dos vieses e os erros-padrão dos estimadores de σ e λ são menores quando consideramos a penalização. Para o modelo PL-Hyp_(1,2) os vieses de $\hat{\sigma}$ e $\hat{\lambda}$ são 0.09 e 1.11, respectivamente, enquanto os vieses de $\tilde{\sigma}$ e $\tilde{\lambda}$ são 0.03 e 0.53. Já os erros-padrão de $\hat{\sigma}$ e $\hat{\lambda}$ são 0.24 e 2.43, respectivamente, e os de $\tilde{\sigma}$ e $\tilde{\lambda}$ são 0.17 e 1.68. Esse comportamento é observado em todos os modelos estudados. Os estimadores de máxima verossimilhança usual e penalizada apresentam desempenhos similares quando o tamanho da amostra é grande ($n = 240$, por exemplo). Em geral, notamos que os estimadores de máxima verossimilhança penalizado apresentam um ganho de precisão considerável, quando comparados com os estimadores obtidos sem a penalização; isso pode ser observado na Figura 2.20, em que são apresentados os boxplots das estimativas de máxima verossimilhança de λ para o modelo PL-Hyp_(1,2). O comportamento das estimativas de máxima verossimilhança de λ é similar para os demais modelos.

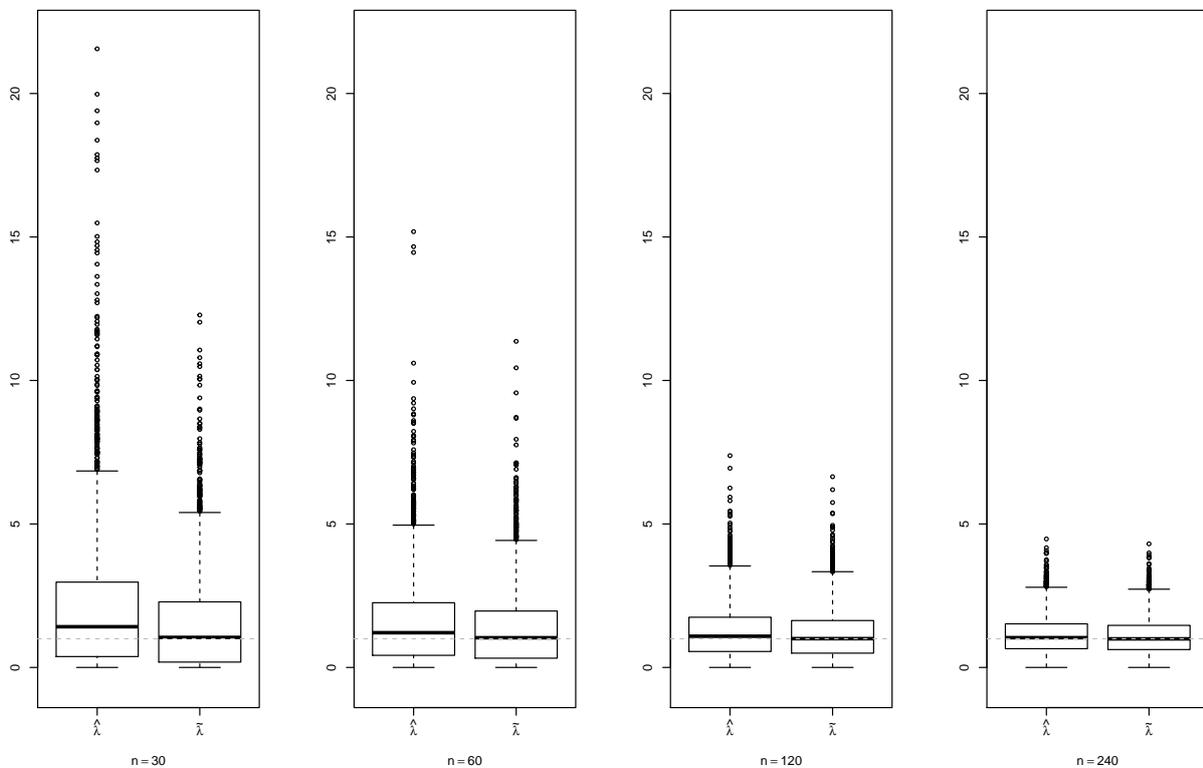


Figura 2.20: Boxplots das estimativas de máxima verossimilhança (usual e penalizada) para λ , considerando o modelo PL-Hyp_(1,2). A linha tracejada nos gráficos representa o valor verdadeiro $\lambda = 1$.

Com base na Figura 2.20, notamos que a variabilidade das estimativas de máxima verossimilhança é bastante alta quando comparada com a das estimativas de máxima verossimilhança penalizada. Além disso, notamos que a distribuição do estimador de máxima verossimilhança para λ (usual e penalizado) é fortemente assimétrica em amostras de tamanho pequeno.

Finalmente, notamos que a correção no estimador de máxima verossimilhança, através da penalização via priori de Jeffrey's, é efetiva. Os estimadores de máxima verossimilhança penalizados para σ e λ apresentam um ganho de precisão significativo e vieses consideravelmente menores do que os obtidos via máxima verossimilhança usual. Resultados similares são observados para as distribuições PL-N, PL-LA e PL-SN_(1,5) (ver Tabela E.1 no Apêndice).

2.4.2 Desempenho do estimador de ζ via minimização de Υ_ζ

Estudos de simulação foram conduzidos a fim de estudar o desempenho do critério proposto para a escolha do parâmetro ζ . Inicialmente, geramos $k = 3000$ amostras de tamanho n de membros da classe *power logit*, cujo gerador de densidade envolve um parâmetro extra. Para cada uma das k amostras, estimamos μ , σ e λ através do método da máxima verossimilhança penalizada e ζ minimizando a medida Υ_ζ . Consideramos diferentes distribuições na classe *power logit* (PL- $t_{(\zeta)}$, PL- $PE_{(\zeta)}$, PL- $SN_{(\zeta)}$ e PL- $Hyp_{(\zeta)}$) e diferentes tamanhos de amostra $n = 40, 80, 150$. Os valores dos parâmetros foram fixados em $\mu = 0.3$, $\sigma = 1$ e $\lambda = 0.5$. Na Tabela 2.4 apresentamos a mediana (M) e a distância interquartilica (IR) das k réplicas para a estimação de ζ . A Tabela E.2, no Apêndice E, apresenta a mediana e a distância interquartilica para os demais parâmetros (μ , σ e λ).

Tabela 2.4: Mediana (M) e distância interquartilica (IR) das M réplicas para a escolha de ζ , considerando diferentes modelos na classe *power logit* e diferentes tamanhos de amostra.

	ζ	$n = 40$		$n = 80$		$n = 150$	
		M	IQ	M	IQ	M	IQ
PL- $t_{(\zeta)}$	2	1.93	0.35	2.00	0.25	2.03	0.19
	4	3.41	0.71	3.67	0.54	3.83	0.41
	6	4.66	1.16	5.20	0.89	5.54	0.65
PL- $Hyp_{(\zeta)}$	1	0.94	0.04	0.97	0.03	0.98	0.02
	1.5	1.41	0.07	1.45	0.05	1.47	0.04
	2	1.87	0.10	1.93	0.07	1.96	0.05
PL- $PE_{(\zeta)}$	1.5	1.59	0.62	1.54	0.44	1.53	0.31
	2	2.08	0.98	2.05	0.73	2.04	0.52
	2.5	2.57	1.34	2.53	1.03	2.53	0.74
PL- $SN_{(\zeta)}$	1.5	1.59	0.09	1.55	0.07	1.53	0.06
	2	2.12	0.13	2.06	0.10	2.03	0.08
	3	3.17	0.20	3.10	0.16	3.05	0.13

Em geral, a mediana das estimativas de ζ obtidas pelo método proposto se aproxima do verdadeiro valor, quando o tamanho da amostra aumenta, para todos os modelos em estudo. A estimativa do viés mediano de ζ , no entanto, é pequena ainda para amostras de tamanho pequeno. Para o modelo PL- $Hyp_{(1)}$ a estimativa do viés mediano é -0.06 , para $n = 40$. Adicionalmente, notamos que a estimativa do viés mediano aumenta quando o valor de ζ aumenta, no modelo PL- $t_{(\zeta)}$, para n fixo. Por exemplo, quando $\zeta = 2$, a estimativa do viés para $n = 40$ é -0.07 e -1.34 quando $\zeta = 6$. Quando o tamanho da amostra é grande ou moderado, a estimativa do viés diminui consideravelmente.

Além disso, a variabilidade em torno da mediana tende a decrescer quando o tamanho da amostra cresce. No modelo PL- $PE_{(2.5)}$, por exemplo, a distância interquartilica é 1.34 quando $n = 40$ e 0.74 quando $n = 150$. Esse comportamento mostra, portanto, uma certa estabilidade na estimação de ζ através do método proposto. Adicionalmente, uma análise da Tabela E.2, no Apêndice E, indica que a estimação de ζ através desse método não influencia o desempenho dos estimadores de μ , σ e λ .

2.4.3 Estudo de robustez

Nesta seção, o interesse é avaliar a influência de observações atípicas nos ajustes de distribuições da classe *power logit* e comparar com outras distribuições comumente utilizadas na modelagem de dados fracionários. Consideramos as distribuições beta, Kumaraswamy e GJS. Inicialmente, geramos um conjunto de dados (y) de $n = 200$ observações e ajustamos as distribuições beta, Kumaraswamy, GJS e PL. Em seguida, 5% das observações são substituídas por observações geradas de uma distribuição beta com parâmetros $\mu = 0.9$ e $\phi = 40$, e os modelos são novamente ajustados.

Dizemos que esses novos dados estão contaminados com observações discrepantes. Consideramos duas distribuições para gerar y : beta (Cenário 1) e GJS-N (Cenário 2).

Para cada cenário, apresentamos os valores de Υ_ζ e do AIC para cada modelo, tal como o gráfico de r_i com envelopes simulados. Essas quantidades são facilmente calculadas para os modelos beta, Kumaraswamy e GJS, conforme apresentado na Seção 2.3.2. Além disso, são apresentadas as estimativas dos parâmetros de cada modelo considerando os dados sem contaminação e contaminados. A estimativa da mediana de Y (denotada por $\widehat{y}_{0.5}$) considerando o ajuste do modelo beta também é apresentada para fins de comparação; para os modelos Kumaraswamy, GJS-N e PL-N, $\widehat{y}_{0.5} = \widehat{\mu}$.

Cenário 1

Neste cenário, geramos os dados da distribuição beta com parâmetros $\mu = 0.3$ e $\phi = 10$. Na Tabela 2.5 temos algumas medidas descritivas para os dois conjuntos de dados. Observe que a média da amostra gerada sem contaminação é inferior à média populacional e aumenta quando contaminamos os dados (isso já é esperado, uma vez que a contaminação é feita com valores próximos de 1). Além disso, o valor máximo quando os dados são contaminados (0.958) é bem superior à média de y , indicando a presença de observações discrepantes.

Tabela 2.5: Medidas descritivas para os dados (contaminados e sem contaminação) gerados de acordo com o Cenário 1.

	Mínimo	Q_1	Mediana	Média	Q_3	Máximo
Dados sem contaminação	0.039	0.187	0.261	0.285	0.370	0.705
Dados contaminados	0.039	0.189	0.274	0.315	0.380	0.958

Na Tabela 2.6 são apresentadas as estimativas dos parâmetros e os valores de Υ_ζ e do AIC para cada modelo ajustado aos dois conjuntos de dados.

Tabela 2.6: Estimativas dos parâmetros (com respectivos erros-padrão), valores da medida Υ_ζ e AIC para cada modelo ajustado aos dados sem e com contaminação – Cenário 1.

	Par.	Dados sem contaminação			Dados contaminados		
		emv (ep)	Υ_ζ	AIC	emv (ep)	Υ_ζ	AIC
beta	ξ	0.285 (0.009)	0.033	-263.5	0.332 (0.013)	0.257	-128.9
	ϕ	11.308 (1.095)			5.453 (0.509)		
	$\widehat{y}_{0.5}$	0.272			0.310		
Kumaraswamy	μ	0.278 (0.010)	0.056	-260.8	0.310 (0.015)	0.275	-121.4
	σ	0.446 (0.009)			0.678 (0.013)		
GJS-N	μ	0.265 (0.010)	0.050	-261.5	0.298 (0.015)	0.224	-139.2
	σ	0.701 (0.035)			0.985 (0.049)		
PL-N	μ	0.271 (0.010)	0.032	-261.7	0.272 (0.011)	0.083	-183.7
	σ	0.515 (0.094)			3.573 (1.151)		
	λ	0.460 (0.291)			6.134 (2.062)		

Quando os dados não têm contaminação, os valores de Υ_ζ e do AIC são bem próximos, sugerindo que qualquer um dos ajustes feitos é razoável. Por simplicidade, a distribuição beta poderia ser usada. Note que, embora o valor de Υ_ζ para a distribuição PL-N seja menor que o da GJS-N, o

AIC de ambos os modelos é praticamente o mesmo. Isso é esperado, pois a medida Υ_ζ , ao contrário do AIC, não leva em consideração o número de parâmetros do modelo. Observe também que a estimativa da mediana de Y , com base no ajuste do modelo beta, é bem próxima da estimativa de μ no modelo PL-N e ambas são próximas da mediana amostral.

Por outro lado, quando os dados são contaminados, os ajustes dos modelos beta, Kumaraswamy e GJS-N são fortemente influenciados. O valor de Υ_ζ para estes modelos é aproximadamente três vezes maior do que o do modelo PL-N. O valor do AIC também sugere que o modelo PL-N é o mais indicado. A Figura 2.21 apresenta as densidades estimadas para cada modelo sobrepostas aos respectivos histogramas dos dados com e sem contaminação. Quando os dados não têm contaminação, as densidades estimadas são bem próximas. A contaminação, no entanto, faz com que a forma das densidades estimadas dos modelos beta, Kumaraswamy e GJS-N mude consideravelmente. Observe que, na tentativa de acomodar as observações extremas, as densidades estimadas destas distribuições acabam piorando o ajuste para a maior parte das observações. O ajuste PL-N, por outro lado, apresenta cauda mais pesada quando os dados estão contaminados, mas não há grandes mudanças em termos de forma para as demais observações. Essa cauda mais pesada é reflexo da mudança nas estimativas de σ e, principalmente, de λ quando os dados são contaminados. A estimativa de λ passa de 0.460 (dados originais) para 6.134 (dados contaminados). Esse comportamento é esperado, uma vez que a assimetria da distribuição deve mudar para que o ajuste acomode bem as observações discrepantes. A estimativa de μ não sofre grande influência, o que é desejável. Já as estimativas do modelo beta são bastante influenciadas por essas observações discrepantes e, ainda assim, a distribuição não parece acomodar bem os dados. A estimativa de ξ , para este modelo, aumenta de 0.285 para 0.332 e a estimativa do parâmetro de precisão ϕ diminui consideravelmente (de 11.308 para 5.453). A estimativa da mediana (μ) do modelo Kumaraswamy também é relativamente influenciada.

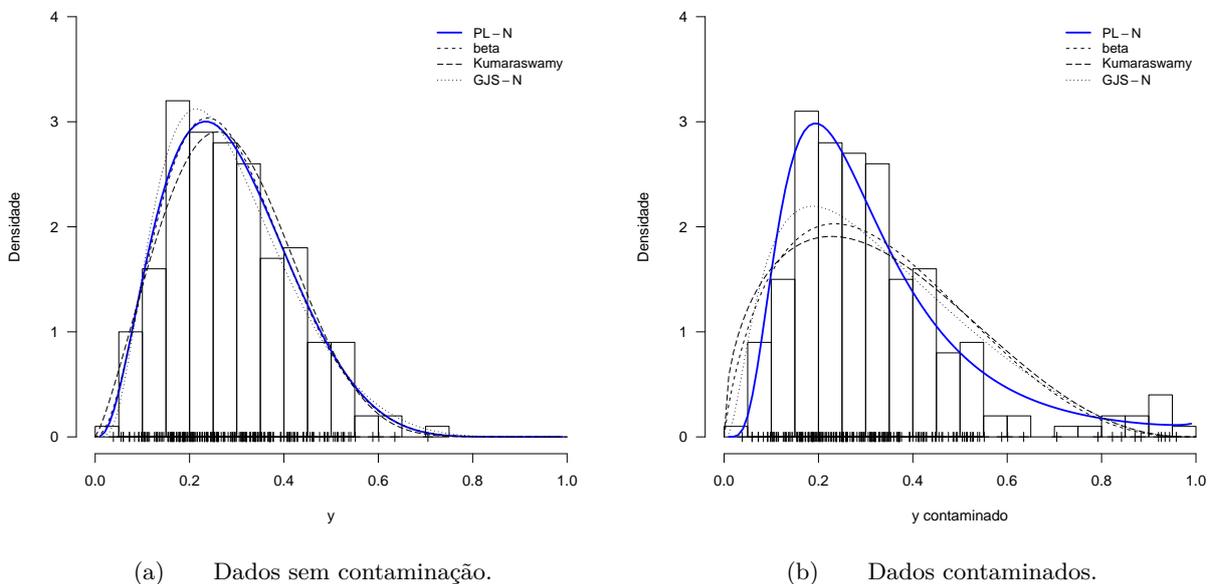
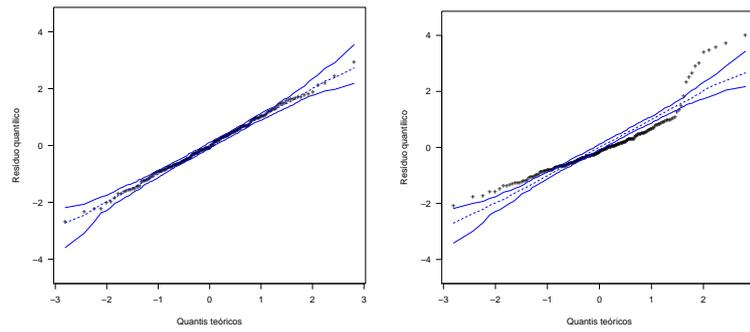


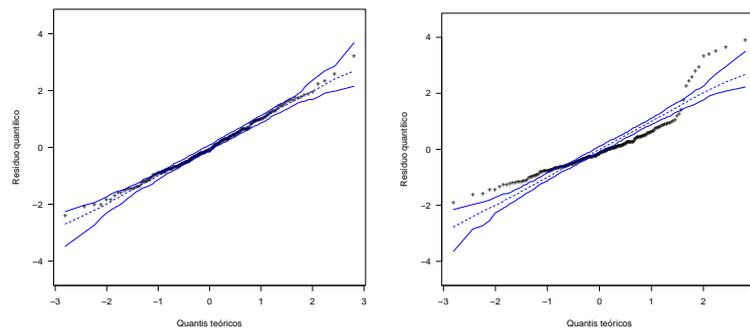
Figura 2.21: Histograma para os dados sem contaminação (y) e com contaminação (y contaminado) com densidades estimadas considerando os modelos beta, Kumaraswamy, GJS-N e PL-N – Cenário 1.

A Figura 2.22 apresenta os gráficos do resíduo r_i com envelopes simulados para os quatro modelos ajustados aos dados originais e com contaminação.

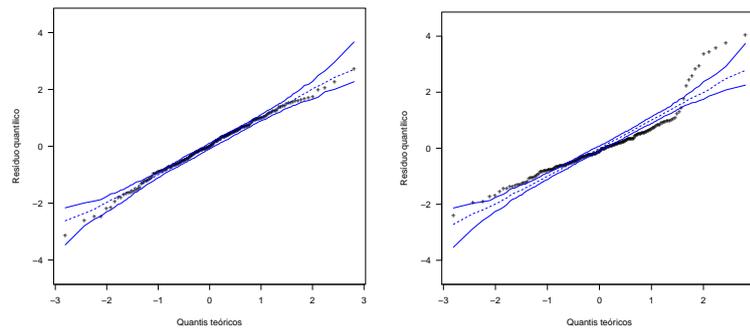
Conforme podemos observar, os ajustes dos modelos beta, Kumaraswamy e GJS-N são fortemente comprometidos pelas observações atípicas, identificadas, através desses gráficos, como *outliers*. Os resíduos indicam que essas distribuições não acomodam bem os *outliers* e, além disso,



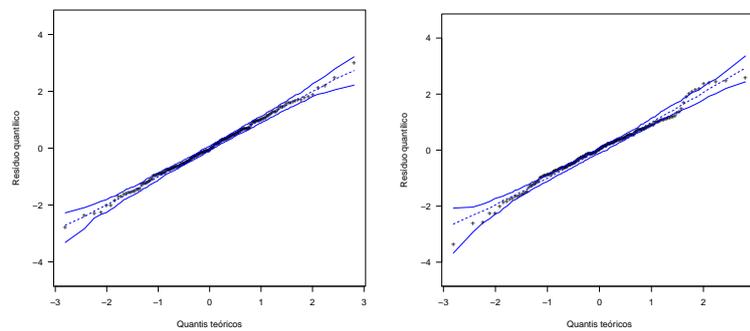
(a) Distribuição beta.



(b) Distribuição Kumaraswamy.



(c) Distribuição GJS-N.



(d) Distribuição PL-N.

Figura 2.22: Gráfico de r_i , $i = 1, \dots, 200$, com envelopes simulados referentes aos ajustes dos modelos beta, Kumaraswamy, GJS-N e PL-N, considerando os dados sem contaminação (gráficos da esquerda) e com contaminação (gráficos da direita) – Cenário 1.

a presença deles faz com que o ajuste às demais observações fique comprometido. A distribuição PL-N, no entanto, consegue se adequar à presença dos *outliers* e o ajuste do modelo continua satisfatório. Isso evidencia, portanto, que o ajuste da distribuição PL-N é menos sensível a observações discrepantes, no sentido de conseguir manter um bom ajuste para a maioria das observações e ainda acomodar as observações discrepantes.

No que segue, esse experimento foi repetido 100 vezes e os valores da medida Υ_ζ foram obtidos para cada ajuste (com e sem contaminação). A Figura 2.23 apresenta os gráficos dos valores de Υ_ζ para cada amostra gerada. Conforme podemos observar, se os dados não estão contaminados, os valores Υ_ζ são próximos para os modelos GJS-N e PL-N, e são um pouco maiores, em geral, para os modelos beta e Kumaraswamy. Por outro lado, se os dados estão contaminados, os ajustes das distribuições beta, Kumaraswamy e GJS-N retornam valores de Υ_ζ muito maiores do que do ajuste da distribuição PL-N, evidenciando, portanto, que o ajuste dessas distribuições é bastante comprometido pela presença de observações discrepantes. Note que a distribuição GJS-N, embora seja mais flexível do que a beta e a Kumaraswamy, não consegue acomodar de forma razoável essas observações. O ajuste da distribuição PL-N, por outro, não parece sofrer grande influência dessas observações. As médias dos valores de Υ_ζ para as distribuições beta, Kumaraswamy, GJS-N e PL-N, considerando os dados contaminados, foi 0.230, 0.250, 0.210 e 0.107, respectivamente.

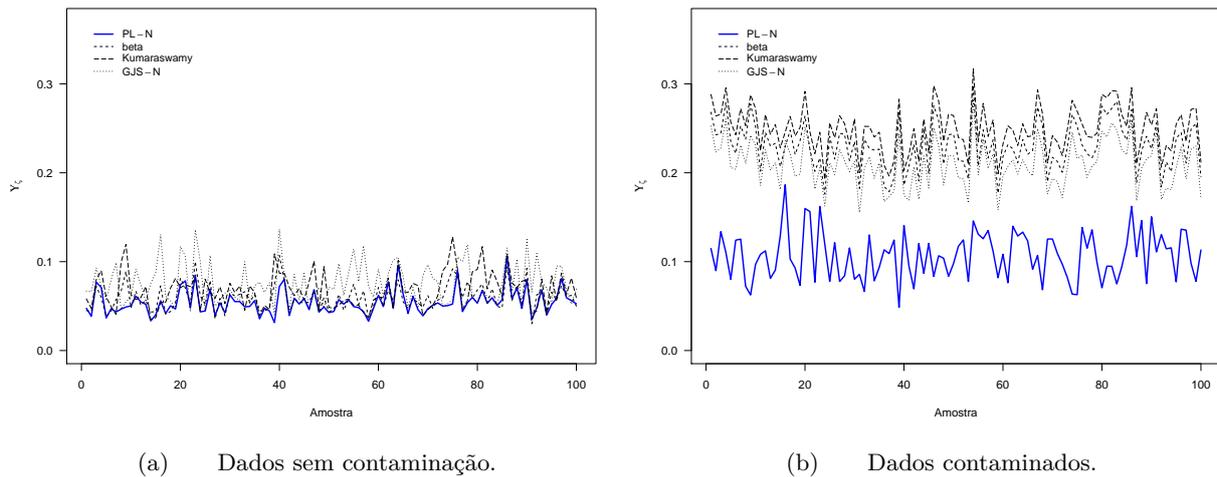


Figura 2.23: Valores de Υ_ζ para os ajustes das distribuições beta, Kumaraswamy, GJS e PL-N para cada uma das 100 amostras geradas de acordo com o Cenário 1.

Finalmente, com o objetivo de verificar o quanto o número de observações contaminadas influencia no ajuste dos modelos em discussão, repetimos o experimento apresentado neste cenário alterando o número de observações contaminadas. Foram contaminadas $k = (0, 1, 3, 5, 7, 10, 20)$ observações na amostra em estudo. A Figura 2.24 apresenta o valor de Υ_ζ para cada contaminação feita.

Com base na Figura 2.24, notamos que a amostra estudada anteriormente é obtida quando $k = 0$ (amostra sem contaminação) e $k = 10$, ou seja, 5% das observações é contaminadas (amostra contaminada). Essa figura mostra que, salvo a distribuição PL-N, todas as outras são fortemente influenciadas pelo número de observações contaminadas na amostra. Quando contaminamos 1% e 2% das observações, os valores de Υ_ζ para todos os ajustes são relativamente próximos, mas quando a proporção de observações contaminadas aumenta, o valor de Υ_ζ para os modelos beta, Kumaraswamy e PL-N tende a crescer, enquanto que o da PL-N apresenta um comportamento constante.

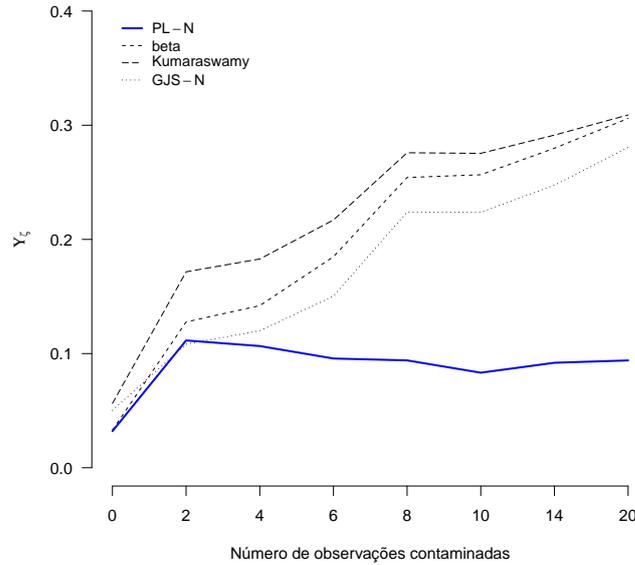


Figura 2.24: Valores de Υ_ζ para os ajustes das distribuição PL-N, beta, Kumaraswamy e GJS para cada uma das 8 amostras geradas de acordo com o Cenário 1, com a porcentagem de observações contaminadas igual a 0%, 1%, 2%, 3%, 4%, 5%, 7% e 10%.

Cenário 2

Neste cenário, geramos $n = 200$ observações de uma distribuição GJS com parâmetros $\mu = 0.2$ e $\sigma = 1$. Algumas medidas resumo para os dados originais e contaminados são apresentadas na Tabela 2.7.

Tabela 2.7: Medidas descritivas para os dados (contaminados e sem contaminação) gerados de acordo com o Cenário 2.

	Mínimo	Q_1	Mediana	Média	Q_3	Máximo
Dados sem contaminação	0.015	0.104	0.179	0.225	0.308	0.843
Dados contaminados	0.015	0.111	0.184	0.256	0.318	0.964

A Tabela 2.8 apresenta as estimativas dos parâmetros dos modelos e respectivos erros-padrão, os valores de Υ_ζ e o AIC para os ajustes considerados. Conforme podemos observar, quando os dados não são contaminados, os modelos GJS-N e PL-N apresentam o melhor ajuste (menor valor de Υ_ζ e menor AIC). As estimativas dos parâmetros dessas distribuições são bem próximas e o obtivemos $\tilde{\lambda} = 0.962$, bem próximo de 1, indicando que a distribuição GJS-N deve ser considerada. Os modelos beta e Kumaraswamy apresentam valores próximos de Υ_ζ e do AIC. Observe que as estimativas da mediana de Y para todos os ajustes são bem próximas do verdadeiro valor (0.2).

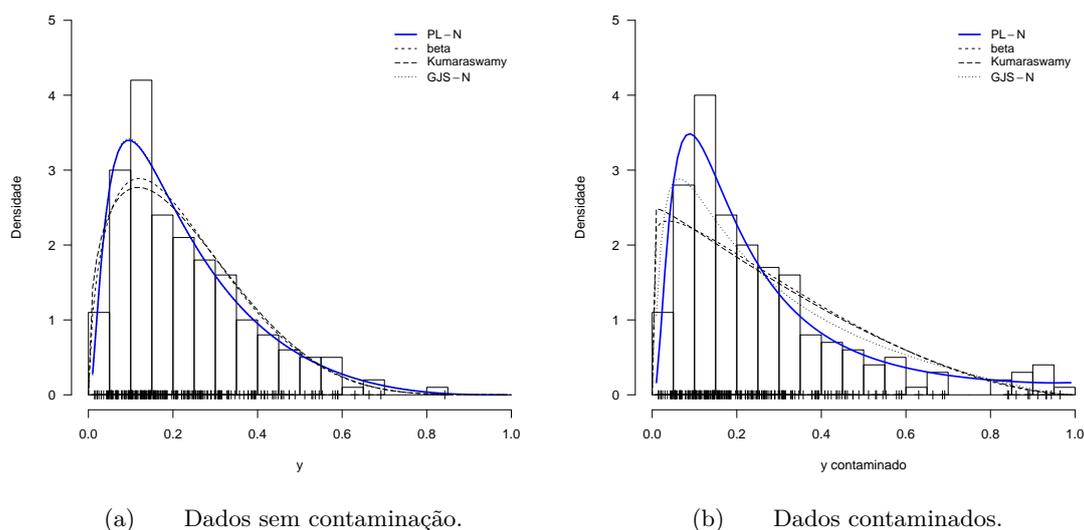
Por outro lado, quando os dados são contaminados, a distribuição PL-N é a que apresenta menores valores de Υ_ζ e de AIC. Novamente, a estimativa de μ no caso da PL-N não sofre grande influência das observações discrepantes, e a estimativa de λ se afasta do valor 1, indicando que o modelo GJS-N não é razoável. As estimativas dos parâmetros dos modelos beta, Kumaraswamy e GJS-N sofrem, na maioria dos casos, grande influência das observações atípicas. Os valores de Υ_ζ para os modelos beta e Kumaraswamy são quatro vezes o valor para o modelo PL-N, evidenciando, portanto, que o modelo PL-N é o que melhor se ajusta aos dados contaminados e que os demais modelos são consideravelmente influenciados pelas observações discrepantes.

Os histogramas dos dados originais e contaminados com as densidades estimadas sobrepostas são apresentados na Figura 2.25. Observe que, quando os dados não são contaminados, as densidades

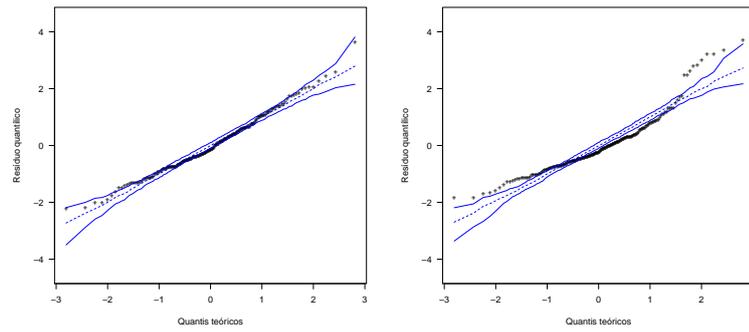
Tabela 2.8: Estimativas dos parâmetros (com respectivos erros-padrão), valores da medida Υ_ζ e AIC para cada modelo ajustado aos dados sem e com contaminação – Cenário 2.

	Par.	Dados sem contaminação			Dados contaminados		
		emv (ep)	Υ_ζ	AIC	emv (ep)	Υ_ζ	AIC
beta	ξ	0.228 (0.010)	0.095	-239.0	0.279 (0.014)	0.241	-134.2
	ϕ	7.009 (0.679)			3.837 (0.359)		
	$\widehat{y}_{0.5}$	0.201			0.238		
Kumaraswamy	μ	0.203 (0.011)	0.108	-235.5	0.234 (0.016)	0.240	-133.7
	σ	0.730 (0.004)			0.986 (0.007)		
GJS-N	μ	0.187 (0.011)	0.043	-246.2	0.215 (0.015)	0.171	-158.1
	σ	0.985 (0.049)			1.265 (0.063)		
PL-N	μ	0.188 (0.011)	0.043	-244.2	0.191 (0.012)	0.060	-182.7
	σ	0.963 (0.187)			3.053 (0.825)		
	λ	0.962 (0.311)			3.469 (1.026)		

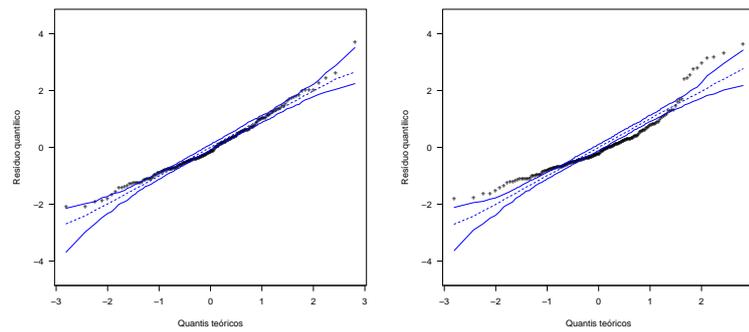
PL-N e GJS-N estimadas são praticamente idênticas. A contaminação faz com que o ajuste do modelo beta e Kumaraswamy fique fortemente comprometido. O modelo PL-N é o que aparenta melhor acomodar as observações, inclusive a cauda da densidade estimada é bem mais pesada do que a dos demais modelos.

**Figura 2.25:** Histograma para os dados sem contaminação (y) e com contaminação (y contaminado) com densidades estimadas considerando os modelos beta, Kumaraswamy, GJS-N e PL-N – Cenário 2.

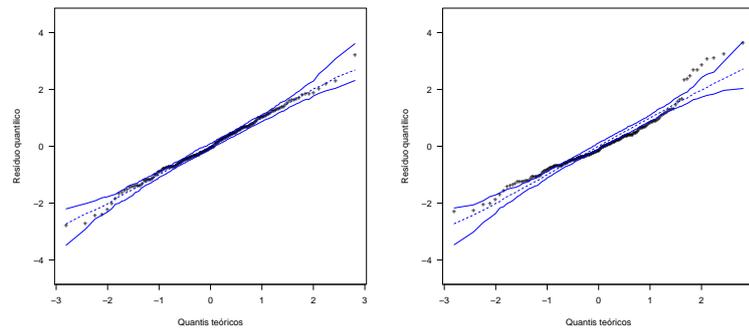
Esse comportamento é observado, também, através dos gráficos de resíduos com envelopes simulados, apresentados na Figura 2.26. Note que as bandas de confiança para os modelos beta, Kumaraswamy e GJS-N não conseguem acomodar bem todas as observações; ao contrário do que acontece com o modelo PL-N.



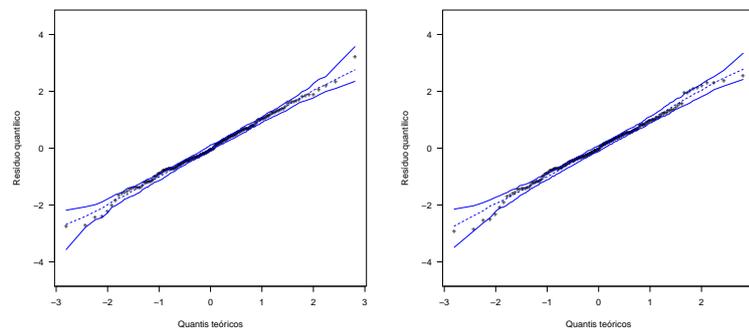
(a) Distribuição beta.



(b) Distribuição Kumaraswamy.



(c) Distribuição GJS-N.



(d) Distribuição PL-N.

Figura 2.26: Gráfico de r_i , $i = 1, \dots, 200$, com envelopes simulados referentes aos ajustes dos modelos beta, Kumaraswamy, GJS-N e PL-N, considerando os dados sem contaminação (gráficos da esquerda) e com contaminação (gráficos da direita) – Cenário 2.

Repetindo esse experimento 100 vezes, os valores da medida Υ_ζ para cada amostra gerada (com e sem contaminação) são apresentados na Figura 2.27. Conforme podemos observar, se os dados não estão contaminados, os valores Υ_ζ são próximos para os modelos GJS-N e PL-N, e são maiores, em geral, para os modelos beta e Kumaraswamy. Neste caso, o modelo GJS-N poderia ser escolhido, uma vez que tem menos parâmetros. Por outro lado, se os dados estão contaminados, o ajuste das distribuições beta, Kumaraswamy e GJS-N retornam valores de Υ_ζ muito maiores do que o ajuste da distribuição PL-N. Evidenciando, mais uma vez, a sensibilidade dessas distribuições a observações atípicas. A distribuição GJS-N, embora mais seja mais flexível do que beta e a Kumaraswamy, não consegue acomodar de forma razoável observações discrepantes. Por outro lado, o ajuste da distribuição PL-N parece não sofrer grande influência dessas observações.

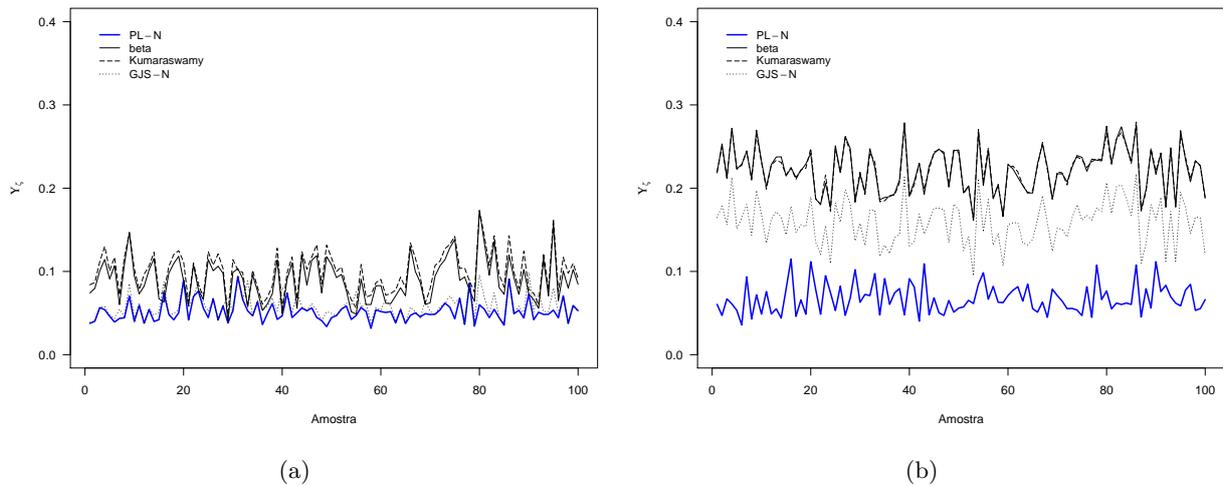


Figura 2.27: Valores de Υ_ζ para os ajustes das distribuições beta, Kumaraswamy, GJS-N e PL-N para cada uma das 100 amostras geradas de acordo com o Cenário 2.

Finalmente, consideramos o experimento apresentado neste cenário alterando o número de observações contaminadas. Contaminamos $k = (0, 1, 3, 5, 7, 10, 20)$ observações na amostra. A Figura 2.28 apresenta o valor de Υ_ζ para cada uma das 7 amostras. Observe que, novamente, o ajuste da distribuição PL-N não é influenciado pelo número de observações contaminadas. Os modelos beta, Kumaraswamy e GJS-N, por outro lado, tendem a apresentar um pior ajuste quando o número de observações contaminadas aumenta.

A partir dessas aplicações a dados simulados, verificamos que, para qualquer cenário com *outliers*, o ajuste do modelo PL-N é superior aos demais modelos em estudo. Esse resultado foi obtido em todos os cenários observados e à medida que o número de observações contaminadas aumenta, essa conclusão é mais evidente. Dessa forma, o modelo proposto é mais indicado nestas situações.

2.5 Aplicações

Nesta seção, ilustramos o uso da metodologia desenvolvida neste capítulo mediante aplicações a conjuntos de dados reais. Para efeito de comparação, consideramos outras distribuições para modelagem de dados no intervalo unitário, a saber: as distribuições beta, Kumaraswamy, simplex e GJS. Os parâmetros dessas distribuições são estimados via método da máxima verossimilhança e os ajustes foram feitos no *software* R. Para ajuste do modelo beta, utilizamos o pacote `betareg`, proposto por Zeileis *et al.* (2020); para ajuste dos modelos simplex e GJS, utilizamos o pacote `gamlss` (Stasinopoulos *et al.*, 2022); e, finalmente, para ajuste das distribuições PL e Kumaraswamy, utilizamos a função `optim` do R. Para verificar a adequação dos modelos em estudo, utilizamos a medida Υ_ζ e os resíduos $r_i = \Phi^{-1}[R(\hat{z}^{(i)})]$.

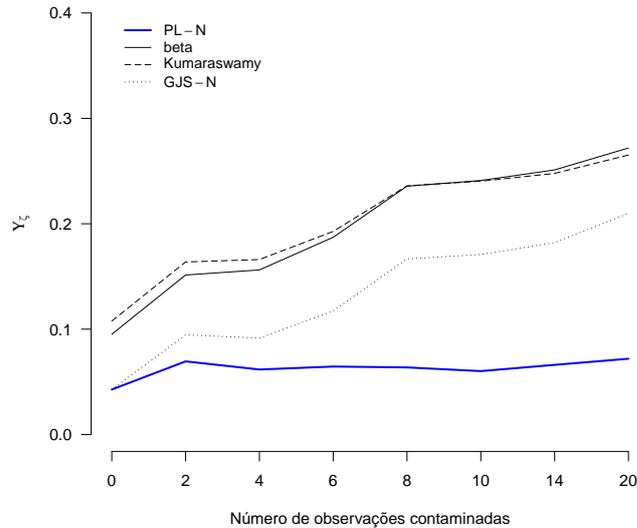


Figura 2.28: Valores de Υ_ζ para os ajustes das distribuição PL-N, beta, Kumaraswamy e GJS para cada uma das 8 amostras geradas de acordo com o Cenário 2, com a porcentagem de observações contaminadas igual a 0%, 1%, 2%, 3%, 4%, 5%, 7% e 10%.

Porcentagem de gordura corporal de morcegos

Nesta primeira aplicação, analisamos os dados de gordura corporal de morcegos coletados em seis locais do nordeste dos Estados Unidos durante o inverno de 2009 (que abrange a estação de hibernação de outubro de 2008 a abril de 2009) e de 2016 (outubro de 2015 a abril 2016). Os locais de coleta foram *Aeolus Cave*, *Hibernia Mine*, *Graphite Mine*, *Williams Preserve Mine*, *Barton Hill Mine*, *Williams Hotel Mine*. Esses dados são estudados em Cheng *et al.* (2019). Os autores investigam se a persistência da síndrome do nariz branco (*white-nose syndrome*) em uma espécie de pequenos morcegos marrons (*M. lucifigus*) pode ser explicada pelo aumento da gordura corporal no início do inverno, o que permitiria aos morcegos tolerar o aumento dos custos energéticos associados à síndrome. Para essa análise, consideramos os dados referentes ao inverno de 2009 e nosso interesse é modelar a proporção de gordura corporal (y) de 342 morcegos coletados nesse período. A Tabela 2.9 apresenta algumas medidas resumo para a variável y e o boxplot robusto (Hubert e Vandervieren, 2008) dessa variável é apresentado na Figura 2.29.

Mínimo	0.025
Q_1	0.084
Mediana	0.123
Média	0.132
Q_3	0.175
Máximo	0.301

Tabela 2.9: Medidas resumo para a porcentagem de gordura corporal de morcegos (y).

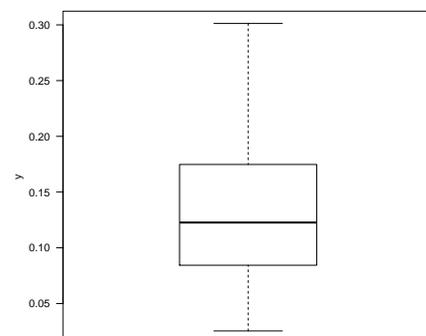


Figura 2.29: Boxplot para a variável y .

Observe que a distribuição dos dados é assimétrica e está bastante concentrada em valores

próximos de zero (25% dos valores estão abaixo de 0.084). O boxplot não identifica nenhuma observação atípica.

A Tabela 2.10 apresenta os valores das estimativas de máxima verossimilhança para os parâmetros de alguns modelos ajustados aos dados de porcentagem de gordura corporal de morcegos. Os parâmetros das distribuições *power logit* foram estimados via máxima verossimilhança penalizada e a estimativa para ζ foi obtida através do método apresentado na Seção 2.3.2. Também nesta tabela, são fornecidos intervalos de 95% confiança (obtidos através da distribuição assintótica do estimador de máxima verossimilhança) e os valores da medida Υ_ζ para cada modelo. Como a estimativa de λ para o modelo PL-PE $_{(\zeta)}$ foi relativamente baixa, sugerindo que o modelo limite poderia ser razoável, foram ajustadas os modelos LogLog-N e LogLog-PE (isto é, as distribuições normal e exponencial potência (PE) à variável $\log(-\log(Y))$).

Tabela 2.10: Estimativas de máxima verossimilhança, intervalos de 95% confiança e valores de Υ_ζ para cada modelo ajustado aos dados de porcentagem de gordura corporal de morcegos.

Distribuição	Parâmetro	Estimativa	IC(95%)	Υ_ζ
beta	ξ	0.132	(0.126, 0.138)	0.075
	ϕ	32.252	(27.420, 37.084)	
Kumaraswamy	μ	0.128	(0.122, 0.135)	0.082
	σ	0.412	(0.403, 0.421)	
simplex	ξ	0.131	(0.124, 0.138)	0.087
	σ	1.702	(1.574, 1.830)	
GJS-N	μ	0.120	(0.114, 0.126)	0.083
	σ	0.546	(0.505, 0.587)	
GJS-PE $_{(3.1)}$	μ	0.117	(0.111, 0.123)	0.092
	σ	0.546	(0.512, 0.580)	
	ζ	3.1	–	
PL-N	μ	0.124	(0.117, 0.131)	0.073
	σ	0.260	(0.141, 0.379)	
	λ	0.138	(0, 0.584)	
PL-PE $_{(3.4)}$	μ	0.125	(0.118, 0.132)	0.023
	σ	0.235	(0.135, 0.335)	
	λ	0.040	(0, 0.442)	
	ζ	3.4	–	
LogLog-N	μ	0.125	(0.119, 0.131)	0.075
	σ	0.225	(0.209, 0.242)	
LogLog-PE $_{(3.5)}$	μ	0.125	(0.120, 0.131)	0.025
	σ	0.225	(0.212, 0.239)	
	ζ	3.5	–	

Note que a estimativa da mediana de y , considerando o ajuste da distribuição PL-N e PL-PE $_{(3.4)}$, é bem próxima da mediana amostral. Também, a estimativa da média de Y , considerando

o ajuste da distribuição beta, é igual à média amostral. Além disso, as estimativas de ζ para os modelos GJS-PE $_{(\zeta)}$ e PL-PE $_{(\zeta)}$ apresentaram valores próximos, o que é razoável. Adicionalmente, note que os intervalos de confiança de 95% de confiança, para todos os modelos da classe *power logit* ajustados, não incluem o valor 1, indicando a rejeição da hipótese $\mathcal{H}_0 : \lambda = 1$ (ao nível de significância de 5%). Ou seja, o ajuste da distribuição GJS não parece ser razoável.

Com base no critério de minimização da estatística Υ_{ζ} , o modelo que produziu o melhor ajuste foi o PL-PE $_{(3.4)}$ com $\Upsilon_{3.4} = 0.023$ e o modelo limite, PE $_{(3.5)}$, com $\Upsilon_{3.5} = 0.025$. Esse critério também sugere que a distribuição GJS-PE $_{(3.1)}$ foi a que apresentou o pior ajuste. A Figura 2.30 apresenta os gráficos de resíduo com envelopes simulados para cada ajuste. Note que os envelopes simulados, considerando a distribuição PL-PE $_{(3.4)}$ e o modelo limite LogLog-PE $_{(3.5)}$, acomodam bem os valores de $\Phi^{-1}[R(\hat{z}^{(i)})]$ (ver Figuras 2.30(g) e 2.30(i)).

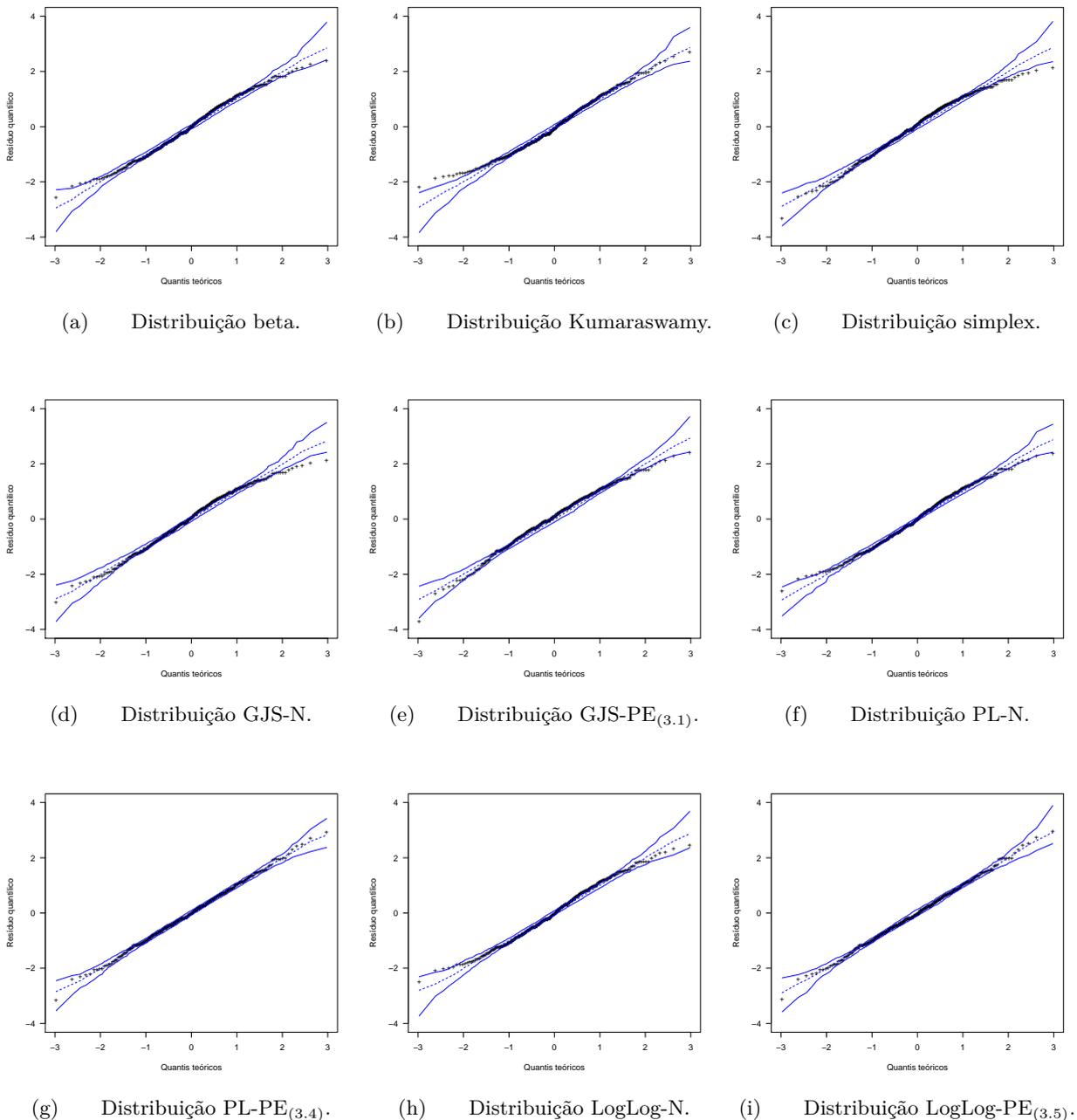


Figura 2.30: Gráficos de resíduos com envelopes simulados para os ajustes apresentados na Tabela 2.10.

A Figura 2.31 apresenta o gráfico de Υ_{ζ} e da função de log-verossimilhança perfilada relativa de

λ para o modelo PL-PE $_{(\zeta)}$. A linha horizontal pontilhada no gráfico da log-verossimilhança perfilada relativa representa um intervalo de confiança de 95% para λ a partir da quantidade $W_p^*(\lambda)$, este intervalo é (0, 0.49). Note que as funções de log-verossimilhança perfilada usual e penalizada para λ não diferem muito, conforme esperado, uma vez que temos um tamanho de amostra relativamente grande.

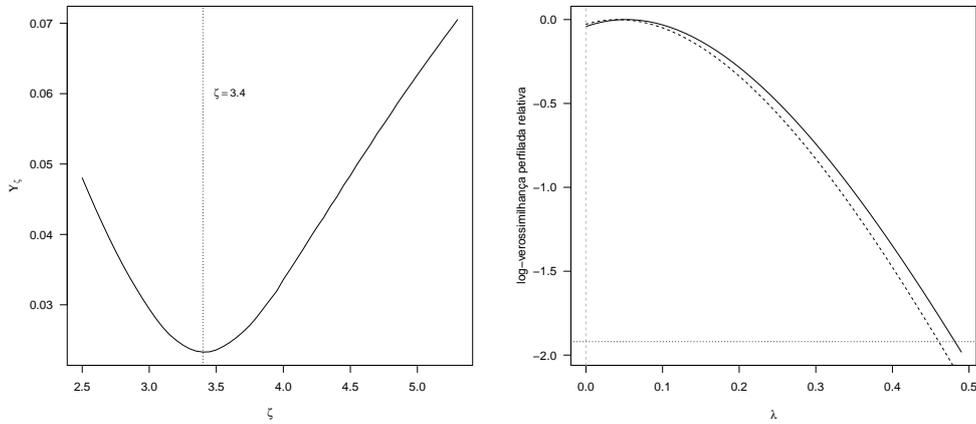


Figura 2.31: Gráfico de Υ_ζ e as versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo PL-PE $_{(\zeta)}$.

As funções densidade de probabilidade e de distribuição ajustadas para alguns dos modelos considerados podem ser visualizadas nas Figuras 2.32(a) e 2.32(b). Note que as densidades estimadas para os modelos PL-PE $_{(3.4)}$ e para o modelo limite LogLog-PE $_{(3.5)}$ são praticamente idênticas, e foram as que melhor acomodaram as observações.

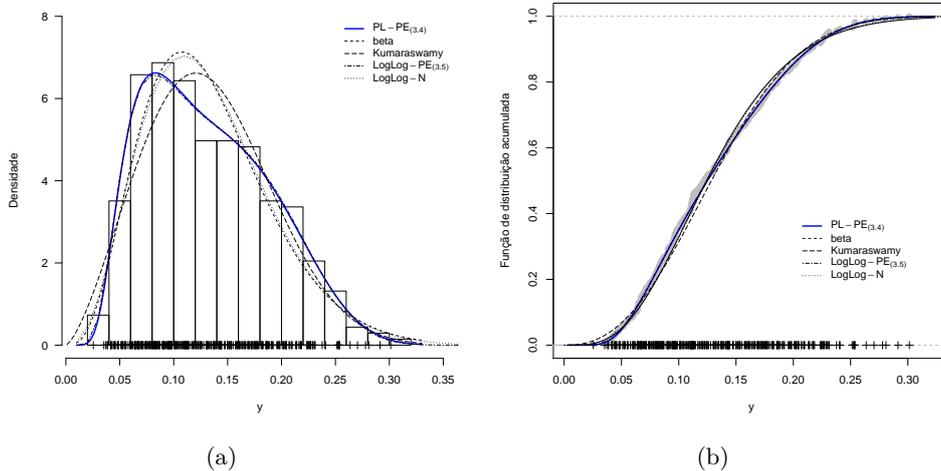


Figura 2.32: Histograma com densidades estimadas sobrepostas (a) e funções densidade de probabilidade empíricas e estimadas (b) para os dados de porcentagem de gordura corporal de morcegos.

Na Figura 2.33 é apresentado o gráfico da discrepância quantílica relativa para os modelos beta, Kumaraswamy, simplex, GJS-PE $_{(3.1)}$ e PL-PE $_{(3.4)}$. Aqui, a discrepância quantílica relativa (DQR_p) é dada por

$$DQR_p = \frac{\omega_p - \vartheta_p}{\vartheta_p},$$

em que ω_p é o quantil de ordem p da distribuição ajustada (no caso das distribuições *power logit*,

$\omega_p = \widehat{y}_p$) e ϑ_p é o quantil empírico de ordem p da variável y . Note que valores de DQR_p próximos de zero indicam que a distribuição ajustada é próxima da distribuição empírica dos dados. Na Figura 2.33, notamos que, em geral, dentre os modelos em estudo, os quantis da distribuição PL-PE_(3,4) são mais próximos dos quantis empíricos da distribuição de Y .

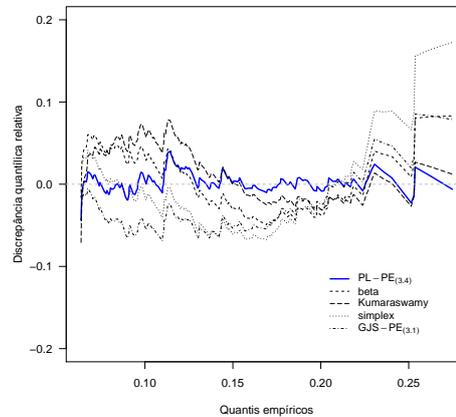


Figura 2.33: Gráfico da distância quantílica relativa para os dados de porcentagem de gordura corporal de morcegos.

Proporção de indivíduos ocupados fora do setor agropecuário

No que segue, consideramos dados das Nações Unidas obtidos através do *Atlas do Desenvolvimento Humano no Brasil*, disponíveis em <https://www.pnud.org.br/atlas/>. Esse atlas possibilita a consulta, em diversos formatos de dados, sobre o panorama do desenvolvimento humano e da desigualdade interna em diferentes territorialidades (estado, região metropolitana, município, etc). Esses índices são úteis para sinalizar o acesso, a ausência ou a insuficiência de recursos que deveriam estar à disposição de todo cidadão. Nesta análise, consideramos dados coletados em 2010 referentes a 200 municípios selecionados aleatoriamente do estado de São Paulo.

A variável de interesse (y) é a proporção de indivíduos ocupados (com mais de 18 anos e que trabalham) fora do setor agropecuário. Essa variável é a razão entre o número de pessoas de 18 anos ou mais de idade ocupadas em um setor diferente do agropecuário e o número total de pessoas ocupadas nessa faixa etária. Algumas medidas descritivas da variável y são apresentadas na Tabela 2.11 e a Figura 2.34 apresenta o boxplot robusto para essa variável.

Mínimo	0.446
Q_1	0.711
Mediana	0.814
Média	0.835
Q_3	0.945
Máximo	0.998

Tabela 2.11: Medidas resumo para a proporção de indivíduos ocupados fora do setor agropecuário (y).

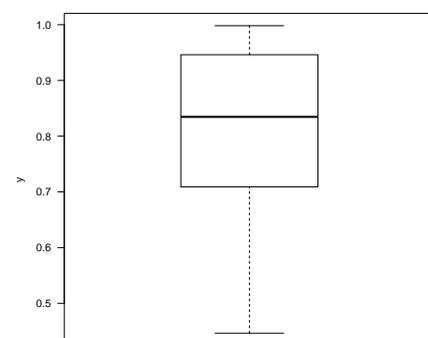


Figura 2.34: Boxplot para a variável y .

São considerados como formalmente ocupados os empregados com carteira de trabalho assi-

nada, os militares do exército, da marinha, da aeronáutica, da polícia militar ou do corpo de bombeiros, os empregados pelo regime jurídico dos funcionários públicos, assim como os empregadores e trabalhadores por conta própria que eram contribuintes de instituto de previdência oficial. Conforme podemos observar, dentre as cidades selecionadas, a proporção de indivíduos ocupados fora do setor agropecuário está concentrada entre 0.711 e 0.945. Ou seja, a maioria das pessoas (com idade superior a 18 anos) nessas cidades está ocupada em setores diferentes do agropecuário. De fato, segundo a *Wikipédia*, o setor primário é o menos relevante para a economia paulista: em 2011, a agropecuária representava somente 2,11% do valor total adicionado à economia de todo o estado. As cidades com maior proporção de indivíduos ocupados no setor agropecuário são Altair e Ubirajara. Ambas estão localizadas no interior do estado e a economia é, predominantemente, baseada na agricultura. Dentre os municípios selecionados, o que tem o menor número de pessoas trabalhando no setor agropecuário, na amostra selecionada, é Carapicuíba (que está localizada na região metropolitana de São Paulo).

Os modelos beta, Kumaraswamy, simplex, GJS-N, GJS-PE_(ζ), PL-N e PL-PE_(ζ) foram ajustados aos dados (outros modelos da classe GJS e PL foram considerados, mas apresentaram desempenho igual ou inferior aos modelos mencionados e, portanto, não serão apresentados). A Tabela 2.12 apresenta os resultados dos ajustes.

Tabela 2.12: Estimativas de máxima verossimilhança, intervalos de 95% de confiança e valores de Υ_ζ para cada modelo ajustado aos dados de proporção de indivíduos ocupados fora do setor agropecuário.

Distribuição	Parâmetro	Estimativa	IC(95%)	Υ_ζ
beta	ξ	0.818	(0.797, 0.839)	0.151
	ϕ	5.427	(4.369, 6.485)	
Kumaraswamy	μ	0.858	(0.838, 0.879)	0.150
	σ	0.226	(-0.793, 1.246)	
simplex	ξ	0.853	(0.847, 0.858)	0.417
	σ	5.598	(4.994, 6.202)	
GJS-N	μ	0.878	(0.857, 0.899)	0.220
	σ	1.420	(1.281, 1.559)	
GJS-PE _(2.7)	μ	0.897	(0.868, 0.925)	0.248
	σ	1.425	(1.302, 1.548)	
	ζ	2.7	–	
PL-N	μ	0.837	(0.810, 0.864)	0.069
	σ	2.406	(1.875, 2.937)	
	λ	9.269	(5.979, 12.559)	
PL-PE _(2.7)	μ	0.840	(0.811, 0.869)	0.054
	σ	2.364	(1.870, 2.859)	
	λ	9.030	(5.891, 12.169)	
	ζ	2.7	–	

O ajuste que resultou no menor valor de Υ_ζ foi o PL-PE_(2.7), seguido da distribuição PL-N. A

distribuição que resultou no pior ajuste foi a simplex ($\Upsilon_\zeta = 0.417$). Todos os modelos considerados apresentaram ajustes inferiores aos dos modelos da classe *power logit*. O valor de Υ_ζ do modelo GJS-N é quase quatro vezes o do modelo PL-N. Além disso, os intervalos de 95% de confiança para λ , considerando os modelos PL-N e PL-PE_(2.7), não incluem o valor 1, indicando que o valor de λ é estatisticamente diferente de 1 (ao nível de significância de 5%); rejeitando, portanto, as hipóteses de que os modelos GJS-N e GJS-PE_(2.7) são razoáveis. Adicionalmente, os valores de Υ_ζ para os ajustes beta e Kumaraswamy são praticamente iguais, indicando ajustes similares.

Os gráficos de resíduos com envelopes simulados para cada um dos ajustes apresentados na Tabela 2.12 são dados na Figura 2.35.

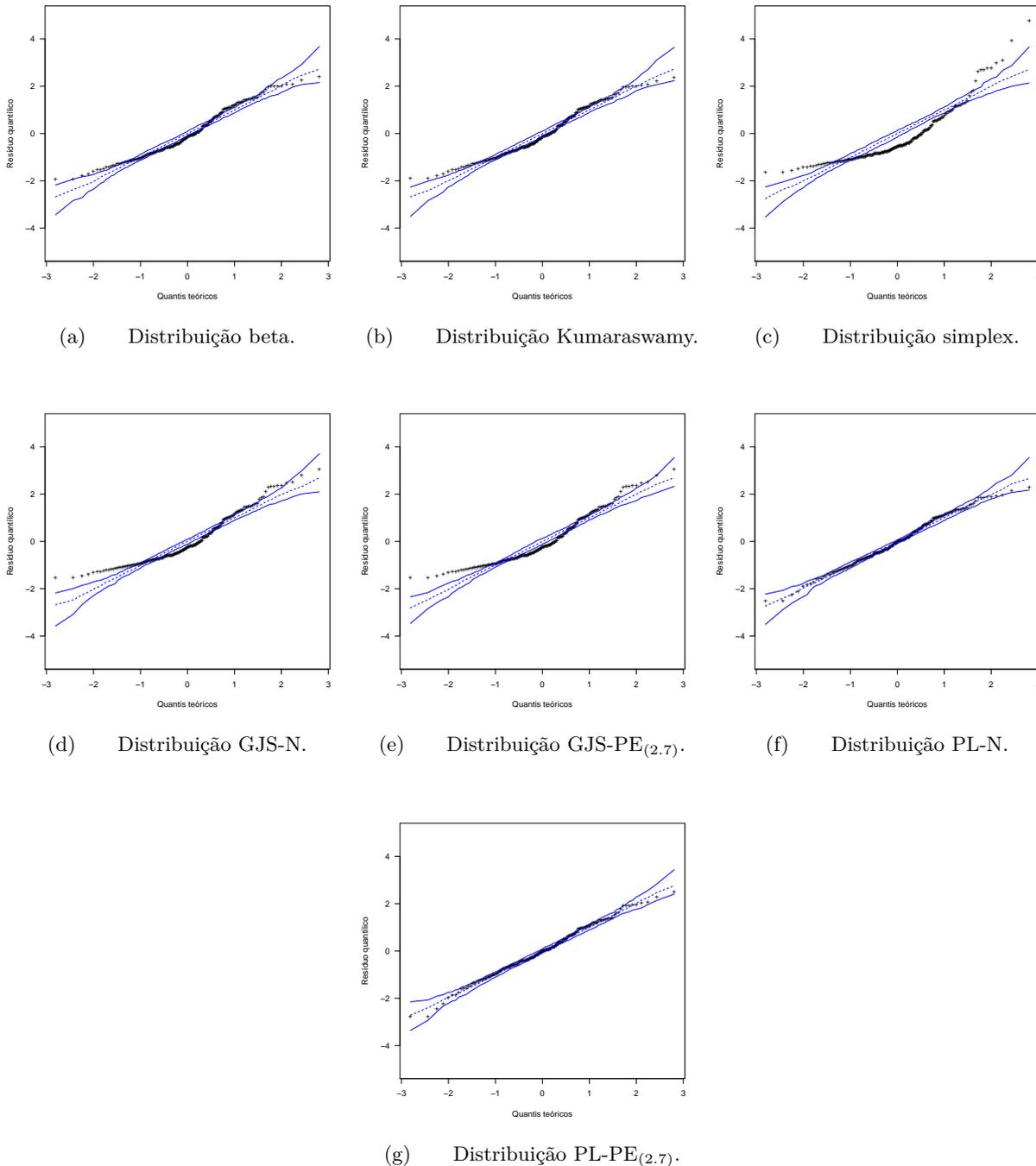


Figura 2.35: Gráficos de resíduo com envelopes simulados para os ajustes apresentados na Tabela 2.12.

Observe que, salvo para os modelos PL, os envelopes simulados não conseguem acomodar a

maioria dos resíduos, indicando falta de ajuste aos dados. Esses gráficos confirmam que o modelo com pior ajuste é o simplex e que os ajustes dos modelos beta e Kumaraswamy são parecidos. Os envelopes simulados para os modelos PL-N e PL-PE_(2.7), por outro lado, acomodam a grande maioria dos resíduos e os ajustes são satisfatórios. Considerando o critério de menor Υ_ζ , o modelo PL-PE_(2.7) deveria ser o escolhido. Por outro lado, o modelo PL-N é mais simples (tem menos parâmetros) e também apresentou baixo valor de Υ_ζ e resíduos com comportamento desejado. Assim, o modelo PL-N é o mais indicado para a análise. Com base neste modelo, a mediana da proporção de indivíduos ocupados fora do setor agropecuário nas cidades do estado de São Paulo é 0.837. Isso indica que em 50% das cidades deste estado, a proporção de pessoas que trabalham fora do setor agropecuário é igual ou superior a 83.7%. O intervalo de confiança observado foi (0.810, 0.864).

Conforme observamos, os dados em estudo não parecem ser bem acomodados pelas distribuições beta, Kumaraswamy, simplex e GJS. No entanto, as distribuições PL, que são mais flexíveis, conseguem acomodar bem a forma desses dados. Isso é confirmado através dos gráficos apresentados na Figura 2.36. Nesta figura são apresentadas as funções densidade de probabilidade ajustas sobrepostas ao histograma de y , funções de distribuição empírica e ajustadas, e gráfico da distância quantílica relativa para os modelos ajustados.

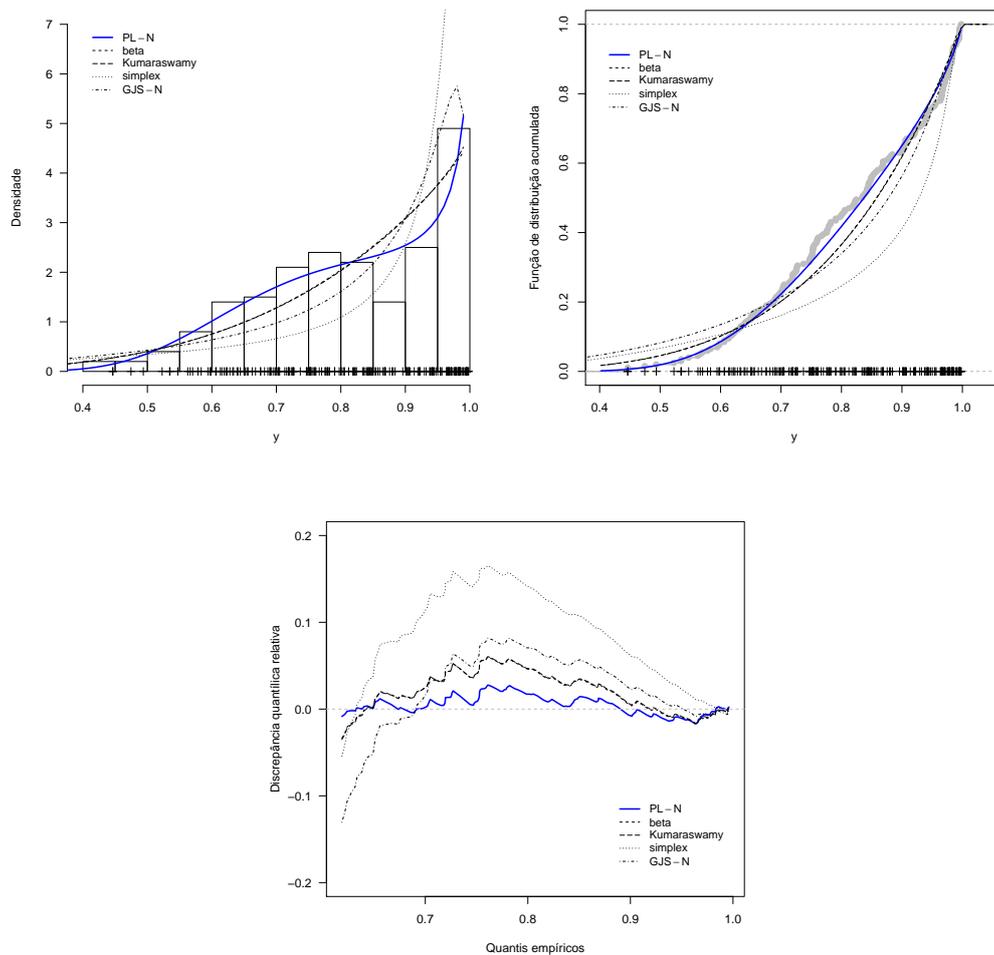


Figura 2.36: Funções densidade de probabilidade ajustas sobrepostas ao histograma de y , funções de distribuição empírica e ajustadas, e gráfico da distância quantílica relativa os modelos considerando os dados da proporção de indivíduos ocupados fora do setor agropecuário.

A Figura 2.37 apresenta o gráfico das versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo PL-N. A linha pontilhada apresenta um intervalo de 95% de confiança

para λ com base na quantidade $W_p^*(\lambda)$. O intervalo obtido foi (6.44, 13.10). A amplitude desse intervalo é menor do que a do intervalo baseado na distribuição assintótica de $\tilde{\lambda}$ e ele, também, não inclui o valor 1. Adicionalmente, observe que não há muita diferença entre as funções de log-verossimilhança usual e penalizada.

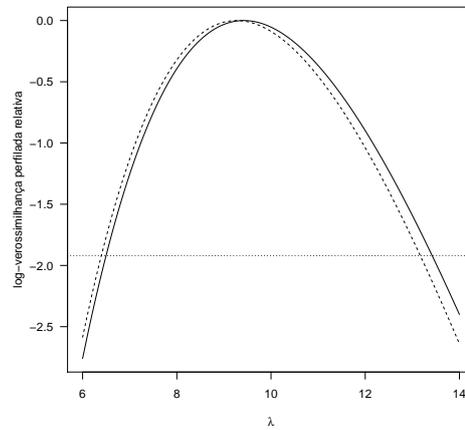


Figura 2.37: Gráfico das versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo PL-N.

Finalmente, observamos que a distribuição dos dados em estudo é fracamente bimodal e tem muitas observações próximas do valor 1. A distribuição PL consegue acomodar bem a bimodalidade, ao contrário das demais distribuições. Observe que o ajuste das distribuições beta, simplex, Kumaraswamy e GJS não conseguem acomodar bem as observações entre os quantis de ordem 0.2 e 0.7 (veja o gráfico das funções de distribuição acumulada ajustadas). Isso também é evidenciado através do gráfico da distância quantílica relativa. Através desse gráfico, concluímos que os quantis da distribuição PL-N ajustada são mais próximos dos quantis empíricos dos dados do que os dos demais ajustes. Portanto, essa aplicação evidencia a maior flexibilidade da classe *power logit* em detrimento aos demais modelos apresentados.

Proporção de indivíduos ocupados com ensino médio completo

Finalmente, consideramos dados do Instituto de Pesquisa Econômica Aplicada (IPEA) obtidos através do *Atlas de Vulnerabilidade Social*, disponíveis em <http://ivs.ipea.gov.br/index.php/pt/>. Esse atlas possibilita a consulta sobre a temática da vulnerabilidade social do Brasil. Nesta análise, consideramos dados coletados em 2010 referentes a 167 cidades do estado do Rio Grande do Norte.

A variável de interesse (y) é a proporção de indivíduos ocupados (com mais de 18 anos e que já concluíram o ensino médio. Essa variável é a razão entre o número de pessoas de 18 anos ou mais de idade ocupadas que já concluíram o ensino médio (regular seriado, regular não seriado, EJA ou supletivo) e o número total de pessoas ocupadas nessa faixa etária. São consideradas como já tendo concluído o médio aquelas pessoas que frequentavam a 4ª série desse nível de ensino. Algumas medidas descritivas da variável y são apresentadas na Tabela 2.13 e a Figura 2.38 apresenta o boxplot robusto para essa variável.

Conforme podemos observar, a proporção de indivíduos ocupados com ensino médio completo está concentrada entre 0.248 e 0.329. Ou seja, a maioria das pessoas ocupadas (com idade superior a 18 anos) nessas cidades não têm o ensino médio completo. Além disso, observamos que três cidades apresentam valores de y bem maiores do que as demais, são elas: Mossoró (0.486), Natal (0.566) e Parnamirim (0.612). Natal é a capital do estado e, juntamente com Mossoró e Parnamirim, formam as três cidades mais populosas do estado. Venha-Ver é a cidade com a menor proporção de indivíduos ocupados com ensino médio completo (0.163). No censo realizado pelo Instituto

Brasileiro de Geografia e Estatística (IBGE) no ano de 2010, a taxa de conclusão do ensino médio (de pessoas com idade entre 18 a 24 anos) nesta cidade era de apenas 26,9%.

Mínimo	0.163
Q_1	0.248
Mediana	0.283
Média	0.290
Q_3	0.329
Máximo	0.612

Tabela 2.13: Medidas resumo para a proporção de indivíduos ocupados com ensino médio completo (y).

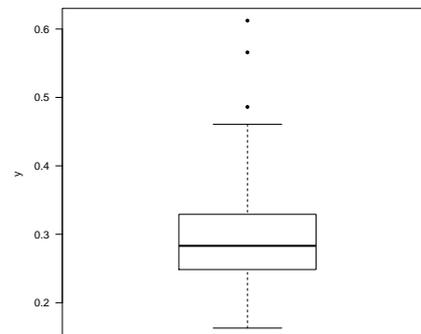


Figura 2.38: Boxplot para a variável y .

Os modelos beta, Kumaraswamy, simplex, GJS-N, GJS- $t_{(\zeta)}$, PL-N e PL- $t_{(\zeta)}$ foram ajustados aos dados. A Tabela 2.14 apresenta os resultados dos ajustes.

Tabela 2.14: Estimativas de máxima verossimilhança, intervalos de 95% de confiança e valores de Υ_{ζ} para cada modelo ajustado aos dados de proporção de indivíduos ocupados com ensino médio completo.

Distribuição	Parâmetro	Estimativa	IC(95%)	Υ_{ζ}
beta	ξ	0.290	(0.280, 0.300)	0.102
	ϕ	45.57	(35.887, 55.259)	
Kumaraswamy	μ	0.290	(0.277, 0.303)	0.227
	σ	0.247	(0.222, 0.272)	
simplex	ξ	0.291	(0.281, 0.301)	0.098
	σ	0.736	(0.657, 0.815)	
GJS-N	μ	0.286	(0.276, 0.296)	0.087
	σ	0.327	(0.292, 0.362)	
GJS- $t_{(7.6)}$	μ	0.284	(0.274, 0.293)	0.053
	σ	0.280	(0.232, 0.326)	
	ζ	7.6	–	
PL-N	μ	0.283	(0.273, 0.292)	0.067
	σ	0.972	(0, 2.080)	
	λ	4.272	(0, 9.314)	
PL- $t_{(6.2)}$	μ	0.283	(0.275, 0.292)	0.051
	σ	0.384	(0.258, 0.510)	
	λ	1.754	(0.851, 2.657)	
	ζ	6.2	–	

O ajuste que resultou em um menor valor de Υ_{ζ} foi o PL- $t_{(6.2)}$, seguido da distribuição GJS- $t_{(7.6)}$. O intervalo de 95% confiança para λ , para os modelos PL, contém o valor 1, o que pode ser um

indicativo de que o ajuste da distribuição GJS é mais razoável. As distribuições beta, Kumaraswamy e simplex apresentam valores de Υ_ζ maiores (ou muito maiores) do que os das distribuições PL e GJS. A estimativa da média de Y no modelo beta é igual à média amostral (0.290) e as estimativas da mediana de Y nos modelos PL-N e PL- $t_{(6,2)}$ são iguais à mediana amostral (0.283). Observe também que as estimativas de σ e λ apresentam variação considerável entre os modelos PL-N e PL- $t_{(6,2)}$; isso pode ser explicado pela presença de um parâmetro extra na distribuição PL- $t_{(\zeta)}$.

Os gráficos das funções densidade de probabilidade, funções de distribuição ajustadas e da discrepância quantílica relativa, para cada um dos modelos ajustados, são apresentados na Figura 2.39. Esses gráficos evidenciam que o modelo PL- $t_{(\zeta)}$ foi o que apresentou o melhor ajuste para os dados. Os gráficos das funções de distribuição ajustadas e da discrepância quantílica relativa mostram que os quantis da distribuição PL- $t_{(6,2)}$ são muito próximos dos quantis empíricos de y .

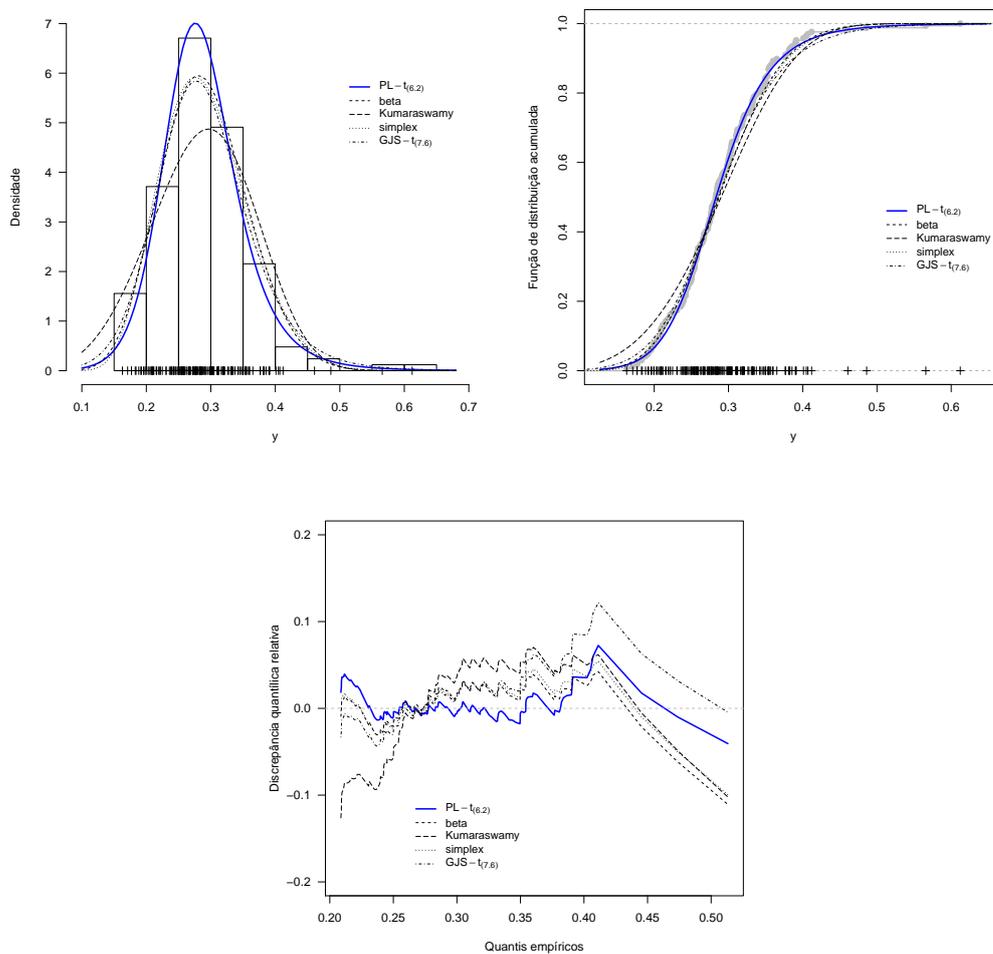


Figura 2.39: Funções densidade de probabilidade ajustas sobrepostas ao histograma de y , funções de distribuição empírica e ajustadas, e gráfico da distância quantílica relativa para alguns modelos considerando os dados da proporção de indivíduos ocupados com ensino médio completo.

Os gráficos de resíduo com envelopes simulados para cada ajuste são apresentados na Figura 2.40. Esses gráficos indicam que algumas das observações discrepantes observadas na análise descritiva (casos #162 e #164, principalmente) foram identificadas como *outliers*. Esses casos apresentam resíduos bem maiores que os demais para os ajustes dos modelos beta, Kumaraswamy, simplex e GJS-N e não são acomodados pelas bandas de confiança dos envelopes. Em alguns casos, essas observações atípicas comprometem o ajuste das demais observações (veja Figura 2.40(b)). Embora sejam apenas duas observações, é importante encontrar um modelo que consiga acomodá-las razoavelmente bem, pois elas representam cidades importantes no estado do Rio Grande do Norte,

tanto no âmbito educacional quanto econômico.

Uma comparação entre os gráficos 2.40(d) e 2.40(f) mostra que o modelo PL-N consegue acomodar melhor essas observações do que o modelo GJS-N. Inclusive, o resíduo referente a elas é menor para o modelo PL-N. Isso foi observado, também, através da medida Υ_ζ (0.053 para o modelo PL-N e 0.087 para o GJS-N). O modelo PL- $t_{(6,2)}$ consegue acomodar bem todas as observações.

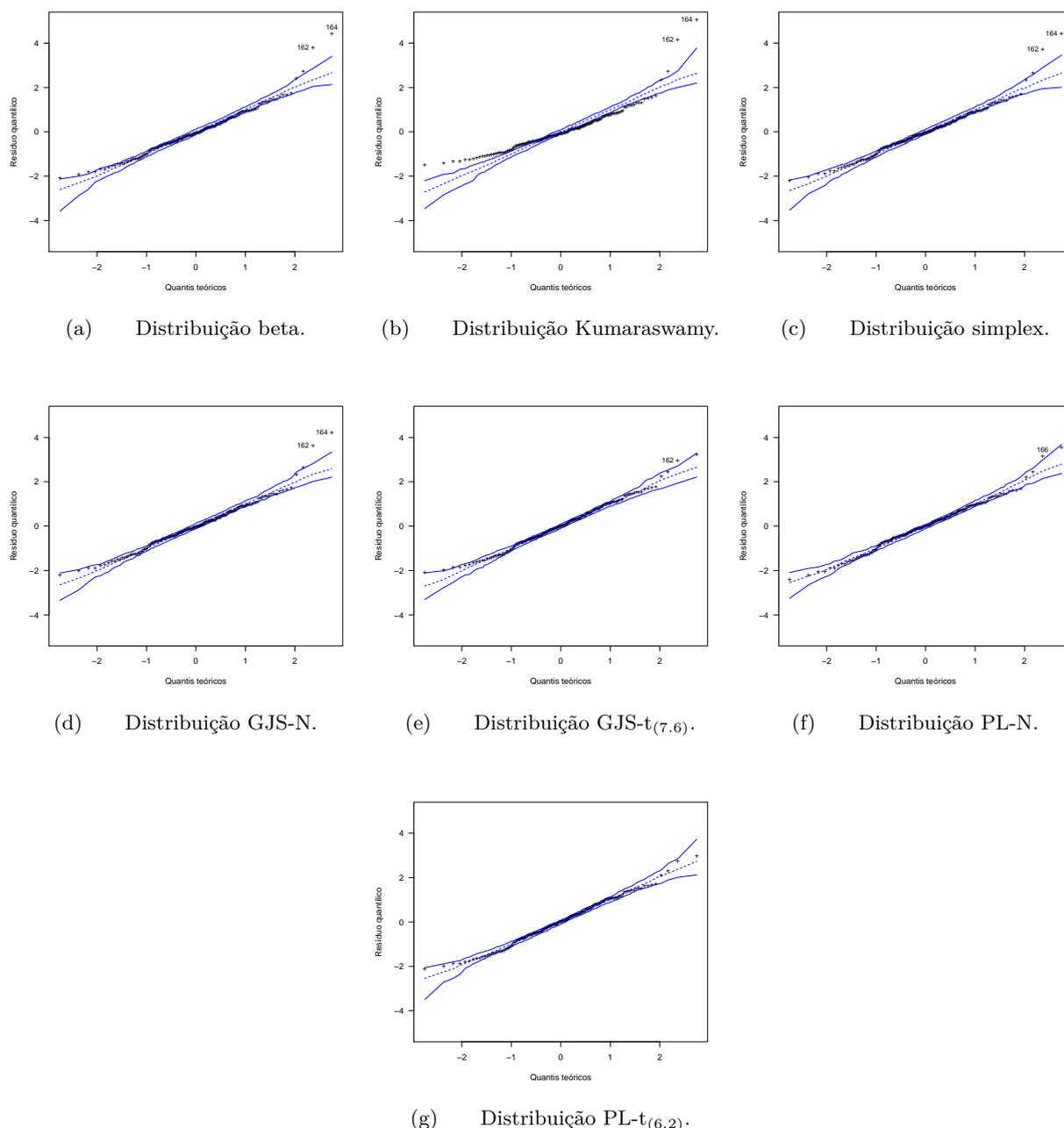


Figura 2.40: Gráfico de resíduo com envelopes simulados para os ajustes apresentados na Tabela 2.14.

A Figura 2.41 apresenta o gráfico de Υ_ζ versus ζ e o gráfico das versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo PL- $t_{(6,2)}$. O intervalo de 95% de confiança para λ com base na quantidade $W_p^*(\lambda)$ é representado pela linha pontilhada e foi (0, 16.1). Notamos que a penalização na função de log-verossimilhança perfilada de λ foi bastante eficaz. Embora a função de log-verossimilhança seja côncava em torno de $\lambda = 2$, ela apresenta comportamento monótono para $\lambda > 5$.

Conforme esperado, esse ajuste sugere que a classe de distribuições *power logit* é menos sensível

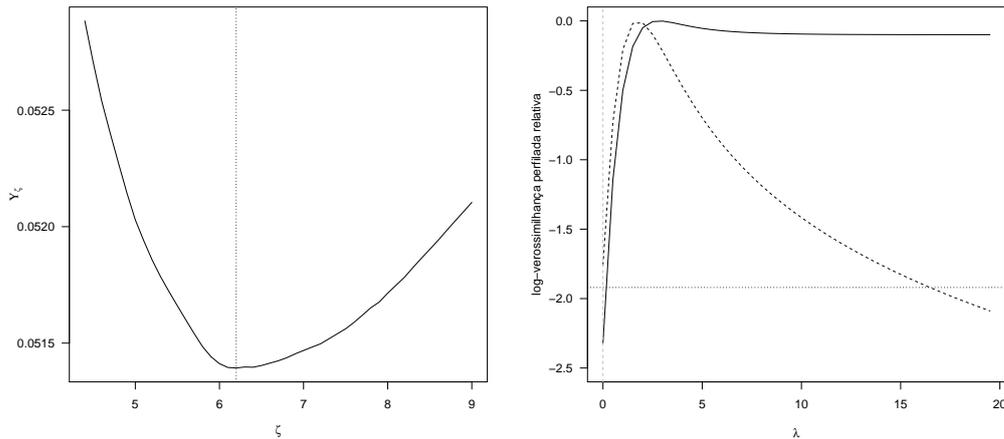


Figura 2.41: Gráfico de Υ_ζ e as versões relativas de $\ell^*(\lambda)$ (linha sólida) e de $\ell_p^*(\lambda)$ (linha tracejada) para o modelo $PL-t_{(\zeta)}$.

a observações discrepantes (*outliers*) do que as distribuições beta, simplex, Kumaraswamy e GJS. De fato, isso já havia sido observado nos estudos de simulação apresentados na Seção 2.4.3.

No que segue, ajustamos novamente os modelos beta, Kumaraswamy, simplex, GJS-N e PL-N aos dados em estudo mas, desta vez, excluindo os casos identificados como *outliers* (casos #162 e #164). Os modelos $GJS-t_{(\zeta)}$ e $PL-t_{(\zeta)}$ não serão apresentados pois os ajustes GJS-N e PL-N foram satisfatórios. A Tabela 2.15 apresenta os resultados dos ajustes.

Tabela 2.15: Estimativas de máxima verossimilhança, intervalos de 95% de confiança e valores de Υ_ζ para cada modelo ajustado aos dados de proporção de indivíduos ocupados com ensino médio completo, após a exclusão dos casos #162 e #164.

Distribuição	Parâmetro	Estimativa	IC(95%)	Υ_ζ
beta	ξ	0.286	(0.277, 0.295)	0.049
	ϕ	56.679	(44.537, 68.820)	
Kumaraswamy	μ	0.289	(0.279, 0.299)	0.116
	σ	0.199	(0.179, 0.220)	
simplex	ξ	0.286	(0.277, 0.295)	0.049
	σ	0.664	(0.592, 0.736)	
GJS-N	μ	0.283	(0.273, 0.292)	0.048
	σ	0.296	(0.247, 0.346)	
PL-N	μ	0.283	(0.273, 0.292)	0.048
	σ	0.290	(0.021, 0.558)	
	λ	0.953	(0, 2.771)	

Note que os valores de Υ_ζ para os modelos beta e simplex diminuíram consideravelmente. Além disso, a estimativa de λ para o modelo PL-N é bem próxima de 1, indicando que a distribuição GJS-N é a mais indicada. Os valores de Υ_ζ são próximos para todos os modelos considerados e, portanto, poderíamos considerar qualquer uma das distribuições ajustadas. Essa conclusão é

confirmada através dos gráficos de resíduo com envelopes simulados, apresentados na Figura 2.42. Esta aplicação sugere que o ajuste da distribuição PL é similar ao das demais distribuições na ausência de *outliers*. Por outro lado, essas observações discrepantes influenciam substancialmente o ajuste das demais distribuições, o que não é observado na distribuição *power logit*.

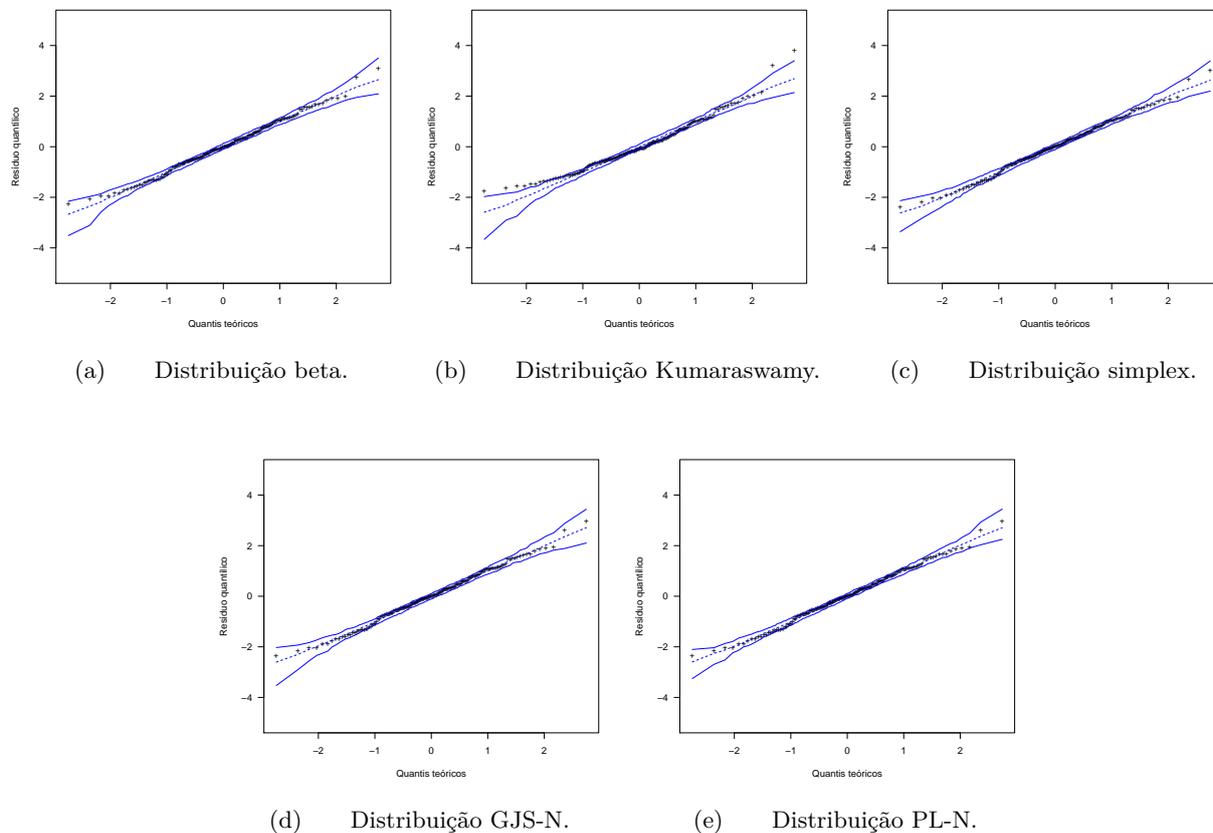


Figura 2.42: Gráfico de resíduo com envelopes simulados para os ajustes apresentados na Tabela 2.15, feitos após a exclusão dos casos #162 e #164.

Capítulo 3

O modelo de regressão *power logit*

Neste capítulo, um novo modelo de regressão para a análise de dados fracionários é proposto. O modelo é baseado na distribuição *power logit* e será denominado *modelo de regressão power logit*. Este novo modelo de regressão é bastante flexível, tem parâmetros interpretáveis e pode ser uma alternativa interessante para a modelagem de dados com presença de observações atípicas. O capítulo encontra-se organizado como segue. Na Seção 3.1 definimos o modelo de regressão *power logit*. Na Seção 3.2 são abordados aspectos inferenciais com base na teoria da verossimilhança, e são apresentadas expressões para o vetor escore e matriz da informação observada. Também, estudos de simulação Monte Carlo são conduzidos a fim de estudar o desempenho dos estimadores propostos em amostras de tamanho finito. Métodos de diagnósticos e medidas de influência para a classe de modelos propostos são apresentados nas Seções 3.3 e 3.4, respectivamente. Na Seção 3.5, apresentamos o modelo de regressão log-log. Uma aplicação em dados reais é conduzida na Seção 3.6.

3.1 Definição

Considere Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes com função densidade *power logit* dada por (2.4), isto é, $Y_i \sim \text{PL}(\mu_i, \sigma_i, \lambda; r)$, para $i = 1, \dots, n$. O modelo de regressão *power logit* é definido por (2.4) e pelos componentes sistemáticos

$$\begin{aligned}d_1(\mu_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} = \eta_{1i}, \\d_2(\sigma_i) &= \mathbf{s}_i^\top \boldsymbol{\tau} = \eta_{2i},\end{aligned}\tag{3.1}$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_q)^\top \in \mathbb{R}^q$ e $\lambda > 0$ são os parâmetros desconhecidos, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ e $\mathbf{s}_i = (s_{i1}, \dots, s_{iq})^\top$ são observações das covariáveis com $p + q < n$, e η_{1i} e η_{2i} são os preditores lineares. Assumimos que as funções de ligação $d_1 : (0, 1) \rightarrow \mathbb{R}$ e $d_2 : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. Diversas funções de ligação podem ser utilizadas. Para μ , possíveis escolhas são $d_1(\mu) = \log\{\mu/(1 - \mu)\}$ (ligação logito); $d_1(\mu) = \Phi^{-1}(\mu)$ (ligação probito), em que $\Phi(\cdot)$ denota a função de distribuição acumulada de uma variável aleatória com distribuição normal padrão; $d_1(\mu) = \log\{-\log(1 - \mu)\}$ (ligação complemento log-log); entre muitas outras. Já para σ e λ , possíveis escolhas são a ligação log, $d_2(\sigma) = \log \sigma$, e ligação raiz quadrada, $d_2(\sigma) = \sqrt{\sigma}$. McCullagh e Nelder (1989) apresentam uma discussão detalhada a respeito da escolha das funções de ligação.

O modelo de regressão *power logit*, definido por (2.4) e (3.1), generaliza alguns modelos já existentes na literatura, entre eles: o modelo de regressão GJS (Lemonte e Bazán, 2016) é obtido quando $\lambda = 1$; se $Y_i \sim \text{PL-LO}(\mu_i, \sigma_i, 1)$, temos o modelo de regressão L-Logístico (da Paz et al., 2019). Além disso, uma vez que a variância de Y_i é função de (μ_i, σ_i) , esse modelo acomoda variáveis respostas com variâncias não constantes. Adicionalmente, os parâmetros do modelo são interpretados em termos da mediana, dispersão e assimetria da distribuição da variável resposta.

3.2 Inferência sobre os parâmetros do modelo

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$ o vetor de parâmetros do modelo. A função de verossimilhança de $\boldsymbol{\theta}$ para o modelo de regressão *power logit* com valores observados y_1, \dots, y_n é da forma

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f_{Y_i}(y_i; \mu_i, \sigma_i, \lambda).$$

Note que μ_i e σ_i são definidos como funções de $\boldsymbol{\beta}$ e $\boldsymbol{\tau}$, respectivamente, através de (3.1), isto é, $\mu_i = d_1^{-1}(\eta_{1i})$, $\sigma_i = d_2^{-1}(\eta_{2i})$ e $\lambda_i = d_3^{-1}(\eta_{3i})$. A função de log-verossimilhança para $\boldsymbol{\theta}$ é dada por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\mu_i, \sigma_i, \lambda), \quad (3.2)$$

em que

$$\ell_i(\mu_i, \sigma_i, \lambda) = \log \lambda - \log \sigma_i - \log\{1 - y_i^\lambda\} + \log\{r(z_i^2)\},$$

sendo $z_i = h(y_i; \mu_i, \sigma_i, \lambda)$. O vetor escore para $\boldsymbol{\theta}$, definido como a primeira derivada da função de log-verossimilhança $\ell(\boldsymbol{\theta})$, é dado por $\mathbf{U}(\boldsymbol{\theta}) = (\mathbf{U}_\beta^\top, \mathbf{U}_\tau^\top, U_\lambda)^\top$. O cálculo dos componentes do vetor escore é feito como segue:

$$U_R = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \beta_R} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} \frac{d\mu_i}{d\eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_R}, \quad R = 1, \dots, p,$$

$$U_l = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \tau_l} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_l}, \quad l = 1, \dots, q,$$

$$U_\lambda = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \lambda} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \lambda},$$

em que

$$\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} = \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i), \quad \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i} = \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1]$$

e

$$\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \lambda} = \frac{1}{\lambda} + \frac{y_i^\lambda \log y_i}{1 - y_i^\lambda} - \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{\log y_i}{(1 - y_i^\lambda)} - \frac{\log \mu_i}{(1 - \mu_i^\lambda)} \right].$$

Além disso, temos que

$$\frac{d\mu_i}{d\eta_{1i}} = \frac{1}{\dot{d}_1(\mu_i)} = \quad e \quad \frac{d\sigma_i}{d\eta_{2i}} = \frac{1}{\dot{d}_2(\sigma_i)},$$

sendo $\dot{d}_j(t) = dd_j(t)/dt$, para $j = 1, 2$. Assim, uma vez que $\partial \eta_{1i}/\partial \beta_R = x_{iR}$ e $\partial \eta_{2i}/\partial \tau_l = s_{il}$, segue que

$$U_R = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \beta_R} = \sum_{i=1}^n \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i) \frac{1}{\dot{d}_1(\mu_i)} x_{iR}, \quad R = 1, \dots, p, \quad (3.3)$$

$$U_l = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \tau_l} = \sum_{i=1}^n \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1] \frac{1}{\dot{d}_2(\sigma_i)} s_{il}, \quad l = 1, \dots, q, \quad (3.4)$$

e

$$U_\lambda = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \lambda} = \sum_{i=1}^n \left\{ \frac{1}{\lambda} + \frac{y_i^\lambda \log y_i}{1 - y_i^\lambda} - \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{\log y_i}{(1 - y_i^\lambda)} - \frac{\log \mu_i}{(1 - \mu_i^\lambda)} \right] \right\}. \quad (3.5)$$

Sejam $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ e $\mathbf{S} = (\mathbf{s}_1, \dots, \mathbf{s}_n)^\top$ matrizes de posto completo (posto p e q , respectivamente), $\mathbf{T}_1 = \text{diag}\{1/\dot{d}_1(\mu_1), \dots, 1/\dot{d}_1(\mu_n)\}$, $\mathbf{T}_2 = \text{diag}\{1/\dot{d}_2(\sigma_1), \dots, 1/\dot{d}_2(\sigma_n)\}$, $\mathbf{W} = \text{diag}\{z_1 v(z_1), \dots, z_n v(z_n)\}$, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^\top$, $\boldsymbol{\sigma}^* = (\sigma_1^*, \dots, \sigma_n^*)^\top$, $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_n^*)^\top$, $\mu_i^* = \lambda/[\sigma_i \mu_i (1 - \mu_i^\lambda)]$, $\sigma_i^* = \sigma^{-1}[z_i^2 v(z_i) - 1]$ e

$$\lambda_i^* = \frac{1}{\lambda} + \frac{y_i^\lambda \log y_i}{1 - y_i^\lambda} - \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{\log y_i}{(1 - y_i^\lambda)} - \frac{\log \mu_i}{(1 - \mu_i^\lambda)} \right].$$

Assim,

$$\mathbf{U}_\beta = \mathbf{X}^\top \mathbf{W} \mathbf{T}_1 \boldsymbol{\mu}^*,$$

$$\mathbf{U}_\tau = \mathbf{S}^\top \mathbf{T}_2 \boldsymbol{\sigma}^*$$

e

$$\mathbf{U}_\lambda = \mathbf{1}_n^\top \boldsymbol{\lambda}^*,$$

em que $\mathbf{1}_n$ denota um vetor de uns de tamanho n . Portanto, o vetor escore para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$ é dado por $\mathbf{U}(\boldsymbol{\theta}) = (\mathbf{U}_\beta^\top, \mathbf{U}_\tau^\top, \mathbf{U}_\lambda)^\top$, isto é,

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}^\top \mathbf{W} \mathbf{T}_1 \boldsymbol{\mu}^* \\ \mathbf{S}^\top \mathbf{T}_2 \boldsymbol{\sigma}^* \\ \mathbf{1}_n^\top \boldsymbol{\lambda}^* \end{bmatrix}. \quad (3.6)$$

A matriz de informação observada para $\boldsymbol{\theta}$ será denotada por $\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$, e é dada pelo negativo da segunda derivada de $\ell(\boldsymbol{\theta})$ com relação a $\boldsymbol{\theta}$. Esses cálculos são apresentados na Seção B.2. Dessa forma, temos que

$$\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \begin{bmatrix} \mathbf{J}_{\beta\beta} & \mathbf{J}_{\beta\tau} & \mathbf{J}_{\beta\lambda} \\ \mathbf{J}_{\beta\tau}^\top & \mathbf{J}_{\tau\tau} & \mathbf{J}_{\tau\lambda} \\ \mathbf{J}_{\beta\lambda}^\top & \mathbf{J}_{\tau\lambda}^\top & \mathbf{J}_{\lambda\lambda} \end{bmatrix}, \quad (3.7)$$

cujos elementos são $\mathbf{J}_{\beta\beta} = \mathbf{X}^\top \mathbf{W}_1 \mathbf{T}_1 \mathbf{X}$, $\mathbf{J}_{\tau\tau} = \mathbf{S}^\top \mathbf{W}_2 \mathbf{T}_2 \mathbf{S}$, $\mathbf{J}_{\lambda\lambda} = \mathbf{1}_n^\top \mathbf{W}_3 \mathbf{1}_n$, $\mathbf{J}_{\beta\tau} = \mathbf{X}^\top \mathbf{W}_4 \mathbf{T}_1 \mathbf{T}_2 \mathbf{S}$, $\mathbf{J}_{\beta\lambda} = \mathbf{X}^\top \mathbf{W}_5 \mathbf{T}_1 \mathbf{1}_n$, $\mathbf{J}_{\tau\lambda} = \mathbf{S}^\top \mathbf{W}_6 \mathbf{T}_1 \mathbf{1}_n$, em que $\mathbf{W}_j = \text{diag}\{w_1^{(j)}, \dots, w_n^{(j)}\}$, com

$$\begin{aligned} w_i^{(1)} &= \frac{\lambda}{[\sigma_i \mu_i (1 - \mu_i^\lambda)]^2} \left\{ \sigma_i [1 - \mu_i^\lambda (1 + \lambda)] z_i v(z_i) + \lambda [v(z_i) + z_i v'(z_i)] \right\} \dot{d}_1(\mu_i)^{-1} \\ &\quad + \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i) \frac{\ddot{d}_1(\mu_i)}{\dot{d}_1(\mu_i)^2}, \\ w_i^{(2)} &= - \left\{ \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^2} z_i^2 [3v(z_i) + z_i v'(z_i)] \right\} \dot{d}_2(\sigma_i)^{-1} + \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1] \frac{\ddot{d}_2(\sigma_i)}{\dot{d}_2(\sigma_i)^2}, \\ w_i^{(3)} &= \frac{1}{\lambda^2} - \frac{y_i^\lambda \log^2 y_i}{(1 - y_i^\lambda)^2} + \frac{1}{\sigma_i^2} [v(z_i) + z_i v'(z_i)] \left(\frac{\log y_i}{1 - y_i^\lambda} - \frac{\log \mu_i}{1 - \mu_i^\lambda} \right)^2 \\ &\quad + \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{y_i^\lambda \log^2 y_i}{(1 - y_i^\lambda)^2} - \frac{\mu_i^\lambda \log^2 \mu_i}{(1 - \mu_i^\lambda)^2} \right], \\ w_i^{(4)} &= \frac{\lambda}{\sigma_i^2 \mu_i (1 - \mu_i^\lambda)} z_i [2v(z_i) + z_i v'(z_i)], \\ w_i^{(5)} &= \frac{-\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} \left\{ \frac{1 - \mu_i^\lambda (1 - \lambda \log \mu_i)}{\lambda (1 - \mu_i^\lambda)} z_i v(z_i) + \frac{1}{\sigma_i} \left(\frac{\log y_i}{1 - y_i^\lambda} - \frac{\log \mu_i}{1 - \mu_i^\lambda} \right) [v(z_i) + z_i v'(z_i)] \right\}, \\ w_i^{(6)} &= - \frac{1}{\sigma_i^2} \left(\frac{\log y_i}{1 - y_i^\lambda} - \frac{\log \mu_i}{1 - \mu_i^\lambda} \right) z_i [2v(z_i) + z_i v'(z_i)]. \end{aligned}$$

em que $\ddot{d}_j(z) = dd_j(z)/dz$, para $j = 1, 2$. Podemos escrever

$$\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \mathbf{P}^\top \mathbf{M} \mathbf{P}$$

em que \mathbf{P} é uma matriz de dimensão $3n \times (p + q + 1)$ e \mathbf{M} uma matriz de dimensão $3n \times 3n$ dadas por

$$\mathbf{P} = \begin{bmatrix} \mathbf{X} & \mathbf{0}_{n,q} & \mathbf{0}_{n,1} \\ \mathbf{0}_{n,p} & \mathbf{S} & \mathbf{0}_{n,1} \\ \mathbf{0}_{n,p} & \mathbf{0}_{n,q} & \mathbf{1}_n \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} \mathbf{M}_{\beta\beta} & \mathbf{M}_{\beta\tau} & \mathbf{M}_{\beta\lambda} \\ \mathbf{M}_{\beta\tau}^\top & \mathbf{M}_{\tau\tau} & \mathbf{M}_{\tau\lambda} \\ \mathbf{M}_{\beta\lambda}^\top & \mathbf{M}_{\tau\lambda}^\top & \mathbf{M}_{\lambda\lambda} \end{bmatrix},$$

sendo $\mathbf{0}_{l,c}$ uma matriz de zeros de dimensão $l \times c$, $\mathbf{M}_{\beta\beta} = \mathbf{W}_1 \mathbf{T}_1$, $\mathbf{M}_{\tau\tau} = \mathbf{W}_2 \mathbf{T}_2$, $\mathbf{M}_{\lambda\lambda} = \mathbf{W}_3$, $\mathbf{M}_{\beta\tau} = \mathbf{W}_4 \mathbf{T}_1 \mathbf{T}_2$, $\mathbf{M}_{\beta\lambda} = \mathbf{W}_5 \mathbf{T}_1$ e $\mathbf{M}_{\tau\lambda} = \mathbf{W}_6 \mathbf{T}_1$.

Os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ podem ser obtidos através da maximização da função de log-verossimilhança dada em (3.2). Por outro lado, conforme apresentado na Seção 2.3, os estimadores de máxima verossimilhança penalizada apresentam melhores propriedades e, portanto, serão considerados no contexto de regressão. Novamente, a penalização será feita apenas na verossimilhança perfilada para λ .

Nesse sentido, a função de log-verossimilhança perfilada penalizada para λ é dada por

$$\ell_p^*(\lambda) = \ell^*(\lambda) + \frac{1}{2} \log \left(\frac{\mathbf{J}_{\lambda\lambda}^*}{n} \right), \quad (3.8)$$

sendo $\ell^*(\lambda) = \ell(\hat{\boldsymbol{\beta}}_\lambda, \hat{\boldsymbol{\tau}}_\lambda, \lambda)$ a função de log-verossimilhança perfilada para λ e $\mathbf{J}_{\lambda\lambda}^* = \mathbf{J}_{\lambda\lambda}(\hat{\boldsymbol{\beta}}_\lambda, \hat{\boldsymbol{\tau}}_\lambda, \lambda)$, em que $\hat{\boldsymbol{\beta}}_\lambda$ e $\hat{\boldsymbol{\tau}}_\lambda$ são as estimativas de máxima verossimilhança para $\boldsymbol{\beta}$ e $\boldsymbol{\tau}$, respectivamente, considerando λ fixo. A penalização apenas na função de log-verossimilhança perfilada para λ torna o processo estimação numericamente menos extensivo no contexto de regressão. Não existem expressões analíticas fechadas para o estimador de máxima verossimilhança penalizada de $\boldsymbol{\theta}$, isto é, $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \tilde{\lambda})^\top$, e estas podem ser obtidas através de métodos numéricos, seguindo os seguintes passos.

i. Encontre $\tilde{\lambda}$ tal que

$$\tilde{\lambda} = \operatorname{argmax}_{\lambda > 0} \ell_p^*(\lambda).$$

ii. Obtenha $\tilde{\boldsymbol{\beta}}$ e $\tilde{\boldsymbol{\tau}}$ maximizando $\ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \tilde{\lambda})$.

Em geral, os algoritmos de maximização necessitam de valores iniciais para os parâmetros do modelo. Como estimativa inicial para o vetor $\boldsymbol{\beta}$, sugerimos as estimativas de mínimos quadrados ordinários obtidos da regressão linear da variável resposta transformada $d_1(y_1), \dots, d_1(y_n)$ em \mathbf{X} , isto é, $\boldsymbol{\beta}^{(0)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{v}$, em que $\mathbf{v} = (d_1(y_1), \dots, d_1(y_n))^\top$. Para o vetor de parâmetros $\boldsymbol{\tau}$, sugerimos $\boldsymbol{\tau}^{(0)} = (\tau_1^{(0)}, 0, \dots, 0)^\top$, em que $\tau_1^{(0)}$ é o desvio padrão da variável $\log[y_i/(1 - y_i)]$; isto é, iniciar o algoritmo com o modelo de regressão *power logit* com dispersão constante. Finalmente, um valor inicial para λ é $\lambda^{(0)} = 1$.

Adicionalmente, os modelos de regressão *power logit* podem depender de parâmetro(s) extra(s), denotado(s) aqui por ζ . Assim como no caso em que as variáveis são independentes e identicamente distribuídas, a estimação desse parâmetro (ou do vetor de parâmetros) será feita através da minimização da quantidade Υ_ζ , definida na Seção 2.3.2. Isto é, $\hat{\zeta}$ é dado por

$$\hat{\zeta} = \operatorname{argmin} \Upsilon_\zeta.$$

3.2.1 Intervalos de confiança e testes de hipóteses

Adicionalmente, sob condições gerais de regularidade, $\tilde{\boldsymbol{\theta}}$ é um estimador consistente para $\boldsymbol{\theta}$ e

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_{p+q+1}(\mathbf{0}_{p+q+1}, \mathbf{K}(\boldsymbol{\theta})^{-1}), \text{ quando } n \rightarrow \infty,$$

em que $\mathbf{K}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ é a matriz da informação de Fisher de $\boldsymbol{\theta}$ para uma única observação. Não existe uma expressão analítica fechada para $\mathbf{K}(\boldsymbol{\theta})$, mas resultados assintóticos mostram que podemos utilizar a matriz $\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ para realizar inferência sobre $\boldsymbol{\theta}$.

Utilizando a distribuição assintótica de $\hat{\boldsymbol{\theta}}$, podemos construir intervalos de confiança aproximados para os parâmetros do modelo. Sejam β_R ($R = 1, \dots, p$) e τ_l ($l = 1, \dots, q$) o R -ésimo e q -ésimo componentes dos vetores $\boldsymbol{\beta}$ e $\boldsymbol{\tau}$, respectivamente. Assim, considerando $0 < \alpha < 1/2$, e nível de confiança $100(1 - \alpha)\%$, os intervalos de confiança assintóticos para β_R , τ_l e λ são dados, respectivamente, por

$$\begin{aligned}\tilde{\beta}_R &\pm z_{1-\alpha/2}(\tilde{\mathbf{J}}_{RR}^{\beta\beta})^{1/2}, \\ \tilde{\tau}_l &\pm z_{1-\alpha/2}(\tilde{\mathbf{J}}_{ll}^{\tau\tau})^{1/2}, \\ \tilde{\lambda} &\pm z_{1-\alpha/2}(\tilde{\mathbf{J}}^{\lambda\lambda})^{1/2},\end{aligned}\tag{3.9}$$

em que $\tilde{\mathbf{J}}_{RR}^{\beta\beta}$, $\tilde{\mathbf{J}}_{ll}^{\tau\tau}$ e $\tilde{\mathbf{J}}^{\lambda\lambda}$ são, respectivamente, as variâncias assintóticas de $\tilde{\beta}_R$, $\tilde{\tau}_l$ e $\tilde{\lambda}$, sendo $\tilde{\mathbf{J}}_{RR}^{\beta\beta}$ o (R, R) -ésimo elemento da matriz $\mathbf{J}^{\beta\beta}$ avaliado em $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \tilde{\lambda})$, $\tilde{\mathbf{J}}_{ll}^{\tau\tau}$ o (l, l) -ésimo elemento da matriz $\mathbf{J}^{\tau\tau}$ avaliado em $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \tilde{\lambda})$ e $\tilde{\mathbf{J}}^{\lambda\lambda}$ o $(p + q + q, p + q + q)$ -ésimo elemento de $\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ avaliado em $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \tilde{\lambda})$. Ainda, para $0 < \alpha < 1/2$, $z_{1-\alpha/2}$ denota o quantil $1 - \alpha/2$ da distribuição normal padrão. Adicionalmente, podemos utilizar a quantidade $W_p^*(\lambda)$, dada por

$$W_p^*(\lambda) = 2\{\ell_p^*(\tilde{\lambda}) - \ell_p^*(\lambda)\},$$

para construir um intervalo de confiança assintótico para λ . Essa quantidade tem distribuição assintótica χ_1^2 e, em geral, produz intervalos de confiança com melhores propriedades para λ .

Por outro lado, podemos estar interessados em testar se um particular β_R e/ou τ_l é igual a zero, com $R = 1, \dots, p$ e $l = 1, \dots, q$, ou se $\lambda = 1$. Neste caso, pode-se utilizar os respectivos intervalos de confiança dados em (3.9). Dessa forma, se o intervalo contém o valor zero ou o valor um (no caso do teste para λ), não rejeitamos a hipótese nula de que o respectivo parâmetro é igual a zero ou o valor um (no caso do teste para λ); caso contrário, a hipótese nula é rejeitada. Outra forma de realizar testes de hipóteses é através de estatísticas de testes. As principais estatísticas de teste são a estatística da razão de verossimilhanças (Wilks, 1938), Wald (Wald, 1943), escore de Rao (Rao, 1948) e gradiente (Terrell, 2002).

3.2.2 Resultados numéricos

Através de simulações Monte Carlo, estudamos o desempenho das estimativas de máxima verossimilhança usual e penalizada em amostras de tamanho finito, para diferentes modelos de regressão *power logit*. As simulações foram feitas no *software* R. Nesse estudo, consideramos um modelo de regressão *power logit* com as seguintes componentes sistemáticas

$$\begin{aligned}\log\left(\frac{\mu_i}{1 - \mu_i}\right) &= 0.5 + 1.5x_i, \\ \log\sigma_i &= -1 + 0.5s_i,\end{aligned}$$

com $\lambda = 1$, e x_i e s_i obtidas, independentemente, de uma variável uniforme no intervalo $(0, 1)$ e mantidas fixas para cada réplica Monte Carlo. Essa configuração resulta em $\mu_i \in (0.61, 0.88)$ e $\sigma_i \in (0.35, 0.60)$, para $i = 1, \dots, n$. Consideramos $n = 40, 120$ e $M = 3000$ réplicas Monte Carlo. A Tabela 3.1 apresenta os resultados de simulação para os modelos de regressão PL- $t_{(5)}$, PL-PE $_{(1.5)}$, PL-Hyp $_{(1.2)}$ e PL-Slshs $_{(1.4)}$. Simulações com diferentes modelos foram conduzidas e resultados similares foram observados.

Em geral, notamos que as estimativas de máxima verossimilhança para os β 's, tanto usuais quanto penalizadas, apresentam vieses próximo de zero e erro quadrático médio menor que para os demais modelos. Note que, para o modelo de regressão PL- $t_{(5)}$, o viés da estimativa de máxima

Tabela 3.1: Estimativas do viés e da raiz do erro quadrático médio (\sqrt{EQM}) das 3000 estimativas de máxima verossimilhança (usual e penalizada) para cada tamanho de amostra e para os modelos de regressão $PL-t_{(5)}$, $PL-PE_{(1.5)}$, $PL-Hyp_{(1.2)}$ e $PL-slash_{(1.4)}$.

		$n = 40$				$n = 120$			
		emv usual		emv penalizada		emv usual		emv penalizada	
		Viés	\sqrt{EQM}	Viés	\sqrt{EQM}	Viés	\sqrt{EQM}	Viés	\sqrt{EQM}
$PL-t_{(5)}$	β_1	-0.00	0.14	-0.00	0.14	-0.00	0.09	-0.00	0.09
	β_2	-0.00	0.28	-0.00	0.28	-0.00	0.18	-0.00	0.18
	τ_1	0.21	0.65	0.05	0.48	0.08	0.34	0.04	0.31
	τ_2	-0.14	0.61	-0.04	0.55	-0.06	0.36	-0.03	0.35
	λ	2.07	5.69	0.79	2.67	0.60	1.79	0.36	1.49
$PL-PE_{(1.5)}$	β_1	-0.00	0.12	-0.00	0.12	-0.00	0.08	-0.00	0.08
	β_2	-0.00	0.24	-0.00	0.24	-0.00	0.15	-0.00	0.15
	τ_1	0.20	0.66	0.05	0.48	0.08	0.35	0.04	0.32
	τ_2	-0.15	0.60	-0.06	0.54	-0.06	0.36	-0.04	0.34
	λ	2.08	5.71	0.80	2.60	0.61	1.91	0.41	1.60
$PL-Hyp_{(1.2)}$	β_1	-0.00	0.16	0.00	0.16	-0.00	0.10	-0.00	0.10
	β_2	-0.01	0.33	-0.01	0.33	-0.00	0.20	-0.00	0.20
	τ_1	0.15	0.54	0.04	0.43	0.05	0.30	0.02	0.28
	τ_2	-0.11	0.57	-0.04	0.53	-0.05	0.35	-0.03	0.34
	λ	1.49	3.88	0.72	2.25	0.43	1.37	0.29	1.20
$PL-slash_{(1.4)}$	β_1	-0.02	0.19	-0.02	0.18	-0.01	0.12	-0.00	0.12
	β_2	-0.01	0.37	-0.01	0.37	-0.00	0.23	-0.00	0.23
	τ_1	-0.01	0.49	-0.10	0.43	-0.01	0.29	-0.04	0.28
	τ_2	0.01	0.57	0.06	0.55	0.01	0.35	0.04	0.34
	λ	0.50	3.79	-0.08	1.70	0.14	1.24	-0.04	1.10

verossimilhança de β_2 , quando $n = 40$, é igual a zero com duas casas decimais. Além disso, notamos que a penalização não parece interferir na estimação dos β 's; observe que o viés e erro quadrático médio não sofrem influência considerável dos métodos de estimação.

As estimativas dos parâmetros associados à dispersão, isto é, τ_1 e τ_2 , em geral, apresentam vieses próximos de zero mas diferentes de zero. Além disso, notamos que a penalização faz com que os vieses dessas estimativas diminuam; no modelo de regressão $PL-t_{(5)}$, por exemplo, o viés da estimativa de máxima verossimilhança de τ_1 , quando $n = 40$, é 0.21, enquanto o viés da estimativa de máxima verossimilhança penalizada é 0.05. Além disso, a estimativa de máxima verossimilhança penalizada resulta em estimativas com EQM menor que as estimativa de máxima verossimilhança usual. No modelo de regressão $PL-PE_{(1.5)}$ e com $n = 40$, a raiz quadrada do EQM de τ_1 é 0.66 para a estimativa de máxima verossimilhança usual e 0.48 para a estimativa de máxima verossimilhança penalizada.

A penalização é bastante efetiva na estimação de λ . Em todos os cenários, quando o tamanho da amostra é pequeno ou moderado ($n = 40$, por exemplo), observamos que o viés da estimativa de máxima verossimilhança de λ é considerável. No modelo de regressão $PL-PE_{(1.5)}$, o viés da estimativa de máxima verossimilhança de λ é 2.08 e a raiz quadrada do erro quadrático médio é 5.71. Dentre os modelos em estudo, apenas a estimativa de máxima verossimilhança de λ para o modelo de regressão $PL-slash_{(1.4)}$ apresentou viés próximo de zero para $n = 40$; por outro lado, o desvio padrão é bem maior do que para os demais modelos (observe que a raiz quadrada do EQM é 3.79). A penalização, neste caso, foi bastante efetiva e resultou em estimativas menos viés e com

menor variabilidade. Observe que para o modelo de regressão PL-Hyp_(1.2), o viés da estimativa de máxima verossimilhança usual de λ , para $n = 40$, é 1.49, enquanto que o da estimativa de máxima verossimilhança penalizada é 0.72. A raiz quadrada do erro quadrático médio também diminui consideravelmente (passa de 3.88 para 2.25). Esse comportamento é observado em todos os modelos estudados.

Quando o tamanho da amostra é grande ($n = 120$), as estimativas de todos os parâmetros apresentam vieses próximos de zero e EQM pequeno. Embora a penalização também funcione conforme esperado, não é tão efetiva quanto em amostras de tamanho pequeno ou moderado. Em resumo, recomendamos a utilização da penalização na função de log-verossimilhança para λ para a obtenção de estimativas mais confiáveis para os parâmetros dos modelos de regressão *power logit*, quando o tamanho da amostra for pequeno ou moderado.

3.3 Diagnóstico no modelo de regressão *power logit*

Os métodos de diagnóstico desempenham papel importante na modelagem de regressão. O interesse principal é verificar a existência de possíveis afastamentos das suposições assumidas para o modelo, bem como a presença de observações atípicas com alguma influência desproporcional no ajuste do modelo. Na prática, diferentes falhas no ajuste de um modelo podem ocorrer, por exemplo: definição incorreta da distribuição da variável resposta ou das componentes sistemáticas, presença de dispersão não constante, correlação entre os erros, etc.. Nesse sentido, a análise de resíduos pode ajudar a identificar essas instabilidades. Existem diferentes formas de definir resíduos e é importante que eles levem em consideração a contribuição que cada observação exerce sobre o ajuste do modelo. Uma das principais referências nesse tema é [Cox e Snell \(1968\)](#), em que é apresentada uma forma geral de definir resíduos. Inúmeros trabalhos têm sido realizados a fim de estudar métodos de diagnóstico em diferentes classes de modelos estatísticos, entre eles: [Dunn e Smyth \(1996\)](#), [Galea et al. \(2005\)](#), [Espinheira et al. \(2008\)](#), [Barros et al. \(2008\)](#), [Espinheira et al. \(2017\)](#), [Lemonte e Moreno-Arenas \(2019\)](#), entre muitos outros.

Esta seção dispõe-se a estudar resíduos em modelos de regressão *power logit*. Consideramos três propostas de resíduos: padronizado, deviance e quantílico. Estes resíduos serão discutidos a seguir.

Resíduo padronizado

Inicialmente, considere o modelo de regressão *power logit* com dispersão constante, isto é, $\sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma$ e assumamos que λ é fixado. Neste caso, temos que

$$z_i = \frac{1}{\sigma} \left[\log \left(\frac{y_i^\lambda}{1 - y_i^\lambda} \right) - \log \left(\frac{\mu_i^\lambda}{1 - \mu_i^\lambda} \right) \right] \sim S(0, 1; r).$$

Assim, o modelo de regressão *power logit* com dispersão constante é equivalente a

$$y_i^\dagger = \mu_i^\dagger(\boldsymbol{\beta}; \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.10)$$

sendo $\varepsilon_i \sim S(0, \sigma^2; r)$, $y_i^\dagger = \log[y_i^\lambda / (1 - y_i^\lambda)]$ e $\mu_i^\dagger(\boldsymbol{\beta}; \mathbf{x}_i) = \mu_i^\dagger = \log[\mu_i^\lambda / (1 - \mu_i^\lambda)]$ uma função não linear de $\boldsymbol{\beta}$ contínua e duas vezes diferenciável, com matriz de derivadas $\mathbf{D} = \partial \boldsymbol{\mu}^\dagger / \partial \boldsymbol{\beta}$ (com posto $p < n$) para todo $\boldsymbol{\beta}$. O modelo (3.10) pode ser visto como um modelo de regressão não linear simétrico ([Cordeiro et al., 2000](#); [Cysneiros e Vanegas, 2008](#); [Galea et al., 2005](#)). Neste caso, temos

$$\mu_i^\dagger(\boldsymbol{\beta}; \mathbf{x}_i) = \log \left(\frac{\mu_i^\lambda}{1 - \mu_i^\lambda} \right) = \log \left\{ \frac{[d_1^{-1}(\eta_{1i})]^\lambda}{1 - [d_1^{-1}(\eta_{1i})]^\lambda} \right\}$$

e matriz de derivadas dada por

$$\mathbf{D} = \sigma \boldsymbol{\mu}^* \mathbf{T}_1 \mathbf{X}.$$

Dessa forma, resíduos para o modelo de regressão *power logit* podem ser desenvolvidos com base em resíduos para o modelo de regressão não linear simétrico. Um resíduo ordinário para o modelo (3.10) pode ser definido da seguinte forma

$$r_i = r_i(y_i, \tilde{\mu}_i; \sigma, \lambda) = y_i^\dagger - \tilde{\mu}_i^\dagger,$$

para $i = 1, \dots, n$. Esse resíduo, no entanto, não tem distribuição normal, mesmo assintoticamente, e o desvio padrão não é 1, o que dificulta a verificação da qualidade do ajuste através dos métodos tradicionais. Uma alternativa seria padronizar r_i de modo a obter resíduos comparáveis. Nesse sentido, seguindo a proposta de Cox e Snell (1968), Galea *et al.* (2005) propõem corrigir, até a ordem n^{-1} , os dois primeiros momentos de r_i a fim de obter propriedades próximas às do i -ésimo erro ε_i .

Através das expansões até a ordem n^{-1} apresentadas em Cox e Snell (1968), quando existem,

$$\mathbb{E}(\mathbf{r}) \approx (\mathbf{I}_n - \mathbf{H})\boldsymbol{\eta}$$

e

$$\text{Var}(\mathbf{r}) \approx \sigma^2 \xi_r \{\mathbf{I}_n - (d_r \xi_r)^{-1} \mathbf{H}\},$$

em que $\mathbf{r} = (r_1, \dots, r_n)^\top$, ξ_r é tal que $\text{Var}(Z) = \xi_r \sigma^2$, $\mathbf{H} = \mathbf{D}(\mathbf{D}^\top \mathbf{D})^{-1} \mathbf{D}^\top$, \mathbf{I}_n é a matriz identidade de ordem n e $\boldsymbol{\eta}$ é a diferença entre a aproximação linear e quadrática de $\mu_i^\dagger(\tilde{\boldsymbol{\beta}}; \mathbf{x}_i)$. Logo, uma forma padronizada do resíduo r_i é dada por

$$r_i^* = \frac{y_i^\dagger - \tilde{\mu}_i^\dagger}{\sigma \sqrt{\xi_r \{1 - (d_r \xi_r)^{-1} h_{ii}\}}}, \quad i = 1, \dots, n.$$

Até o momento, assumimos que σ e λ são fixos ou conhecidos. Na prática, essas quantidades devem ser substituídas por suas respectivas estimativas; neste trabalho, utilizamos as estimativas de máxima verossimilhança penalizada, isto é, $\tilde{\sigma}$ e $\tilde{\lambda}$. Da mesma forma, as quantidades ξ_r e d_r são avaliadas em uma estimativa de ζ . Finalmente, definimos o resíduo padronizado da seguinte forma:

$$r_i^p = \frac{\tilde{y}_i^\dagger - \tilde{\mu}_i^\dagger}{\tilde{\sigma} \sqrt{\tilde{\xi}_r \{1 - (\tilde{d}_r \tilde{\xi}_r)^{-1} \tilde{h}_{ii}\}}} = \frac{\tilde{z}_i}{\sqrt{\tilde{\xi}_r \{1 - (\tilde{d}_r \tilde{\xi}_r)^{-1} \tilde{h}_{ii}\}}}, \quad i = 1, \dots, n,$$

em que $\tilde{y}_i^\dagger = \log(y_i^{\tilde{\lambda}} / (1 - y_i^{\tilde{\lambda}}))$, \tilde{h}_{ii} é o i -ésimo elemento diagonal de \mathbf{H} avaliado na estimativa de máxima verossimilhança penalizada.

No caso do modelo de regressão *power logit* com dispersão variável teremos, analogamente, o respectivo modelo simétrico não linear heterocedástico (ver, por exemplo, Cysneiros *et al.* (2010)). Neste caso, pode-se mostrar que

$$\text{Var}(\mathbf{r}) \approx \xi_r \boldsymbol{\Sigma} \{\mathbf{I}_n - (d_r \xi_r)^{-1} \mathbf{H}\},$$

em que $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$ e $\mathbf{H} = \boldsymbol{\Sigma}^{-1/2} \mathbf{D}(\mathbf{D}^\top \boldsymbol{\Sigma}^{-1} \mathbf{D})^{-1} \mathbf{D}^\top \boldsymbol{\Sigma}^{-1/2}$. Dessa forma, o resíduo padronizado é dado por

$$r_i^p = \frac{\tilde{y}_i^\dagger - \tilde{\mu}_i^\dagger}{\tilde{\sigma}_i \sqrt{\tilde{\xi}_r [1 - (\tilde{d}_r \tilde{\xi}_r)^{-1} \tilde{h}_{ii}]}}, \quad i = 1, \dots, n, \quad (3.11)$$

sendo \tilde{h}_{ii} o i -ésimo elemento diagonal de $\mathbf{H} = \boldsymbol{\Sigma}^{-1/2} \mathbf{D}(\mathbf{D}^\top \boldsymbol{\Sigma}^{-1} \mathbf{D})^{-1} \mathbf{D}^\top \boldsymbol{\Sigma}^{-1/2}$ avaliado na estimativa de máxima verossimilhança penalizada.

Uma desvantagem do resíduo padronizado é que ele depende da existência da variância de Z e da função d_r e, em alguns casos, elas podem não existir. Felizmente, para a grande maioria dos casos, esse resíduo está bem definido. A Tabela 3.2 apresenta os valores de ξ_r e de d_r para os principais modelos simétricos estudados aqui.

Tabela 3.2: Valores de ξ_r e d_r para alguns modelos na classe *power logit*.

Distribuição	ξ_r	d_r
PL-N	1	1
PL-t _(ζ)	$\frac{\zeta}{\zeta - 2}$	$\frac{\zeta + 1}{\zeta + 3}$
PL-PE _(ζ)	1	$\zeta^2 \Gamma\left(\frac{1}{\zeta}\right)^{-2} \Gamma\left(\frac{3}{\zeta}\right) \Gamma\left(\frac{2\zeta - 1}{\zeta}\right)$
PL-LO	0.79569	1.47724
PL-slash _(ζ)	$\frac{\zeta}{\zeta - 1}$	$\frac{4\zeta(\zeta + 1/2)[(\zeta + 3/2)(\zeta + 5/2) + \zeta + 1]}{(\zeta + 1)(\zeta + 3/2)^2(\zeta + 5/2)}$
PL-Hyp _(ζ)	$\frac{K_2(\zeta)}{\zeta K_1(\zeta)}$	$\frac{\zeta^2 h_1(\zeta)}{K_1(\zeta)}$
PL-SN _(ζ)	$q(\zeta)$	$2 + \frac{4}{\zeta^2} - \frac{\sqrt{2\pi}}{\zeta} \left[1 - \operatorname{erf}\left(\frac{\sqrt{2}}{\zeta}\right) \right] \exp\left\{\frac{2}{\zeta^2}\right\}$

Na Tabela 3.2, os valores de ξ_r e d_r para os modelos PL-LO, PL-slash_(ζ) e PL-SN_(ζ) são aproximados. Além disso, $K_2(\cdot)$ denota a função Bessel modificada de terceira ordem e índice 2, $h_1(\zeta) = \int_1^\infty (\sqrt{x^2 - 1}/x) \exp\{-\zeta x\} dx$, $\operatorname{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp\{-t^2\} dt$ é a função erro e

$$q(\zeta) = \begin{cases} \frac{\zeta^2}{4} \left(1 - \frac{\zeta^2}{4}\right), & \text{para } 0 < \zeta < 2, \\ 2.197543451 - 1.963510026 \log(\zeta\sqrt{2}) + [\log(\zeta\sqrt{2})]^2, & \text{para } \zeta > 2. \end{cases}$$

Conforme mencionado, para alguns valores de ζ , as funções ξ_r e d_r podem não existir para algumas distribuições na classe *power logit*. Dentre os modelos em estudo, essas funções estão bem definidas para os modelos PL-N; PL-t_(ζ), para $\zeta > 2$; PL-PE_(ζ), para $\zeta > 1/2$; PL-LO; PL-slash_(ζ), para $\zeta > 1$; PL-Hyp_(ζ); e PL-SN_(ζ).

Resíduo deviance

Seguindo a ideia dos modelos lineares generalizados (McCullagh e Nelder, 1989), propomos o resíduo deviance, que é baseado em

$$\operatorname{sign}(y_i - \tilde{\mu}_i) \{2[\ell_i(\bar{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda}) - \ell_i(\tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda})]\}^{1/2}, \quad i = 1, \dots, n,$$

sendo $\bar{\mu}_i$ é a estimativa de máxima verossimilhança de μ_i no modelo saturado. Pode-se verificar que, para os modelos *power logit*, $\bar{\mu}_i = y_i$. Assim, o resíduo deviance é dado por

$$r_i^d = \operatorname{sign}(y_i - \tilde{\mu}_i) \{2[\ell_i(y_i, \tilde{\sigma}_i, \tilde{\lambda}_i) - \ell_i(\tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda}_i)]\}^{1/2}, \quad i = 1, \dots, n.$$

Note que esse resíduo é baseado na diferença entre a log-verossimilhança maximizada no modelo saturado e no modelo sob investigação. Após algumas simplificações, o resíduo r_i^d pode ser escrito como

$$r_i^d = \operatorname{sign}(\tilde{z}_i) \left\{ 2 \log \left[\frac{r(0)}{r(\tilde{z}_i^2)} \right] \right\}^{1/2}, \quad i = 1, \dots, n.$$

Resíduo quantílico

Baseando-se na proposta de Dunn e Smyth (1996), consideramos o resíduo quantílico para a classe de modelos de regressão *power logit*. Esse resíduo será denotado por r_i^q e é definido por

$$r_i^q = \Phi^{-1}[F_{Y_i}(y_i; \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda})], \quad i = 1, \dots, n,$$

em que $F_{Y_i}(\cdot)$ é a função de distribuição acumulada de Y_i , definida em (2.5). Também podemos escrever r_i^q da forma

$$r_i^q = \Phi^{-1}[R(\tilde{z}_i)], \quad i = 1, \dots, n,$$

em que $R(\cdot)$ é a função de distribuição acumulada de Z .

Se a distribuição especificada para os dados é correta e os parâmetros do modelo são estimados consistentemente, o resíduo quantílico tem distribuição assintótica normal com média 0 e variância 1. Já a distribuição dos resíduos r_i^p e r_i^d não é conhecida e uma forma de verificar a adequação do modelo é fazer o gráfico de probabilidade normal envelopes simulados (Atkinson, 1981). Tais bandas podem auxiliar na avaliação de simetria para os resíduos, adequação das caudas, presença de observações discrepantes, etc.. Outros métodos gráficos podem ser úteis para detectar diferentes falhas no ajuste do modelo, por exemplo: (i) gráfico dos resíduos *versus* valores ajustados (ou índice das observações) para detectar dispersão não constante; (ii) gráfico dos resíduos *versus* covariáveis para verificar a forma que as covariáveis devem entrar no modelo; (iii) gráfico de índices de \tilde{h}_{ii} pode ajudar a detectar observações que tem grande influência no seu valor predito; entre outros.

3.3.1 Aplicações em dados simulados

Nesta seção, consideramos três aplicações em dados simulados com o objetivo de avaliar o desempenho dos resíduos propostos na detecção de três diferentes tipos de perturbações no modelo: presença de observação discrepante (Experimentação I), suposição incorreta de dispersão constante (Experimentação II) e especificação incorreta da componente sistemática associada à mediana da distribuição (Experimentação III).

Experimentação I

Neste cenário, geramos $n = 40$ observações de um modelo de regressão PL-N com

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = 3.5 - 3.5x_i, \quad i = 1, \dots, 40,$$

em que x_i , para $i = 1, \dots, 38$, são obtidos de uma variável aleatória $U \sim U(0, 1)$ e $x_{39} = x_{40} = 1$, $\log \sigma = -2$ e $\lambda = 0.5$. Em seguida, contaminamos os dados trocando os valores de y_{39} e y_{40} por y_{39}^* e y_{40}^* , que são obtidos de uma variável aleatória $W \sim \text{beta}(0.9, 90)$, e fazendo $x_{40} = 1.4$. Esses dados, contaminados e sem contaminação, foram ajustados utilizando o modelo de regressão PL-N e PL-t₍₅₎, os resultados são apresentados nas Figuras 3.1 e 3.2.

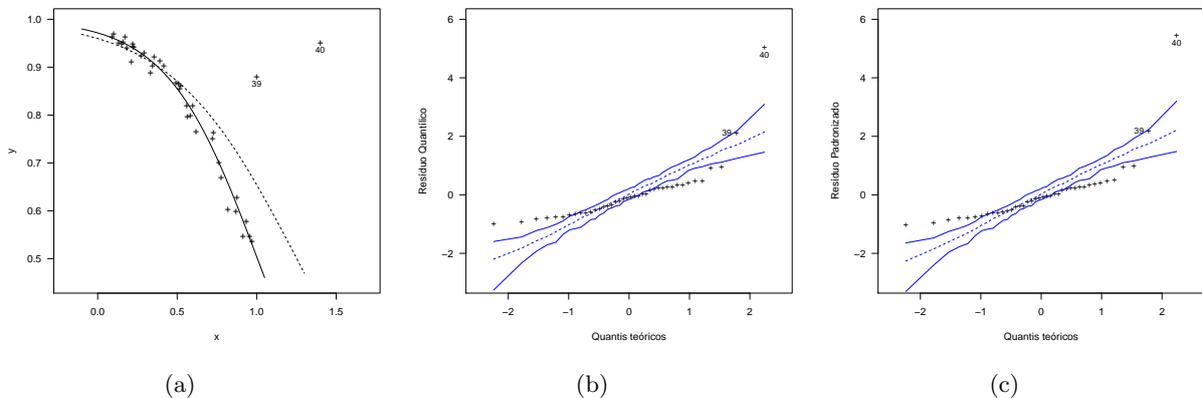


Figura 3.1: Diagrama de dispersão com as curvas ajustadas (via máxima verossimilhança) para os dados completos (linha sólida) e para os dados contaminados (linha tracejada) (a), e gráficos dos resíduos quantílicos (b) e padronizados (c) com envelopes simulados para os dados contaminados; modelo de regressão PL-N.

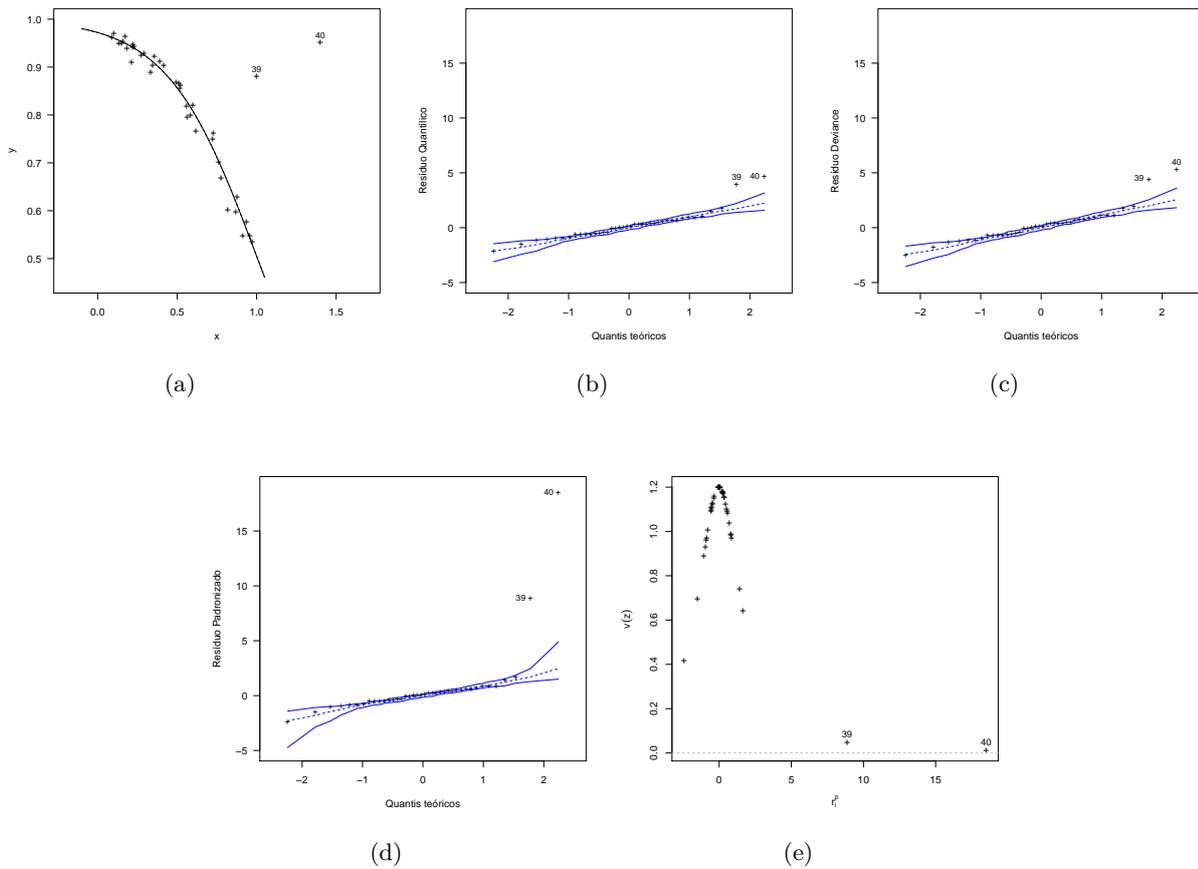


Figura 3.2: Diagrama de dispersão com as curvas ajustadas (via máxima verossimilhança) para os dados completos (linha sólida) e para os dados contaminados (linha tracejada) (a), gráficos dos resíduos quantílico (b), deviance (c) e padronizados (d) com envelopes simulados para os dados contaminados, e gráfico da função $v(z)$ versus o resíduo padronizado (e) para os dados contaminados; o modelo de regressão $PL-t_{(5)}$.

Conforme esperado, o ajuste do modelo de regressão $PL-N$ é influenciado pelas observações 38 e 40, e os resíduos conseguem detectar que tais observações são pontos atípicos; veja, por exemplo, as Figuras 3.1(b) e 3.1(c). Além disso, a Figura 3.1(a) mostra que a reta ajustada via máxima verossimilhança penalizada muda consideravelmente quando os dados são contaminados, indicando que as estimativas dos parâmetros do modelo são sensíveis a essas observações.

Por outro lado, quando o modelo de regressão $PL-t_{(5)}$ é considerado para analisar os dados, notamos que as observações 39 e 40, apesar de serem detectadas como casos atípicos pelos resíduos propostos, não influenciam as estimativas dos parâmetros do modelo (note, por exemplo, que as retas estimadas dos dados com e sem contaminação são praticamente idênticas, veja a Figura 3.2(a)). Os gráficos de resíduos com envelopes simulados mostram que o modelo de regressão $PL-t_{(5)}$ ajusta bem a maioria das observações (ver Figuras 3.2(b), 3.2(c) e 3.2(d)). Além disso, os resíduos quantílico e deviance apresentam valores próximos para as observações 39 e 40, ao contrário do resíduo padronizado. O resíduo padronizado para a observação 40 (que é um ponto de alavanca) é maior do que o da observação 39. Isso ocorre, essencialmente, porque este resíduo leva em consideração a matriz \mathbf{H} . Ainda, a Figura 3.2(e) apresenta o gráfico da função $v(z)$ versus o resíduo r_i^p . Observe que observações com resíduos grandes têm pouco peso na estimação dos β 's.

Finalmente, esse experimento mostra que os resíduos propostos são eficazes na identificação de observações atípicas, sendo o resíduo padronizado mais indicado para a detecção de pontos de alavanca. Além disso, observamos que o modelo de regressão $PL-t_{(5)}$ é menos sensível a observações discrepantes do que o modelo de regressão $PL-N$, conforme esperado. Isso se dá porque a função $v(z)$ retorna pesos pequenos para observações distantes do centro da distribuição de Z e essa

função é levada em consideração no processo de estimação via máxima verossimilhança penalizada dos modelos *power logit*. Esse experimento foi feito em outros modelos de regressão *power logit* e resultados similares foram observados.

Experimentação II

No segundo experimento, nosso objetivo é verificar se os resíduos propostos conseguem detectar a suposição incorreta de dispersão constante. Para tanto, geramos $n = 50$ observações do modelo de regressão *power logit*

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = -1.5 + 2.2x_i,$$

$$\log \sigma_i = -2 + 1.5s_i,$$

para $i = 1, \dots, 40$, $\lambda = 0.5$, em que x_i são obtidos de uma variável aleatória $U \sim U(0, 1)$ e s_i é uma variável indicadora, que recebe o valor 0 para as 30 primeiras observações e 1 para as demais. Em seguida, ajustamos o modelo de regressão *power logit* com dispersão constante. Consideramos, neste experimento, os modelos $PL-PE_{(1.5)}$ e $PL-slash_{(2.3)}$. A Figura 3.3 apresenta os gráficos de resíduos *versus* os índices das observações para os dois modelos em estudo.

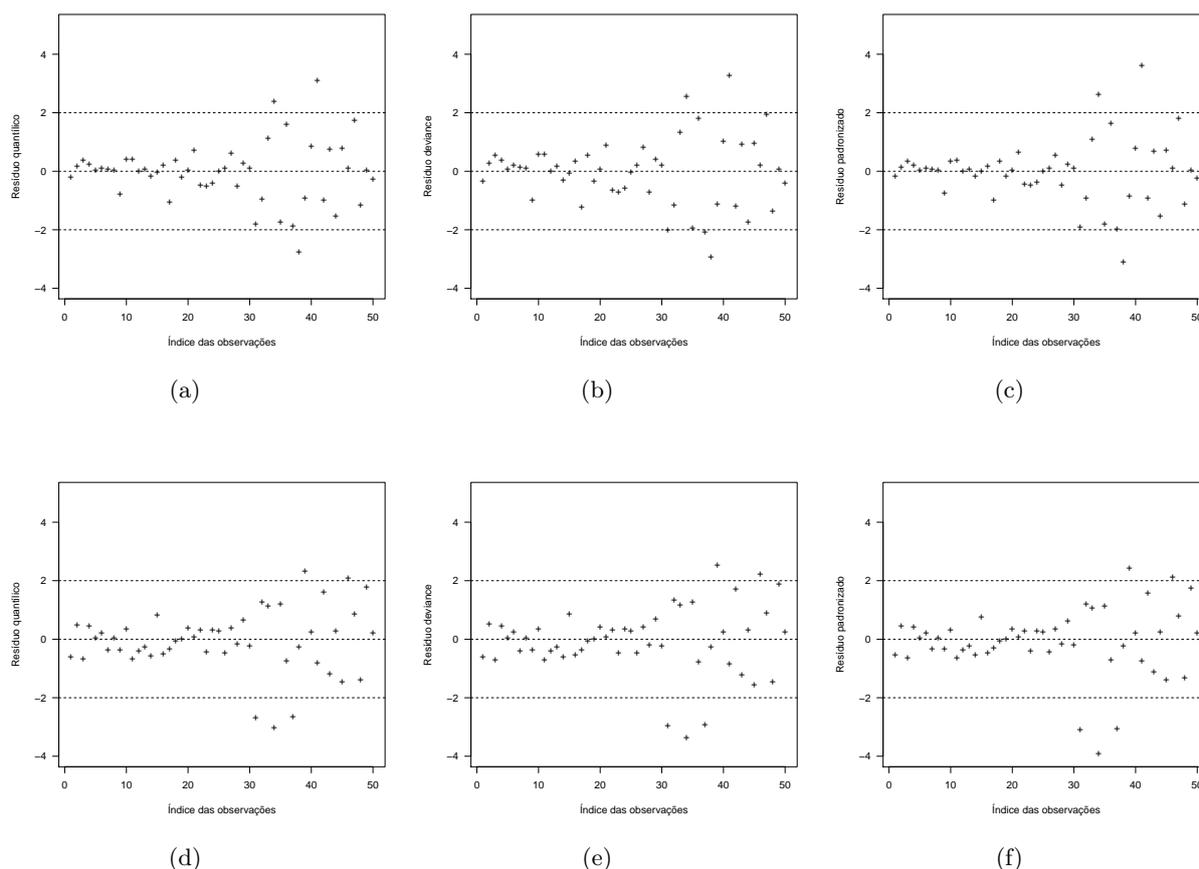


Figura 3.3: Gráficos de resíduo *versus* o índice das observações para os ajustes dos modelos de regressão $PL-PE_{(1.5)}$ (primeira linha) e $PL-slash_{(2.3)}$ (segunda linha).

Os gráficos mostram que os resíduos conseguem detectar a presença de dispersão não constante tanto no modelo $PL-PE_{(1.5)}$ (Gráficos 3.3(a), 3.3(b) e 3.3(c)) quanto no modelo $PL-slash_{(2.3)}$ (Gráficos 3.3(d), 3.3(e) e 3.3(f)). Experimentos com outros modelos de regressão da classe *power logit*, não apresentados aqui, resultaram em comportamentos similares.

Experimentação III

Finalmente, geramos $n = 50$ observações do modelo de regressão *power logit*

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = 2.5 - 2.4x_i^2,$$

$$\log \sigma_i = -1.2 + 1.5s_i,$$

para $i = 1, \dots, 40$, $\lambda = 1$, em que x_i são obtidos de uma variável aleatória $U \sim U(-1.5, 1.5)$ e s_i obtidos de uma variável $W \sim U(0, 1)$. Em seguida, ajustamos o modelo de regressão *power logit* considerando a componente sistemática da forma

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_1 + \beta_2 x_i,$$

$$\log \sigma_i = \tau_1 + \tau_2 s_i,$$

para $i = 1, \dots, 40$. Isto é, desconsideramos o termo quadrático na covariável x . Consideramos, nesse experimento, os modelos PL-SN_(1.6) e PL-Hyp_(1.5). A Figura 3.4 apresenta os gráficos de resíduo *versus* a covariável x para os dois modelos em estudo.

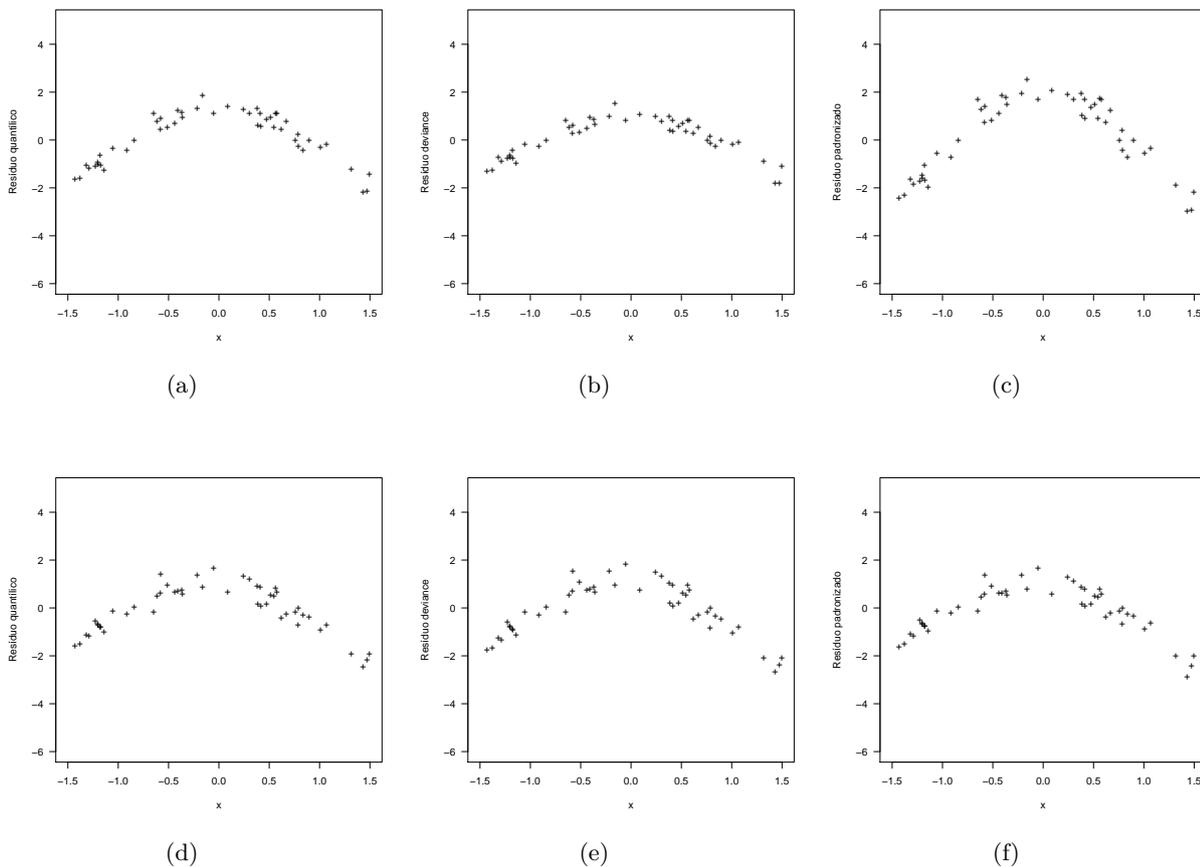


Figura 3.4: Gráficos de resíduo *versus* a covariável x para os ajustes dos modelos de regressão PL-SN_(1.6) (primeira linha) e PL-Hyp_(1.5) (segunda linha).

Os resíduos sugerem que os modelos não ajustam bem os dados e, em todos eles, os resíduos *versus* a covariável x apresentam o padrão de uma curva, sugerindo a necessidade de se considerar um termo quadrático para a variável x . Esse padrão é observado tanto nos resíduos para o modelo de regressão PL-SN_(1.6) (Figuras 3.4(a), 3.4(b) e 3.4(c)) quanto para o modelo de regressão PL-Hyp_(1.5) (Figuras 3.4(d), 3.4(e) e 3.4(f)). Novamente, experimentações com outros modelos de regressão na

classe *power logit* resultaram em comportamentos similares.

As aplicações em dados simulados apresentadas nesta seção indicam que os resíduos propostos para o modelo de regressão *power logit* são eficientes para detectar diferentes falhas no ajuste do modelo. No caso de observações atípicas (*outliers*, pontos de alavanca, etc.), o resíduo padronizado, r_i^p , apresentou desempenho melhor do que os demais. Adicionalmente, estudos de simulação para avaliar o comportamento destes resíduos em amostras de tamanho finito foram conduzidos. Por brevidade, esses resultados não serão apresentados aqui, mas indicaram que os resíduos propostos apresentam média próxima de zero e desvio padrão próximo de um, conforme esperado. Além disso, o resíduo quantílico apresenta distribuição próxima da normal padrão, principalmente quando o tamanho da amostra é grande. A distribuição dos resíduos padronizado e deviance, no entanto, não é próxima da normal (principalmente quando a variável resposta está concentrada próxima dos extremos do intervalo unitário) e, portanto, o uso das bandas de confianças simuladas se faz necessário.

3.3.2 Ponto de alavanca generalizado

A medida de alavanca generalizada é definida por $GL_{ij} = \partial \hat{y}_i / \partial y_j$ e reflete a taxa de mudança instantânea no i -ésimo valor predito de y , quando a j -ésima observação da variável resposta sofre um aumento infinitesimal. Essa medida foi proposta por [Wei et al. \(1998\)](#) e a ideia principal do uso de GL_{ii} é avaliar a influência de y_i no seu próprio valor predito. Pontos de alavanca altos sugerem que a observação pode ser um *outlier* nas covariáveis.

Seja $\hat{\boldsymbol{\theta}}$ o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ e seja $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ o vetor de valores preditos para \mathbf{y} . A matriz de alavanca generalizada é dada por

$$GL(\boldsymbol{\theta}) = \dot{\mathbf{L}}_{\boldsymbol{\beta}} \mathbf{J}_n^{-1} \ddot{\mathbf{L}}_{\boldsymbol{\theta} \mathbf{y}},$$

em que $\dot{\mathbf{L}}_{\boldsymbol{\beta}} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}$ e $\ddot{\mathbf{L}}_{\boldsymbol{\theta} \mathbf{y}} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \mathbf{y}^T$. Na prática, trabalhamos com a matriz $GL(\hat{\boldsymbol{\theta}})$.

Note que, para $R = 1, \dots, p$, temos que

$$\frac{\partial \mu_i}{\partial \beta_R} = \frac{x_{iR}}{\dot{d}_1(\mu_i)},$$

de modo que $\partial \boldsymbol{\mu} / \partial \boldsymbol{\beta} = \mathbf{T}_1 \mathbf{X}$ e, portanto, $\dot{\mathbf{L}}_{\boldsymbol{\beta}} = [\mathbf{T}_1 \mathbf{X}, \mathbf{0}_{n,q}, \mathbf{0}_{n,1}]$, sendo $\mathbf{0}_{a,b}$ uma matriz de zeros de dimensão $a \times b$. Além disso, para $j = 1, \dots, n$, temos que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_R \partial y_j} &= \frac{\partial}{\partial y_j} \sum_{i=1}^n \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i) \frac{1}{\dot{d}_1(\mu_i)} x_{iR} \\ &= \frac{\lambda}{\sigma_j \mu_j (1 - \mu_j^\lambda)} \frac{1}{\dot{d}_1(\mu_j)} x_{jR} \frac{\lambda}{\sigma_j y_j (1 - y_j^\lambda)} [v(z_j) + z_j v'(z_j)], \quad R = 1, \dots, p, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \tau_l \partial y_j} &= \frac{\partial}{\partial y_j} \sum_{i=1}^n \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1] \frac{1}{\dot{d}_2(\sigma_i)} s_{il} \\ &= \frac{1}{\dot{d}_2(\sigma_j)} s_{jl} \frac{\lambda}{\sigma_j y_j (1 - y_j^\lambda)} \frac{1}{\sigma_j} z_j [2v(z_j) + z_j v'(z_j)], \quad l = 1, \dots, q, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \lambda \partial y_j} &= \frac{\partial}{\partial y_j} \sum_{i=1}^n \left\{ \frac{1}{\lambda} + \frac{y_i^\lambda \log y_i}{1 - y_i^\lambda} - \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{\log y_i}{(1 - y_i^\lambda)} - \frac{\log \mu_i}{(1 - \mu_i^\lambda)} \right] \right\} \\ &= \frac{\lambda}{\sigma_j y_j (1 - y_j^\lambda)} \left\{ \frac{\sigma_j y_i^\lambda (1 + \lambda \log y_j - y_j^\lambda)}{\lambda (1 - y_j^\lambda)} - \frac{1}{\sigma_j} \left(\frac{\log y_j}{1 - y_j^\lambda} - \frac{\log \mu_j}{1 - \mu_j^\lambda} \right) [v(z_j) + z_j v'(z_j)] \right. \\ &\quad \left. - z_j v(z_j) \frac{y_i^\lambda (\lambda \log y_j - 1) + 1}{\lambda (1 - y_j^\lambda)} \right\}. \end{aligned}$$

Dessa forma, defina $\mathbf{D}_{\boldsymbol{\beta}} = \text{diag}\{\mu_1^*, \dots, \mu_n^*\}$, $\mathbf{D}_y = \text{diag}\{y_1^*, \dots, y_n^*\}$, $\mathbf{W}_{\boldsymbol{\beta}} = \text{diag}\{v(z_1) +$

$z_1 v'(z_1), \dots, v(z_n) + z_n v'(z_n)\}$, $\mathbf{W}_\tau = \text{diag}\{\sigma_1^{-1} z_1 [2v(z_1) + z_1 v'(z_1)], \dots, \sigma_n^{-1} z_n [2v(z_n) + z_n v'(z_n)]\}$, $\mathbf{W}_\lambda = \text{diag}\{w_1^\lambda, \dots, w_n^\lambda\}$, em que $y_i^* = \lambda / [\sigma_i y_i (1 - y_i^\lambda)]$ e

$$w_i^\lambda = \frac{\sigma_j y_i^\lambda (1 + \lambda \log y_j - y_j^\lambda)}{\lambda (1 - y_j^\lambda)} - \frac{1}{\sigma_j} \left(\frac{\log y_j}{1 - y_j^\lambda} - \frac{\log \mu_j}{1 - \mu_j^\lambda} \right) [v(z_j) + z_j v'(z_j)] - z_j v(z_j) \frac{y_i^\lambda (\lambda \log y_j - 1) + 1}{\lambda (1 - y_j^\lambda)},$$

para $i = 1, \dots, n$. Assim,

$$\ddot{\mathbf{L}}_{\theta y} = \begin{bmatrix} \mathbf{X}^\top \mathbf{T}_1 \mathbf{D}_\beta \mathbf{D}_y \mathbf{W}_\beta \\ \mathbf{S}^\top \mathbf{T}_2 \mathbf{D}_y \mathbf{W}_\tau \\ \mathbf{1}_n^\top \mathbf{D}_y \mathbf{W}_\lambda \end{bmatrix}.$$

Finalmente, o gráfico dos elementos diagonais de $\text{GL}(\hat{\boldsymbol{\theta}})$ versus o índice das observações pode revelar pontos com grande influência nos seus próprios valores preditos.

3.4 Influência local no modelo de regressão *power logit*

Um dos principais tópicos na análise de diagnóstico é a detecção de observações influentes, isto é, pontos que exercem variações desproporcionais nas estimativas do modelo e até mesmo na significância dos coeficientes. Existem diferentes medidas propostas na literatura que têm como objetivo detectar observações influentes. Neste trabalho, consideramos o método de influência local, proposto por Cook (1986). Basicamente, a ideia é avaliar a influência conjunta das observações sob pequenas perturbações nos dados ou no modelo. Essa metodologia tem sido utilizada em diferentes modelos (ver, por exemplo, Leiva *et al.* (2007), Espinheira *et al.* (2008) e Lemonte e Bazán (2016), Queiroz (2018)).

Seja $\boldsymbol{\omega}$ um vetor k -dimensional de perturbações assumindo valores em um conjunto aberto $\Omega \subset \mathbb{R}^k$; em geral, tem-se que $k = n$. Denote por $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ a função de log-verossimilhança do modelo perturbado. Assuma que existe um vetor de não perturbação $\boldsymbol{\omega}_0 \in \Omega$ tal que $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$. Se o interesse é avaliar a influência de pequenas perturbações na estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$, pode-se utilizar a a função $LD_\omega = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_\omega)]$, chamada afastamento pela verossimilhança (*likelihood displacement*), em que $\hat{\boldsymbol{\theta}}_\omega$ é a estimativa de máxima verossimilhança sob $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$.

A ideia de influência local é analisar o comportamento local de LD_ω em torno de $\boldsymbol{\omega}_0$. Para isso, o método estuda a curvatura do gráfico de $LD_{\boldsymbol{\omega}_0 + a\mathbf{h}}$ versus a , em que $a \in \mathbb{R}$ e \mathbf{h} é um vetor unitário em \mathbb{R}^k diferente de zero. Cook (1986) mostrou que a curvatura normal na direção \mathbf{h} é dada por

$$C_{\mathbf{h}} = 2|\mathbf{h}^\top \boldsymbol{\Delta}^\top \ddot{\ell}_{\theta\theta}^{-1} \boldsymbol{\Delta} \mathbf{h}|, \quad (3.12)$$

em que $\boldsymbol{\Delta}$ é a matriz de perturbação de dimensão $(p + q + 1) \times k$, dada por

$$\boldsymbol{\Delta} = \left. \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \quad \text{e} \quad \ddot{\ell}_{\theta\theta} = \left. \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

A expressão dada em (3.12) é denominada influência local da perturbação na direção \mathbf{h} sobre a estimativa de $\boldsymbol{\theta}$. Uma das formas de estudar $C_{\mathbf{h}}$ é considerar a direção \mathbf{h}_{\max} que corresponde à maior curvatura normal $C_{\mathbf{h}_{\max}}$. Especificamente, a curvatura normal $C_{\mathbf{h}_{\max}}$ equivale ao maior autovalor da matriz $-\boldsymbol{\Delta}^\top \ddot{\ell}_{\theta\theta}^{-1} \boldsymbol{\Delta}$ e a direção \mathbf{h}_{\max} corresponde ao autovetor associado a esse maior autovalor. O gráfico de índices de \mathbf{h}_{\max} pode indicar as observações que têm grande influência em LD_ω sob pequenas perturbações. Além disso, os gráficos de \mathbf{h}_{\max} versus as covariáveis podem ser úteis para identificar padrões atípicos.

Uma outra ideia é estudar a curvatura na direção do i -ésimo indivíduo (Lesaffre e Verbeke, 1998) através de

$$C_i = 2|\boldsymbol{\Delta}_i^\top \ddot{\ell}_{\theta\theta}^{-1} \boldsymbol{\Delta}_i|, \quad i = 1, \dots, n, \quad (3.13)$$

em que $\boldsymbol{\Delta}_i$ é a i -ésima coluna da matriz $\boldsymbol{\Delta}$. O gráfico de índices de C_i pode indicar observações

individualmente influentes.

Para o modelo de regressão *power logit*, temos interesse em avaliar a influência de pequenas perturbações na estimação dos parâmetros associados à mediana e à dispersão. Assim, consideramos o parâmetro de assimetria, λ , conhecido. Na prática, esse parâmetro é substituído por sua estimativa de máxima verossimilhança penalizada. Assim, seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$. A matriz $\boldsymbol{\Delta}$ tem dimensão $(p + q) \times n$ e é dada por

$$\boldsymbol{\Delta} = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_1 \partial \beta_1} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_n \partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_1 \partial \beta_p} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_n \partial \beta_p} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_1 \partial \tau_1} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_n \partial \tau_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_1 \partial \tau_q} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_n \partial \tau_q} \end{bmatrix}.$$

Note que essa matriz pode ser particionada e escrita na forma $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\tau^\top)^\top$, sendo

$$\begin{aligned} \boldsymbol{\Delta}_\beta &= \left. \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \\ \boldsymbol{\Delta}_\tau &= \left. \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}. \end{aligned}$$

De maneira geral, definimos como perturbação qualquer mudança feita nas suposições do modelo e/ou nos dados de modo a alterar, de forma substancial, os resultados da análise (para uma discussão mais aprofundada, consultar [Billor e Loynes \(1993\)](#)). Neste trabalho, consideramos dois esquemas de perturbação: ponderação de casos e perturbação nas covariáveis. Nesta sessão, as quantidades avaliadas na estimativa de máxima verossimilhança penalizada serão escritas com acento til.

Ponderação de casos

A ponderação de casos é um dos principais esquemas de perturbação e tem sido utilizada em diversos trabalhos. Neste esquema de perturbação, seja $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ um vetor $n \times 1$ de ponderações, em que $0 \leq \omega_i \leq 1$, $i = 1, \dots, n$. O vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$. Neste caso, a função de log-verossimilhança perturbada é da forma

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \ell_i(\mu_i, \sigma_i, \lambda),$$

em que $\ell_i(\mu_i, \sigma_i, \lambda)$ é definida em (3.2). Observe que se $\omega_i = 0$ temos que o i -ésimo caso é eliminado. Neste esquema de perturbação, a curvatura normal é dada por (3.12), em que $\ddot{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ é dada por $-\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$. Dessa forma, usando as quantidades dadas em (C.1), a matriz $\boldsymbol{\Delta}$ é dada por

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{X}^\top \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 \widehat{\mathbf{D}}_\beta \\ \mathbf{S}^\top \widehat{\mathbf{T}}_2 \widehat{\mathbf{D}}_\tau \end{bmatrix}.$$

Perturbação nas variáveis explicativas

Neste esquema de perturbação, estamos interessados em avaliar a influência, individual ou coletiva, das variáveis explicativas (contínuas) na estimação dos parâmetros do modelo.

Perturbação individual das covariáveis da mediana

Assuma que é feita uma perturbação aditiva em uma variável explanatória contínua da mediana, digamos \mathbf{x}_j , com $j = 1, \dots, p$. Seguindo [Thomas e Cook \(1989\)](#), trocamos x_{ij} por

$$x_{ij\omega} = x_{ij} + \sigma_x \omega_i, \quad (3.14)$$

sendo σ_x o fator de escala dado pelo desvio padrão amostral de \mathbf{x}_j . Neste caso, por exemplo, se $j \neq p$

$$\eta_{1i\omega} = \beta_1 x_{i1} + \dots + \beta_j (x_{ij} + \sigma_x \omega_i) + \dots + \beta_p x_{ip} = \mathbf{x}_{i\omega}^\top \boldsymbol{\beta},$$

em que $\mathbf{x}_{i\omega} = (x_{i1}, \dots, x_{ij\omega}, \dots, x_{ip})^\top$. A função de log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\mu_{i\omega}, \sigma_i, \lambda), \quad (3.15)$$

em que $\mu_{i\omega} = d_1^{-1}(\eta_{1i\omega})$. O vetor de não perturbação é dado por $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$. Com base nas expressões dadas em (C.3) e (C.5) na Seção C.1.2, temos que

$$\Delta = \begin{bmatrix} -\sigma_x \hat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1 \widehat{\mathbf{T}}_1 + \sigma_x \mathbf{c}_\beta^j \widehat{\boldsymbol{\mu}}^{*\top} \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 \\ -\sigma_x \hat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \end{bmatrix},$$

em que o vetor \mathbf{c}_β^j foi definido na Seção C.1.2. As demais quantidades foram definidas anteriormente.

Perturbação individual das covariáveis da dispersão

Agora, considere uma perturbação aditiva em uma variável explanatória contínua da dispersão, digamos \mathbf{s}_k , com $k = 1, \dots, q$. Seguindo [Thomas e Cook \(1989\)](#), trocamos s_{ik} por

$$s_{ik\omega} = s_{ik} + \sigma_s \omega_i, \quad (3.16)$$

sendo σ_s o fator de escala dado pelo desvio padrão amostral de \mathbf{s}_k . Neste caso, por exemplo, se $k \neq q$, temos que

$$\eta_{2i\omega} = \tau_1 s_{i1} + \dots + \tau_k (s_{ik} + \sigma_s \omega_i) + \dots + \tau_q s_{iq} = \mathbf{s}_{i\omega}^\top \boldsymbol{\tau},$$

em que $\mathbf{s}_{i\omega} = (s_{i1}, \dots, s_{ik\omega}, \dots, s_{iq})^\top$. A função de log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\mu_i, \sigma_{i\omega}, \lambda), \quad (3.17)$$

em que $\sigma_{i\omega} = d_2^{-1}(\eta_{2i\omega})$. O vetor de não perturbação é $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$.

Com base nas expressões dadas em (C.7) e (C.9) na Seção C.1.2, temos que

$$\Delta = \begin{bmatrix} -\sigma_s \hat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \\ -\sigma_s \hat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2 \widehat{\mathbf{T}}_2 + \sigma_s \mathbf{c}_\tau^k \widehat{\boldsymbol{\sigma}}^{*\top} \widehat{\mathbf{T}}_2 \end{bmatrix},$$

em que o vetor \mathbf{c}_τ^k foi definido na Seção C.1.2. As demais quantidades foram definidas anteriormente.

Perturbação simultânea das covariáveis da mediana e da dispersão

Assuma que é feita uma perturbação aditiva em uma variável explanatória contínua particular da mediana e da dispersão, digamos \mathbf{x}_j ($j = 1, \dots, p$) e \mathbf{s}_k ($k = 1, \dots, q$), respectivamente. Seguindo

Thomas e Cook (1989), trocamos x_{ij} e s_{ik} , respectivamente, por

$$\begin{aligned}x_{ij\omega} &= x_{ij} + \sigma_x \omega_i, \\s_{ik\omega} &= s_{ik} + \sigma_s \omega_i,\end{aligned}\tag{3.18}$$

sendo σ_x e σ_s os fatores de escala dados pelos desvios padrão amostrais de \mathbf{x}_j e \mathbf{s}_k , respectivamente. Neste caso, temos que

$$\begin{aligned}\eta_{1i\omega} &= \beta_1 x_{i1} + \cdots + \beta_j (x_{ij} + \sigma_x \omega_i) + \cdots + \beta_p x_{ip} = \mathbf{x}_{i\omega}^\top \boldsymbol{\beta}, \\ \eta_{2i\omega} &= \tau_1 s_{i1} + \cdots + \tau_k (s_{ik} + \sigma_s \omega_i) + \cdots + \tau_q s_{iq} = \mathbf{s}_{i\omega}^\top \boldsymbol{\tau},\end{aligned}$$

em que $\mathbf{x}_{i\omega} = (x_{i1}, \dots, x_{ij\omega}, \dots, x_{ip})^\top$ e $\mathbf{s}_{i\omega} = (s_{i1}, \dots, s_{ik\omega}, \dots, s_{iq})^\top$. A função de log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\mu_{i\omega}, \sigma_{i\omega}, \lambda),\tag{3.19}$$

sendo $\mu_{i\omega} = d_1^{-1}(\eta_{1i\omega})$ e $\sigma_{i\omega} = d_2^{-1}(\eta_{2i\omega})$. O vetor de não perturbação é dado por $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$. Assim, temos

$$\boldsymbol{\Delta} = \begin{bmatrix} -\sigma_x \hat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1 \widehat{\mathbf{T}}_1 + \sigma_x \mathbf{c}_\beta^j \hat{\boldsymbol{\mu}}^{*\top} \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 - \sigma_s \hat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \\ -\sigma_s \hat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2 \widehat{\mathbf{T}}_2 + \sigma_s \mathbf{c}_\tau^k \hat{\boldsymbol{\sigma}}^{*\top} \widehat{\mathbf{T}}_2 - \sigma_x \hat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \end{bmatrix}.$$

O cálculo da matriz $\boldsymbol{\Delta}$, para este esquema de perturbação, será omitido por brevidade, mas pode ser feito de maneira análoga ao esquema de perturbação individual das covariáveis.

3.5 Caso limite: o modelo de regressão log-log

A classe de modelos de regressão *power logit* tem como caso limite, quando $\lambda \rightarrow 0$, a (nomeada) classe de modelos de regressão log-log. As distribuições log-log foram apresentadas na Definição 2.3. Nesta seção, definimos brevemente a classe de modelos de regressão log-log e alguns resultados. Considere Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes com $Y_i \sim \text{log-log}(\mu_i, \sigma_i; r)$, para $i = 1, \dots, n$. O modelo de regressão log-log é definido por (2.6) e pelos componentes sistemáticos em (3.1).

No que segue, para definir algumas quantidades relacionadas ao modelo de regressão log-log, utilizamos a mesma notação das seções anteriores, porém com modificação nas seguintes quantidades, que passam a ser definidas por

$$\begin{aligned}z_i &= -\sigma_i^{-1} \{-\log(-\log y) - [-\log(-\log \mu)]\}, \quad \mu_i^* = [\sigma_i \mu_i (-\log \mu_i)]^{-1}, \\ w_i^{(1)} &= -\mu_i^{*2} \{\sigma_i (\log \mu_i + 1) z_i v(z_i) - [v(z_i) + z_i v'(z_i)]\} \frac{1}{\dot{d}_1(\mu_i)} + \mu_i^* z_i v(z_i) \frac{\ddot{d}_1(\mu_i)}{\dot{d}_1(\mu_i)^2}, \\ w_i^{(2)} &= -\left\{ \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^2} z_i^2 [3v(z_i) + z_i v'(z_i)] \right\} \frac{1}{\dot{d}_2(\sigma_i)} + \sigma_i^* \frac{\ddot{d}_2(\sigma_i)}{\dot{d}_2(\sigma_i)^2}, \\ w_i^{(3)} &= \mu_i^* z_i [2v(z_i) + z_i v'(z_i)] / \sigma_i, \quad y_i^\dagger = -\log(-\log y_i), \quad \mu_i^\dagger = -\log(-\log \mu_i), \\ y_i^* &= [\sigma_i \mu_i (-\log y_i)]^{-1},\end{aligned}$$

para $i = 1, \dots, n$. Assim, por exemplo, a matriz \mathbf{W} que aparece no vetor escore para $\boldsymbol{\theta}$ no modelo *power logit* em (3.6), fica agora dada por $\mathbf{W} = \text{diag}\{z_1 v(z_1), \dots, z_n v(z_n)\}$, com $z_i = -\sigma_i^{-1} \{-\log(-\log y_i) - [-\log(-\log \mu_i)]\}$, para $i = 1, \dots, n$.

Dessa forma, seja, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$ o vetor de parâmetros do modelo. A função de log-verossimilhança para $\boldsymbol{\theta}$ é dada por $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\mu_i, \sigma_i)$, sendo $\ell_i(\mu_i, \sigma_i) = -\log[-\sigma_i y_i \log(y_i)] + \log\{r(z_i^2)\}$. O vetor

escore e a matriz da informação observada de θ são dados por

$$\mathbf{U}(\theta) = \begin{bmatrix} \mathbf{X}^\top \mathbf{W} \mathbf{T}_1 \boldsymbol{\mu}^* \\ \mathbf{S}^\top \mathbf{T}_2 \boldsymbol{\sigma}^* \end{bmatrix} \quad \text{e} \quad \mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \begin{bmatrix} \mathbf{J}_{\beta\beta} & \mathbf{J}_{\beta\tau} \\ \mathbf{J}_{\beta\tau}^\top & \mathbf{J}_{\tau,\tau} \end{bmatrix},$$

respectivamente, em que $\mathbf{J}_{\beta\beta} = \mathbf{X}^\top \mathbf{W}_1 \mathbf{T}_1 \mathbf{X}$, $\mathbf{J}_{\tau\tau} = \mathbf{S}^\top \mathbf{W}_2 \mathbf{T}_2 \mathbf{S}$, $\mathbf{J}_{\beta\tau} = \mathbf{X}^\top \mathbf{W}_3 \mathbf{T}_1 \mathbf{T}_2 \mathbf{S}$.

Assim como no modelo de regressão *power logit*, os estimadores de máxima verossimilhança para θ podem ser obtidos via maximização da função de log-verossimilhança através de métodos numéricos. Em geral, os algoritmos de maximização necessitam de valores iniciais para os parâmetros do modelo e, neste caso, sugerimos os mesmos pontos iniciais do modelo *power logit*, isto é, $\boldsymbol{\beta}^{(0)}$ e $\boldsymbol{\tau}^{(0)}$. Adicionalmente, a estimação dos parâmetros extras também é feita através da minimização da quantidade Υ_ζ .

No contexto de diagnóstico, os resíduos propostos para o modelo de regressão *power logit* podem ser definidos de maneira similar para o modelo de regressão log-log, fazendo as devidas adaptações. Assim, temos os resíduos quantílico, deviance e padronizado dados, respectivamente, por

$$\begin{aligned} r_i^q &= \Phi^{-1}[R(\hat{z}_i)], \\ r_i^d &= \text{sign}(\hat{z}_i) \left\{ 2 \log \left[\frac{r(0)}{r(\hat{z}_i^2)} \right] \right\}^{1/2}, \\ r_i^p &= \frac{\hat{y}_i^\dagger - \hat{\mu}_i^\dagger}{\hat{\sigma}_i \sqrt{\hat{\xi}_r \{1 - (\hat{d}_r \hat{\xi}_r)^{-1} \hat{h}_{ii}\}}}, \end{aligned}$$

para $i = 1, \dots, n$. A matriz de alavancagem generalizada é dada por $\text{GL}(\theta) = \dot{\mathbf{L}}_\beta \mathbf{J}_n^{-1} \ddot{\mathbf{L}}_{\theta y}$, em que $\dot{\mathbf{L}}_\beta = [\mathbf{T}_1 \mathbf{X}, \mathbf{0}_{n,q}]$ e

$$\ddot{\mathbf{L}}_{\theta y} = \begin{bmatrix} \mathbf{X}^\top \mathbf{T}_1 \mathbf{D}_\beta \mathbf{D}_y \mathbf{W}_\beta \\ \mathbf{S}^\top \mathbf{T}_2 \mathbf{D}_y \mathbf{W}_\tau \end{bmatrix}.$$

De forma análoga, obtemos as matrizes Δ para cada um dos esquemas de perturbação apresentados na Seção 3.4, adaptando as respectivas quantidades definidas no início desta seção. Essas matrizes são apresentadas a seguir.

- Ponderação de casos

$$\Delta = \begin{bmatrix} \mathbf{X}^\top \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 \widehat{\mathbf{D}}_\beta \\ \mathbf{S}^\top \widehat{\mathbf{T}}_2 \widehat{\mathbf{D}}_\tau \end{bmatrix}.$$

- Perturbação individual das covariáveis da mediana

$$\Delta = \begin{bmatrix} -\sigma_x \hat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1 \widehat{\mathbf{T}}_1 + \sigma_x \mathbf{c}_\beta^j \hat{\boldsymbol{\mu}}^{*\top} \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 \\ -\sigma_x \hat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \end{bmatrix}.$$

- Perturbação individual das covariáveis da dispersão

$$\Delta = \begin{bmatrix} -\sigma_s \hat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \\ -\sigma_s \hat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2 \widehat{\mathbf{T}}_2 + \sigma_s \mathbf{c}_\tau^k \hat{\boldsymbol{\sigma}}^{*\top} \widehat{\mathbf{T}}_2 \end{bmatrix}.$$

- Perturbação simultânea das covariáveis da mediana e da dispersão

$$\Delta = \begin{bmatrix} -\sigma_x \hat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1 \widehat{\mathbf{T}}_1 + \sigma_x \mathbf{c}_\beta^j \hat{\boldsymbol{\mu}}^{*\top} \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 - \sigma_s \hat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \\ -\sigma_s \hat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2 \widehat{\mathbf{T}}_2 + \sigma_s \mathbf{c}_\tau^k \hat{\boldsymbol{\sigma}}^{*\top} \widehat{\mathbf{T}}_2 - \sigma_x \hat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2 \end{bmatrix}.$$

Podemos observar que tomando o limite, quando $\lambda \rightarrow 0$, das quantidades definidas para o modelo de regressão *power logit* obtemos as expressões para o modelo de regressão log-log (desde

que o limite exista e esteja bem definido). Este modelo é simples e tem um parâmetro a menos, o que o leva a ser bastante atrativo. Além disso, observamos em aplicação prática que, para alguns conjuntos de dados, a estimativa de λ é bem pequena (próxima de zero). Nestes casos, sugerimos o ajuste do modelo limite, por ser mais simples e parcimonioso. O desenvolvimento dos principais resultados referentes a esse modelo é apresentado no Apêndice F.

3.6 Aplicação

Nesta aplicação, consideramos dados sobre práticas de gestão de risco de 73 empresas. Esses dados foram introduzidos por Schmit e Roth (1990) e analisados recentemente em Ribeiro e Ferrari (2022) utilizando estimadores robustos em um modelo de regressão beta. Os dados estão disponíveis na página pessoal do Professor E. Frees (Wisconsin School of Business Research)¹. A variável resposta (y) é definida como o total de prêmios de seguro mais perdas não seguradas como uma proporção do ativo total da firma. Valores menores de y são atribuídos às empresas que apresentam um bom desempenho de gestão de risco. Consideramos duas variáveis explicativas: `ind_risco`, uma medida de risco do ramo de atividade da firma, e `log_at`, o logaritmo do ativo total da firma. Algumas estatísticas descritivas da variável resposta são apresentadas na Tabela 3.3. Notamos que a variável y é concentrada próximo de zero (mediana igual a 0.061) e o valor máximo é 0.976.

Tabela 3.3: Medidas resumo para as variáveis y , `ind_risco` e `log_at`.

	Mínimo	Q_1	Mediana	Média	Q_3	Máximo	Desvio padrão
y	0.002	0.035	0.061	0.110	0.127	0.976	0.162
<code>ind_risco</code>	0.090	0.330	0.340	0.418	0.500	1.220	0.216
<code>log_at</code>	5.270	7.650	8.270	8.332	8.950	10.600	0.963

Ajustaremos o modelo de regressão *power logit* aos dados em estudo. Assim, assumimos que $Y_i \sim \text{PL}(\mu_i, \sigma_i, \lambda; r)$, para $i = 1, \dots, 73$, são variáveis aleatórias independentes, tais que

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1 \text{ind_risco}_i + \beta_2 \text{log_at}_i,$$

$$\log \sigma_i = \tau_0 + \tau_1 \text{ind_risco}_i + \tau_2 \text{log_at}_i.$$

Diferentes modelos na classe *power logit* foram considerados e o que apresentou o melhor ajuste foi o modelo PL-slash. Este modelo foi o que apresentou menor valor para a medida Υ_ζ e melhor análise de resíduos. As estimativas para os parâmetros do modelo são apresentadas Tabela 3.4.

Tabela 3.4: Estimativas (*Est.*), erros padrão assintóticos (*e.p.*) e valores p para o modelo de regressão PL-slash com dispersão variável.

	Est.	e.p.	valor p
$\log[\mu/(1 - \mu)]$			
intercepto	3.822	1.020	< 0.001
<code>ind_risco</code>	2.312	0.806	0.004
<code>log_at</code>	-0.908	0.122	< 0.001
$\log \sigma$			
intercepto	-0.569	0.789	0.471
<code>ind_risco</code>	0.366	0.541	0.498
<code>log_at</code>	0.074	0.100	0.455
λ	2.04	1.196	
ζ	1.88		
AIC	-232.1		

Conforme observamos na Tabela 3.4, nenhuma das covariáveis é estatisticamente significante

¹Disponível em <https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/CSVData/RiskSurvey.csv>

para explicar a dispersão do modelo sugerindo, portanto, que um modelo com dispersão constante é mais indicado. Dessa forma, assumiremos que $Y_i \sim \text{PL-slash}(\mu_i, \sigma, \lambda; r)$, para $i = 1, \dots, 73$. As estimativas dos parâmetros do modelo são apresentadas na Tabela 3.5. As covariáveis `ind_risco` e `log_at` são estatisticamente significantes para explicar a mediana da variável resposta, considerando os níveis de significância usuais (por exemplo, 5%). Note que existe uma relação positiva entre a mediana da variável resposta e a variável `ind_risco`, relacionada ao risco da empresa; e uma relação negativa entre a mediana da variável resposta e a variável `log_at`, isto é, o ativo total da empresa.

Tabela 3.5: Estimativas (*Est.*), erros padrão assintóticos (*e.p.*) e valores *p* para o modelo de regressão PL-slash com dispersão constante.

	Est.	e.p.	valor <i>p</i>
$\log[\mu/(1-\mu)]$			
intercepto	3.867	0.999	< 0.01
ind_risco	2.133	0.584	< 0.01
log_at	-0.905	0.112	< 0.01
$\log \sigma$	0.133	0.497	0.788
λ	1.79	1.01	
ζ	2.29		
AIC	-234.0		

A estimativa de λ é 1.79 e o intervalo de confiança assintótico de 95% baseado na estatística de razão de verossimilhanças perfilada penalizada é (0.59, 8.50), indicando que o modelo de regressão GJS-slash ($\lambda = 1$) pode ser adequado (veja a Figura 3.5; a linha tracejada horizontal fornece o intervalo de confiança de 95%).

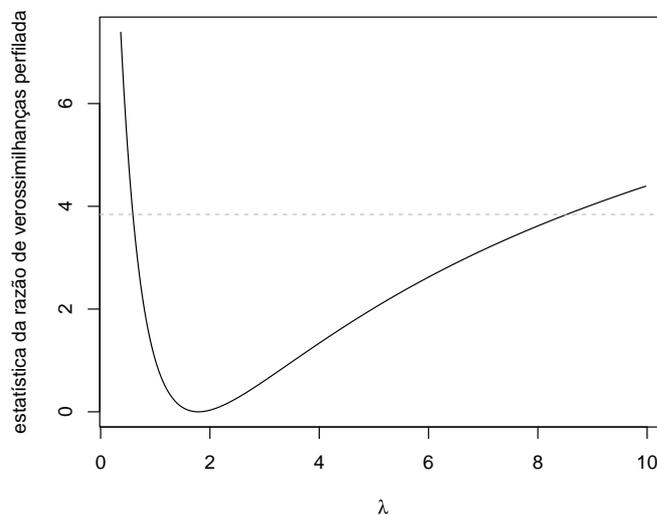


Figura 3.5: Gráfico da estatística da razão de verossimilhanças penalizadas perfilada para λ , $W_p^*(\lambda)$.

Os gráficos de resíduos com envelopes simulados para os modelos PL-slash e GJS-slash com dispersão constante são apresentados na Figura 3.6. Os gráficos de resíduos com envelopes simulados indicam que o ajuste do modelo PL-slash é mais razoável do que o do modelo GJS-slash. O caso #15, que apresenta o maior resíduo, é referente à empresa com maior proporção dos ativos totais.

O gráfico de influência considerando o esquema de ponderação de casos é apresentado na Figura 3.7(a). Note que o caso #10 foi o único que apareceu destacado no esquema de ponderação de casos; o caso #16, que apresentou o maior resíduo, não apareceu na análise de influência. Os gráficos de influência considerando o esquema de perturbação nas covariáveis, que não são apresentados aqui, não indicaram a presença de observações influentes. Na Figura 3.7 também são apresentados o

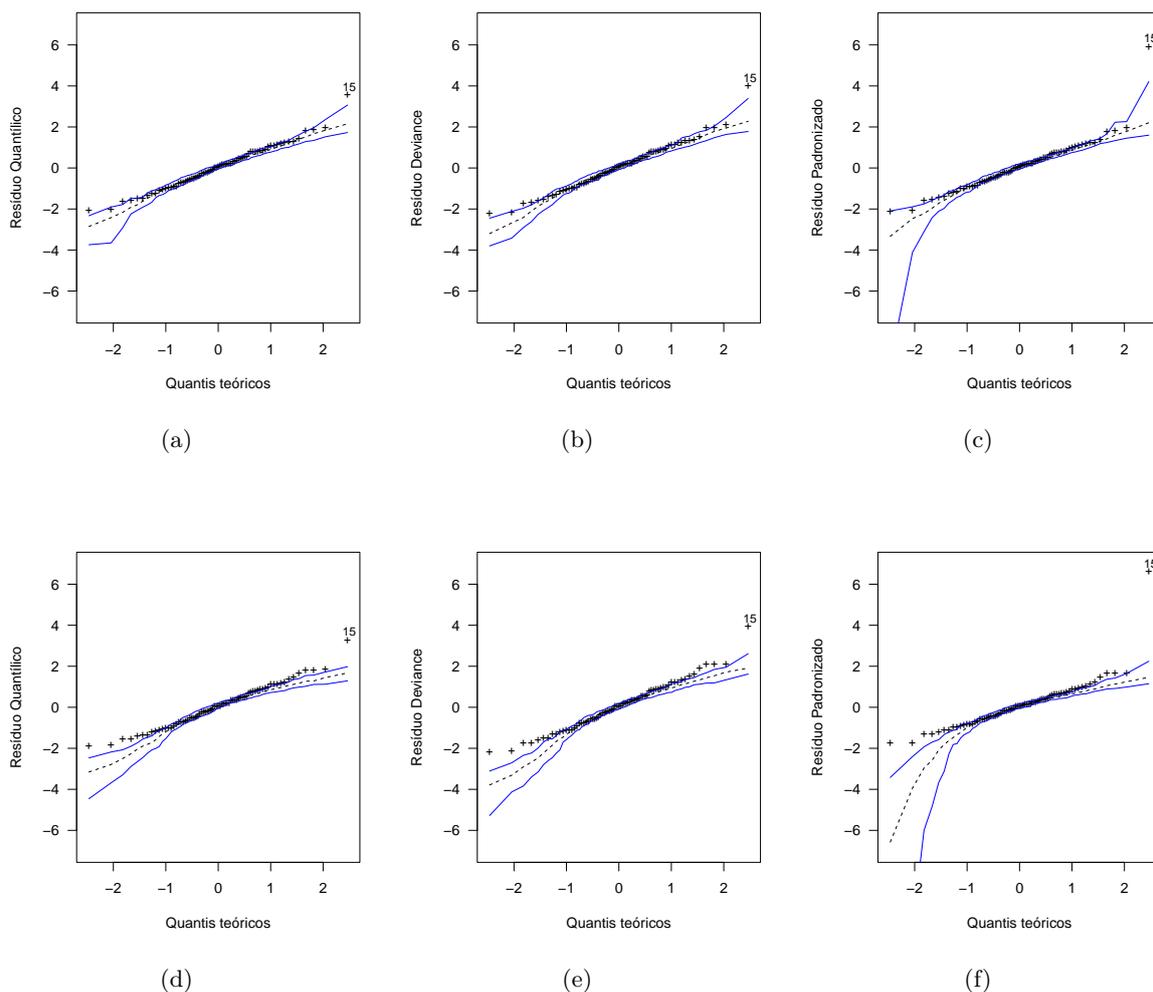


Figura 3.6: Gráficos dos resíduos quantílico (a), deviance (b) e padronizados (c) com envelopes simulados para os modelos de regressão *PL-slash* (primeira linha) e *GJS-slash* (segunda linha) com dispersão constantes.

gráfico de índices de GL_{ii} e o diagrama de dispersão de $v(z)$ versus o resíduo padronizado. Observe que a observação com menor peso na estimação foi aquela com maior resíduo (caso #15). Além disso, o gráfico índices de GL_{ii} não indicou a presença de pontos de alavanca.

Na análise feita em Ribeiro e Ferrari (2022), os casos #15, #16 e #72 foram detectados como influentes. As autoras verificaram que a exclusão individual e/ou coletiva desses casos altera de forma substancial o ajuste do modelo de regressão beta com precisão variável. A título de comparação, reproduzimos o ajuste feito em Ribeiro e Ferrari (2022). Na Figura 3.8, são apresentados os diagramas de dispersão entre y e a variável `ind_risco` com as curvas estimadas, considerando os dados completos e os dados sem *outliers*, para os modelos de regressão beta (a) e *PL-slash* (b). Os gráficos de dispersão foram produzidos definindo o valor da covariável `log_at` em sua mediana amostral. Além dos casos #15, #16 e #72, consideramos a exclusão do caso #10, detectado como possível influente no modelo *PL-slash*.

Conforme observamos na Figura 3.8, a retirada individual dos casos #15 e #72 mudam de forma considerável o ajuste do modelo beta; os casos #10 e #16 não parecem influenciar consideravelmente o ajuste. Para o modelo *PL-slash*, no entanto, a exclusão individual dos casos não altera as curvas estimadas. As curvas estimadas para os dados completos e com a exclusão dos casos #15, #16 e #72, para este modelo, são praticamente idênticas. Ribeiro e Ferrari (2022) propõem um método de estimação robusto para o modelo de regressão beta e, com essa abordagem, as estimativas robustas deixam de ser influenciadas por esses *outliers*. A ideia, intuitivamente, é atribuir pesos para cada

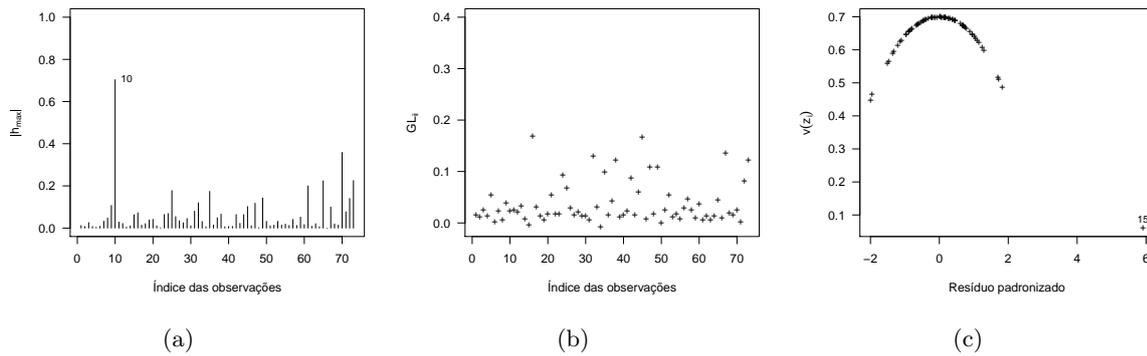


Figura 3.7: Gráficos de índices de $|h_{max}|$ para o esquema de ponderação de casos (a) e de GL_{ii} (b), e gráfico da função $v(z)$ versus o resíduo padronizado (c) para o modelo de regressão PL-slash com dispersão constante.

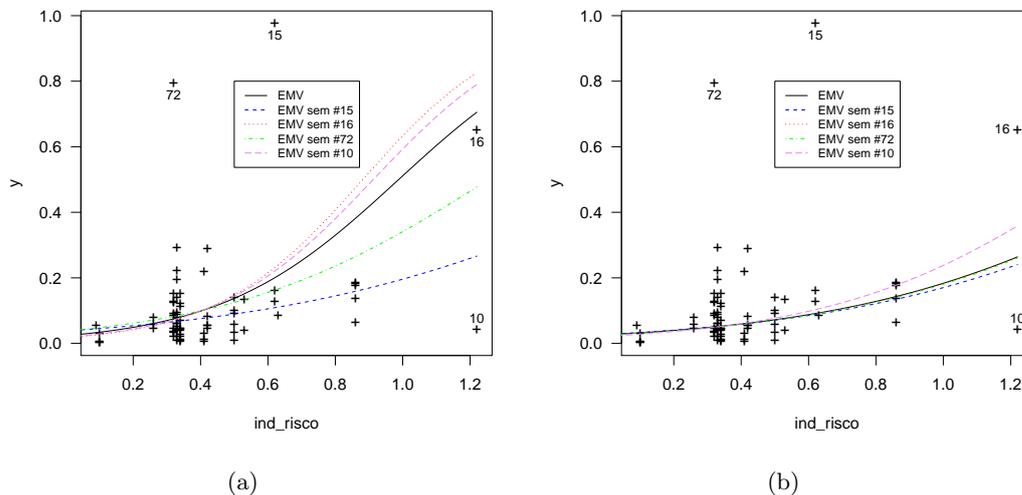


Figura 3.8: Gráficos de dispersão do ind_risco versus y , juntamente com as linhas ajustadas com base na estimativa de máxima verossimilhança (EMV) para os dados completos e os dados sem outliers, considerando o modelo de regressão beta com precisão variável (a) e o modelo de regressão PL-slash com dispersão constante (b).

observação no processo de estimação. Nesta análise, o caso #15 é o único que recebe peso baixo e as autoras concluem que isso ocorre porque os casos #16 e #72 não têm influência substancial nas estimativas de máxima verossimilhança quando o caso #15 é excluído dos dados. Da mesma forma, o caso #15 é o único que recebe peso baixo no processo de estimação do modelo PL-slash (veja Figura 3.7(a)), concordando com a análise feita em [Ribeiro e Ferrari \(2022\)](#).

Capítulo 4

O pacote PLreg

Neste capítulo, apresentamos o pacote `PLreg`, desenvolvido no *software* R, que permite o ajuste dos modelos de regressão *power logit*. Esse pacote foi desenvolvido com o objetivo de facilitar o uso dos modelos propostos neste trabalho e é similar a outros pacotes disponíveis para análise de dados de suporte limitado como, por exemplo, os pacotes `betareg` (Zeileis *et al.*, 2020) e `simplexreg` (Zhang *et al.*, 2016). Este capítulo é organizado como segue. O pacote `PLreg` é apresentado na Seção 4.1 e, na Seção 4.2, apresentamos algumas funções básicas referentes as distribuições *power logit*. A implementação do modelo de regressão *power logit* no pacote é descrita na Seção 4.3. Finalmente, na Seção 4.4, diversas aplicações em dados reais são conduzidas a fim de exemplificar o uso do pacote desenvolvido.

4.1 PLreg: apresentação e instalação

O pacote `PLreg` permite ajustar modelos de regressão *power logit* em que o gerador de densidade pode ser normal, Student-t, exponencial de potência, slash, hiperbólica, sinh-normal ou logística do tipo II. Ferramentas de diagnóstico associadas ao modelo ajustado, como os resíduos, medidas de influência local, medidas de alavancagem e de qualidades de ajuste são implementadas. O processo de estimação é baseado em máxima verossimilhança e, atualmente, o pacote suporta dois tipos de estimadores: o estimador usual de máxima verossimilhança e o estimador de máxima verossimilhança penalizada. O parâmetro de assimetria λ pode ser fixado, de modo que o pacote também permite ajustar os modelos de regressão GJS ($\lambda = 1$) e log-log ($\lambda = 0$). A principal função do pacote é `PLreg()`, que segue a abordagem padrão para implementação de modelos de regressão em R. Uma vez realizado o processo de ajuste, um objeto da classe `S3` “`PLreg`” é produzido para a qual vários métodos estão disponíveis. Detalhes técnicos são apresentados nas seções seguintes.

O pacote `PLreg` está disponível no repositório do R (CRAN) e pode ser acessado em <https://cran.r-project.org/web/packages/PLreg/index.html>. A instalação pode ser feita através da função `install.packages()`:

```
install.packages('PLreg')
```

Também é possível instalar a versão de desenvolvimento do pacote através do GitHub, fazendo:

```
# install.packages('devtools')
devtools::install_github('ffqueiroz/PLreg')
```

A página do pacote no GitHub pode ser acessada em <https://github.com/ffqueiroz/PLreg>.

4.2 A distribuição *power logit* no pacote PLreg

Atualmente, o pacote `PLreg` inclui sete membros da classe de distribuições *power logit*: as distribuições *power logit* normal, t-Student, logística do tipo I, exponencial potência, slash, hiperbólica e

seno hiperbólica normal. Neste pacote, temos as funções `dPL`, `pPL` e `qPL`, que permitem a avaliação da função densidade de probabilidade, função de distribuição acumulada e função quantílica. Além disso, a função `rPL` pode ser usada para gerar amostras aleatórias de variáveis com distribuição *power logit*. Essas funções são apresentadas a seguir.

```
dPL(x, mu, sigma, lambda, zeta = 2, family, log = FALSE)
```

```
pPL(q, mu, sigma, lambda, zeta = 2, family, lower.tail = TRUE, log.p = FALSE)
```

```
qPL(p, mu, sigma, lambda, zeta = 2, family, lower.tail = TRUE, log.p = FALSE)
```

```
rPL(n, mu, sigma, lambda, zeta = 2, family)
```

Os principais argumentos para essas funções são `mu`, `sigma`, `lambda` e `family`, especificando os parâmetros μ , σ , λ e a função geradora de densidade r ; se a função do gerador de densidade depende de um parâmetro extra, o valor deve ser especificado no argumento ζ . Essas funções também podem ser usadas para a classe log-log, basta fazer `lambda = 0`. Atualmente, o pacote `PLreg` inclui 7 membros da classe de distribuições *power logit* (ou log-log): power logit normal (`family = "NO"`), power logit t-Student (`family = "TF"`), power logit logística tipo II (`family = "LO"`), power logit exponencial potência (`family = "PE"`), power logit sinh-normal (`family = "SN"`), power logit hiperbólica (`family = "Hyp"`) e power logit slash (`family = "SLASH"`). Os argumentos `x` e `q` são vetores de quantis, `p` é um vetor de probabilidades e `n` é o número de números aleatórios a serem gerados. Outros argumentos são `log`, `log.p` e `lower.tail`. Se `log = TRUE`, o log da função de densidade de probabilidade será retornado. Se `log.p = TRUE`, então o log da função de distribuição acumulada será retornado e a função quantil será calculada para $\exp(p)$. Se `lower.tail = FALSE` então um menos a função de distribuição acumulada será retornada e a função de quantil será calculada para $1 - p$.

No que segue, apresentamos alguns usos dessas funções.

Gerando dados e plotando um histograma

```
R> mu = 0.3; sigma = 1; lambda = 2
R> set.seed(1)
R> y = rPL(1000, mu, sigma, lambda,
+       family = "SN", zeta = 2.5)
R> hist(y, prob = TRUE, breaks = 15,
+       main = "", las = 1, xlab = "y",
+       ylim = c(0, 4.5),
+       ylab = "Densidade")
R> curve(dPL(x, mu, sigma, lambda,
+       family = "SN", zeta = 2.5),
+       from = 0.01, to = 0.8, add = TRUE,
+       col = "red")
R> rug(y)
```

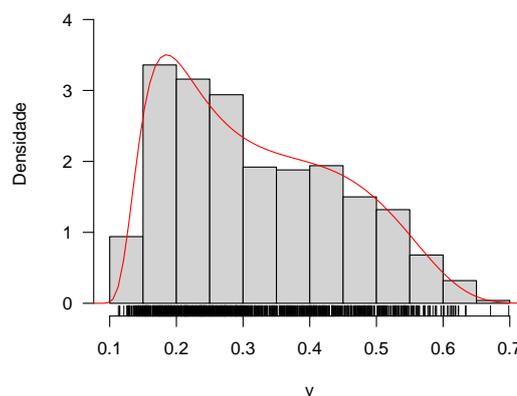


Figura 4.1: Histograma de uma amostra aleatória de uma distribuição $PL-SN_{(2.5)}$.

Função densidade de probabilidade

```
R> mu = 0.3; sigma = 1; lambda = 2
R> x = seq(0.01, 0.9, 0.01)
R> y = dPL(x, mu, sigma, lambda,
+         family = "Hyp", zeta = 2)
R> plot(x, y, type = "l", lwd = 2,
+       las = 1, ylab = expression(f(y)),
+       xlab = "y")

R> x1 = seq(0.01, 0.4, 0.01)
R> y1 = dPL(x1, mu, sigma, lambda,
+         family = "Hyp", zeta = 2)
R> polygon(c(x1, 0.4, 0), c(y1, 0, 0),
+         col = "lightblue")

R> text(mu-0.025, 1,
+       paste("P(Y<0.4) = ",
+             round(pPL(0.4, mu, sigma,
+                       lambda,
+                       family = "Hyp",
+                       zeta = 2),
+             2)), font = 1,
+       cex = 0.6)
```

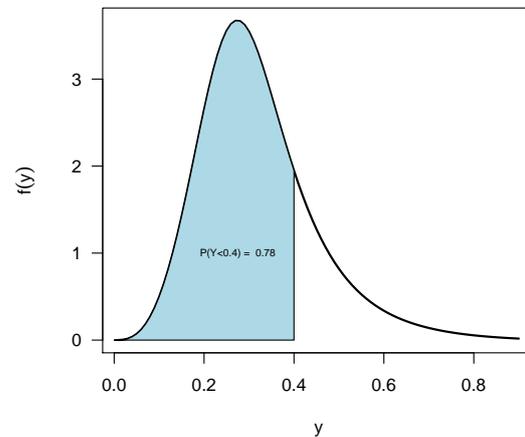


Figura 4.2: Função densidade de probabilidade de uma distribuição PL-Hyp₍₂₎.

Função de distribuição acumulada

```
R> mu = 0.3; sigma = 1; lambda = 2
R> x = seq(0.01, 0.9, 0.01)
R> plot(x, pPL(x, mu, sigma, lambda,
+         family = "PE", zeta = 1.3),
+       type = "l", las = 1, lwd = 2,
+       ylab = expression(P(Y<y)),
+       xlab = "y")
R> p = pPL(0.5, mu, sigma, lambda,
+         family = "PE", zeta = 1.3)
R> q = qPL(p, mu, sigma, lambda,
+         family = "PE", zeta = 1.3)
R> points(q, p, pch = 16, col = 2,
+        cex = 1.5)
R> text(0.55, 0.83,
+       paste("(", 0.5, ")",
+             round(p, 2), ")"),
+       font = 2, cex = 0.8, col = "red")
```

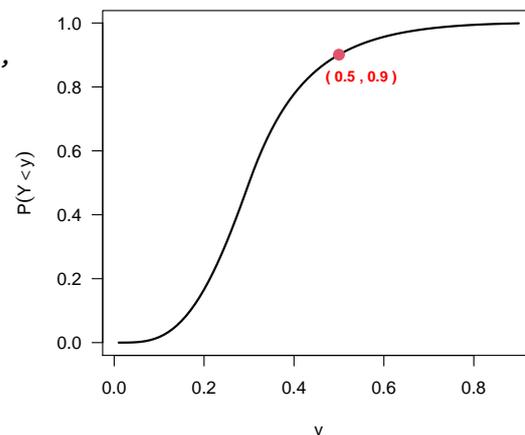


Figura 4.3: Função de distribuição acumulada de uma distribuição PL-PE_(1.3).

4.3 Implementação do modelo de regressão *power logit* no pacote PLreg

Conforme mencionado anteriormente, o pacote PLreg permite ajustar um modelo de regressão *power logit*, cujo gerador de densidades pode ser normal, t-Student, exponencial potência, logística tipo II, slash, hiperbólica e sinh-normal. O cálculo de medidas de diagnóstico associadas ao modelo

ajustado, como resíduos, medidas de influência local e medidas de qualidade de ajuste, também está implementado no pacote. O processo de estimação é baseado no método de máxima verossimilhança e atualmente dois tipos de estimadores são suportados: o estimador de máxima verossimilhança usual e o estimador de máxima verossimilhança penalizado. Além disso, o parâmetro de assimetria λ pode ser fixado, portanto, este pacote também permite o ajuste do modelo de regressão GJS. Além disso, o pacote `PLreg` também permite o ajuste dos modelos de regressão log-log, o modelo limite discutido na Seção 3.5.

A principal função do pacote `PLreg` é `PLreg()`, que segue a abordagem padrão para implementar modelos de regressão em R. Uma vez que o processo de ajuste foi realizado, um objeto da classe `S3` “`PLreg`” é produzido, para o qual uma série de métodos são disponibilizados. Os argumentos da função `PLreg()` são semelhantes aos das funções de outros pacotes para modelos de regressão em R, como a função `betareg()`:

```
PLreg(formula, data, subset, na.action,
      family = c("NO", "LO", "TF", "PE", "SN", "SLASH", "Hyp"),
      zeta = NULL,
      link = c("logit", "probit", "cloglog", "cauchit", "log", "loglog"),
      link.sigma = NULL, type = c("pML", "ML"), control = PLreg.control(...),
      model = TRUE, y = TRUE, x = FALSE, ...)
```

O argumento `formula` pode conter três partes (separadas pelos símbolos “ \sim ” e “ $|$ ”), a saber: variável resposta observada no intervalo unitário, preditor linear do submodelo da mediana e da dispersão; (veja [Zeileis e Croissant \(2020\)](#) para mais detalhes sobre o argumento `formula`). Por exemplo, se `formula = y ~ x1 + x2 | s1 + s2 + s3` é fornecido, ele descreve y para a variável resposta, x_1 e x_2 para o submodelo da mediana e s_1 , s_2 e s_3 para o submodelo da dispersão do modelo de regressão *power logit*. A ausência da terceira parte do argumento `formula` indica para `PLreg()` que o modelo deve ser ajustado assumindo dispersão constante. Assim, um modelo de regressão *power logit* com dispersão constante pode ser especificado por `formula = y ~ x1 + x2`. A função fornece seis opções para a função de ligação para o submodelo da mediana, ou seja, para a função $d_1(\cdot)$: “`logit`”, “`probit`”, “`cloglog`”, “`cauchit`”, “`log`”, “`loglog`”; para o submodelo da dispersão, ou seja, para a função $d_2(\cdot)$, duas funções de link são permitidas: “`log`”, “`sqrt`”. As funções de ligação *default* são “`logit`” (para o submodelo da mediana) e “`log`” (para o submodelo da dispersão). Existem dois outros argumentos importantes: `family` e `zeta`. O argumento `family` especifica a distribuição simétrica usada para gerar o modelo *power logit*. Para algumas distribuições, um parâmetro extra deve ser especificado no argumento `zeta`.

A função `PLreg.fit()` fornece uma abordagem alternativa para a definição do modelo:

```
PLreg.fit(X, y, S = NULL, family, type = "pML", zeta = zeta,
          link = "logit", link.sigma = "log",
          control = PLreg.control(...))
```

Note que essa função não precisa da especificação de uma fórmula, basta que o usuário forneça as matrizes de covariáveis do modelo e o vetor de variável resposta. O processo de estimação é realizado através da função `optim` com opções de controle definidas pela função `PLreg.control()`. Conforme mencionado, a estimação pode ser feita através do estimador de máxima verossimilhança usual (“`ML`”) e do estimador de máxima verossimilhança penalizado (“`pML`”); isso deve ser especificado no argumento `type`, por *default*, `type = "pML"`. Se o parâmetro de assimetria (λ) for fixado, apenas o estimador de máxima verossimilhança usual é suportado. Além disso, para λ fixado, deve-se especificar o valor no argumento `control` através da função `PLreg.control()`. Considerando λ fixo, dois modelos especiais podem ser ajustados: o modelo de regressão GJS, fazendo `control = PLreg.control(lambda = 1)` e o modelo de regressão log-log, fazendo `control = PLreg.control(lambda = 0)` (vale enfatizar que esta é apenas uma notação, na verdade, $\lambda = 0$ significa $\lambda \rightarrow 0^+$). Adicionalmente, se `type = "ML"`, a função `optim` usa gradientes analíticos no

processo iterativo; se `type = "pML"`, gradientes analíticos são usados apenas no processo iterativo para estimar os parâmetros relacionados aos submodelos da mediana e da dispersão. Por *default*, os valores iniciais dos coeficientes de regressão são obtidos conforme descrito na Seção 3.2. Por outro lado, valores iniciais para β e τ podem ser passados através da função `PLreg.control()`.

A função `PLreg()` retorna um objeto da classe “`PLreg`”, cujos componentes são apresentados na Tabela 4.1. O objeto da classe “`PLreg`” tem alguns métodos disponíveis. O método `summary()` apresenta uma saída padrão, com estimativas de coeficientes, erros padrão, testes Wald parciais e valores p para os coeficientes de regressão; a medida global de qualidade de ajuste (Υ_ζ); o pseudo R^2 ; etc. O argumento `type` na função `summary()` especifica o tipo de resíduos incluídos na saída, atualmente três resíduos são suportados: “`padronizado`”, “`quantílico`” e “`deviance`”. O método `plot()` desenha gráficos úteis para a análise de diagnóstico e análises de influência. A Tabela 4.2 apresenta os principais métodos e funções disponíveis no pacote `PLreg`.

Uma função importante do pacote `PLreg` é `extra.parameter()`. Esta função pode ser usada para estimar o parâmetro extra (ou vetor de parâmetros extras) de alguns modelos da classe *power logit*. O uso básico desta função é:

```
extra.parameter(object, lower, upper, grid = 10)
```

Esta função fornece gráficos da medida Υ_ζ e de $-2\ell(\hat{\theta})$ como função de ζ . O argumento `object` é um objeto da classe `S3` obtido de um ajuste da função `PLreg()`; `lower` e `upper` são os limites inferior e superior para o parâmetro extra, respectivamente; e `grid` é o número de valores de ζ que serão avaliados no intervalo. Além disso, essa função retorna uma lista com cinco elementos: “`zeta.Ups`”, “`zeta.loglik`”, “`zeta.values`”, “`Upsilon.values`” e “`loglik.values`”, que representam, respectivamente, as estimativas de ζ pelos dois métodos, os valores de ζ utilizados e os valores obtidos de Υ_ζ e de $-2\ell(\hat{\theta})$ para cada valor de ζ . Esses elementos são úteis para a elaboração de gráficos com configurações diferentes das disponibilizadas pela função `extra.parameter()`.

Alguns conjuntos de dados estão disponíveis no pacote `PLreg` e são descritos a seguir.

1. `bodyfat_Aeolus`. Este conjunto de dados foi coletado por Cheng *et al.* (2019) e a variável de resposta é `percentualfat`, a proporção de gordura corporal de morcegos. Nestes dados, consideramos morcegos amostrados em *Aeolus Cave*, localizada em East Dorset, Vermont, no nordeste dos Estados Unidos, mas no trabalho original existem cinco outras regiões. O conjunto de dados tem 159 observações e três covariáveis: `sex` (sexo do morcego), `year` (ano que o morcego foi amostrado) e `days` (tempo de hibernação, definido como dias desde o equinócio de outono).
2. `Firm`. Este conjunto de dados foi apresentado e analisado por Schmit e Roth (1990) e está disponível na página pessoal do Professor E. Frees (Wisconsin School of Business Research). A variável resposta é `firmcost`, valores pequenos desta variável são atribuídos às empresas que apresentam um bom desempenho de gestão de risco. O conjunto de dados tem 73 observações e seis covariáveis: `assume` (montante de retenção de ocorrência como uma porcentagem dos ativos totais), `cap` (indica que a empresa possui uma seguradora cativa; 1 se sim, 0 caso contrário), `sizelog` (logaritmo do ativo total), `indcost` (uma medida do risco setorial da empresa), `central` (uma medida da importância dos gerentes locais na escolha da quantidade de risco a ser retida), `soph` (uma medida do grau de importância no uso de ferramentas analíticas).
3. `PeruVotes`. Este conjunto de dados foi coletado por Bayes *et al.* (2012) e a variável de resposta é `votes`, proporção de votos em branco nas eleições gerais peruanas de 2006. O conjunto de dados contém 194 observações e a covariável é `HDI`, índice de desenvolvimento humano.

No que segue, apresentamos algumas aplicações a fim de exemplificar o uso do pacote `PLreg` na prática.

Tabela 4.1: Componentes retornados pela função *PLreg()*.

Componente	Descrição
<code>coefficients</code>	uma lista com os coeficientes da mediana, dispersão e assimetria (se <code>lambda = NULL</code>).
<code>residuals</code>	um vetor com os resíduos <i>raw</i> (a diferença entre os valores observados e esperados da variável resposta).
<code>fitted.values</code>	um vetor com os valores ajustados (mediana ajustada para cada observação).
<code>optim</code>	uma lista com o <i>output</i> da função <code>optim</code> . Para λ não fixado, se <code>type = "pML"</code> , o <i>output</i> é baseado no processo iterativo associado aos parâmetros do submodelo da mediana e da dispersão; e, se <code>type = "ML"</code> , é baseado no processo iterativo associado a estimação de todos os parâmetros do modelo.
<code>family</code>	um caracter especificando o gerador de densidade utilizado.
<code>method</code>	o método utilizado pela função <code>optim</code> (<i>default</i> é "BFGS").
<code>control</code>	os argumentos controles passados para a função <code>optim</code> .
<code>start</code>	um vetor contendo os pontos iniciais utilizados no processo iterativo.
<code>nobs</code>	número de observações.
<code>df.null</code>	graus de liberdade do resíduo do modelo nulo (mediana e dispersão constantes), isto é, $n - 3$.
<code>df.residual</code>	graus de liberdade do resíduo do modelo ajustado.
<code>lambda</code>	valor numérico indicando o valor do parâmetro de assimetria λ (NULL para λ não fixado).
<code>loglik</code>	função de log-verossimilhança do modelo ajustado (função de log-verossimilhança usual).
<code>vcov</code>	matriz de variância e covariância de todos os parâmetros do modelo.
<code>pseudo.r.squared</code>	valor do pseudo R^2 (definido como o quadrado da correlação linear entre o preditor linear a variável resposta transformada através da função $d_1(\cdot)$).
<code>Upsilon.zeta</code>	medida de bondade de ajuste global (ver Seção 2.3.2).
<code>link</code>	uma lista contendo os elementos "median" e "dispersion" com os respectivos objetos associados as funções de ligação do modelo ajustado.
<code>converged</code>	valor lógico indicando se houve sucesso na convergência do processo iterativo.
<code>zeta</code>	valor numérico especificando o valor de ζ .
<code>type</code>	um caracter especificando o tipo de estimador utilizado.
<code>v</code>	um vetor com os valores da função $v(z)$ para todas as observações.
<code>call</code>	chamada da função original.
<code>formula</code>	fórmula utilizada.
<code>terms</code>	uma lista com elementos "median", "dispersion" e "full" contendo objetos com os termos dos respectivos submodelos.
<code>levels</code>	uma lista com elementos "median", "dispersion" e "full" contendo os níveis das covariáveis categóricas.
<code>contrasts</code>	uma lista com elementos "median" e "dispersion" contendo os constrantes referentes aos níveis dos respectivos modelos.
<code>model</code>	o <i>model frame</i> completo (se <code>y == TRUE</code>).
<code>y</code>	variável resposta (se <code>y == TRUE</code>).
<code>x</code>	uma lista com elementos "median" e "dispersion" contendo as matrizes dos submodelos da mediana e da dispersão (se <code>x == TRUE</code>).

Tabela 4.2: Funções e métodos para objetos da classe “PLreg”.

Função	Descrição
<code>print()</code>	apresenta as estimativas dos coeficientes do modelo.
<code>summary()</code>	<i>output</i> para o ajuste do modelo (estimativas dos coeficientes, erros padrão, Wald testes).
<code>coef()</code>	extrai todos os coeficientes do modelo em um único vetor.
<code>vcov()</code>	matriz de variância e covariância.
<code>logLik()</code>	extrai a função de log-verossimilhança ajustada
<code>model.matrix()</code>	extrai a matriz do modelo de todos os componentes
<code>AIC()</code>	calcula os critérios de informação (AIC, BIC, ...).
<code>residuals()</code>	extrai os resíduos associados ao modelo ajustado (quantílico, padronizado e deviance). O <i>default</i> é o resíduo padronizado.
<code>plot()</code>	apresenta gráficos úteis para a análise de diagnóstico. Atualmente, sete tipos de gráficos estão disponíveis: resíduos <i>versus</i> índice das observações, gráfico de índices de $ h_{max} $ sob o esquema de ponderação de casos, diagrama de dispersão dos elementos diagonais de elementos da diagonal de $GL(\hat{\theta})$ <i>versus</i> valores preditos, resíduos <i>versus</i> preditor linear, gráfico de probabilidade normal dos resíduos, valores preditos <i>versus</i> valores observados, resíduos <i>versus</i> função $v(z)$.
<code>influence()</code>	fornece medidas de influência associadas ao modelo ajustado. Atualmente, a função retorna uma lista com três objetos: “ <code>case.weights</code> ”, “ <code>totalLI</code> ” e “ <code>GL</code> ”, contendo, respectivamente, os valores de $ h_{max} $ sob o esquema de ponderação de casos, os valores de C_i (curvatura na direção do i -ésimo caso e os elementos da diagonal de $GL(\hat{\theta})$.
<code>envelope()</code>	fornece o gráfico de probabilidade normal com envelopes simulados.
<code>extra.parameter()</code>	ferramenta para estimar o parâmetro extra (se existir).

4.4 Aplicações utilizando o pacote PLreg

4.4.1 Dados `bodyfat_Aeolus`: Ajustando a distribuição *power logit*

Nesta aplicação, consideramos os dados `bodyfat_Aeolus`. Ajustamos distribuições da classe *power logit* à variável porcentagem de gordura corporal (y).

```
R> data("bodyfat_Aeolus", package = "PLreg")
R> head(bodyfat_Aeolus, n = 5)
```

```
  sex percentfat year days
1  F   0.1487539 2016   48
2  F   0.3203463 2016   48
3  F   0.3233591 2016   48
4  F   0.2032902 2016   48
5  M   0.2014726 2016   48
```

No que segue, ajustamos as distribuições PL-N, PL-PE $_{(\zeta)}$ e PL-SN $_{(\zeta)}$ à variável y . Recomendamos utilizar a função `extra.parameter()` para se obter o melhor valor de ζ para os modelos PL-PE $_{(\zeta)}$ e PL-SN $_{(\zeta)}$.

```
R> # Ajustando a distribuição PL-N
R> fitPL_NO <- PLreg(percentfat ~ 1, data = bodyfat_Aeolus, family = "NO")
```

```
R> # Ajustando a distribuição PL-PE
R> fitPL_PE.aux <- PLreg(percentfat ~ 1, data = bodyfat_Aeolus,
```

```

+                               family = "PE", zeta = 3)
R> # extra.parameter(fitPL_PEAux, lower = 1, upper = 6) #escolhendo zeta = 3.22
R> fitPL_PE <- PLreg(percentfat ~ 1, data = bodyfat_Aeolus,
+                   family = "PE", zeta = 3.22)

R> # Ajustando a distribuição PL-SN
R> fitPL_SN.aux <- PLreg(percentfat ~ 1, data = bodyfat_Aeolus,
+                       family = "SN", zeta = 1)
R> # extra.parameter(fitPL_SN.aux, lower = 1, upper = 4) #escolhendo zeta = 1.67
R> fitPL_SN <- PLreg(percentfat ~ 1, data = bodyfat_Aeolus,
+                   family = "SN", zeta = 1.67)

```

A Figura 4.4 apresenta os gráficos fornecidos pela função `extra.parameter()` para o ajuste `fitPL_PE.aux`.

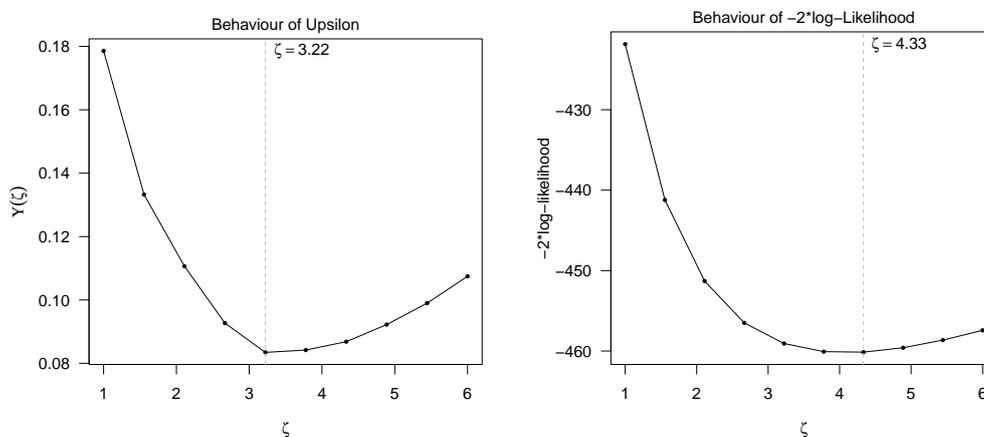


Figura 4.4: Gráficos fornecidos pela função `extra.parameter()` para o ajuste `fitPL_PE.aux`.

Uma forma de escolher o melhor modelo é através da medida Υ_ζ ou do AIC e BIC, por exemplo. Essas medidas, para cada um dos modelos ajustados, são apresentadas a seguir.

```

R> PL_NO <- round(c(fitPL_NO$Upsilon.zeta, AIC(fitPL_NO), BIC(fitPL_NO)), 3)
R> PL_PE <- round(c(fitPL_PE$Upsilon.zeta, AIC(fitPL_PE), BIC(fitPL_PE)), 3)
R> PL_SN <- round(c(fitPL_SN$Upsilon.zeta, AIC(fitPL_SN), BIC(fitPL_SN)), 3)

```

```

R> medidas <- rbind(PL_NO, PL_PE, PL_SN)
R> colnames(medidas) <- c("Upsilon", "AIC", "BIC")
R> medidas

```

	Upsilon	AIC	BIC
PL_NO	0.114	-443.771	-434.564
PL_PE	0.083	-453.058	-443.851
PL_SN	0.082	-453.942	-444.736

Com base nesses critérios, notamos que os modelos $PL-PE_{(3.44)}$ e $PL-SN_{(1.67)}$ apresentaram um melhor ajuste. Consideramos, portanto, o modelo *power logit* sinh-normal.

```

R> summary(fitPL_SN)

```

Call:

```
PLreg(formula = percentfat ~ 1, data = bodyfat_Aeolus, family = "SN", zeta = 1.67)
```

Standardized residuals:

Min	1Q	Median	3Q	Max
-2.4654	-0.8318	-0.2065	0.7453	2.0551

Coefficients (median model with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.11477	0.05416	-39.05	<2e-16 ***

Sigma coefficients (dispersion model with log link):

	Estimate	Std. Error	z value	Pr(> z)
(sigma)	0.2107	0.4506	0.468	0.64

Lambda coefficient:

	Estimate	Std. Error
(lambda)	1.438	0.764

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Family: PL - SN (1.67) (Power logit sinh-normal)

Estimation method: pML (penalized maximum likelihood)

Log-likelihood: 230 on 3 Df

Upsilon statistic: 0.08151

AIC: -453.9

Number of iterations in BFGS optimization: 9

A porcentagem de gordura corporal mediana estimada, com base nesse ajuste, é, aproximadamente, 11%. A função densidade de probabilidade estimada por esse modelo é apresentada na Figura 4.5.

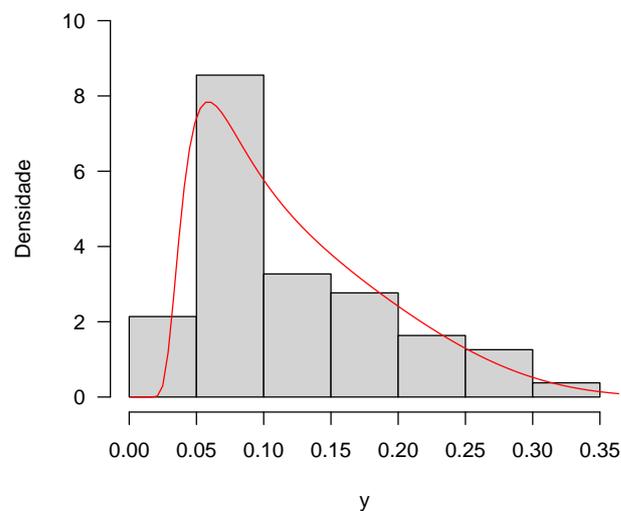


Figura 4.5: Função densidade de probabilidade estimada com base no ajuste fitPL_{SN} .

4.4.2 Dados Firm: Ajustando o modelo de regressão *power logit* com dispersão constante

Neste exemplo, utilizamos os dados Firm para ajustar um modelo de regressão *power logit* slash com dispersão constante. A variável resposta é `firmcost` e as covariáveis são `sizelog` e `indcost`. A fim de escolher o melhor valor para ζ , consideramos um ajuste inicial com $\zeta = 2$ (slash canônica) e, depois, utilizamos o objeto criado para escolher o valor de ζ através da função `extra.parameter()`.

```
R> fitPL_slash.aux <- PLreg(FIRMCOST ~ SIZELOG + INDCOST, data = Firm,
+                           family = "SLASH", zeta = 2)
R> extra.parameter(fitPL_slash.aux, lower = 1, upper = 3)
```

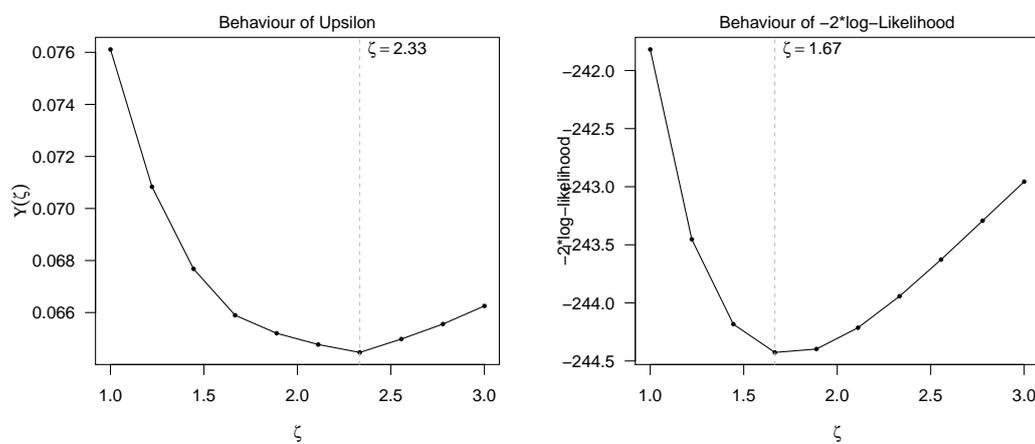


Figura 4.6: Gráficos fornecidos pela função `extra.parameter()` para o ajuste `fitPL_slash.aux`.

Com base na Figura 4.6, escolhemos $\zeta = 2.33$.

```
R> fitPL_slash <- PLreg(FIRMCOST ~ SIZELOG + INDCOST, data = Firm,
+                       family = "SLASH", zeta = 2.33)
R> summary(fitPL_slash)
```

Call:

```
PLreg(formula = firmcost ~ sizelog + indcost, data = Firm, family = "SLASH", zeta = 2.33)
```

Standardized residuals:

	Min	1Q	Median	3Q	Max
	-2.1196	-0.6613	0.0548	0.7173	5.9008

Coefficients (median model with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.8625	0.9994	3.865	0.000111 ***
sizelog	-0.9049	0.1120	-8.077	6.66e-16 ***
indcost	2.1345	0.5835	3.658	0.000254 ***

Sigma coefficients (dispersion model with log link):

	Estimate	Std. Error	z value	Pr(> z)
(sigma)	0.1481	0.5344	0.277	0.782

Lambda coefficient:

```

          Estimate Std. Error
(lambda)   1.809      1.022
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Family: PL - SLASH ( 2.33 ) (Power logit slash)
Estimation method: pML (penalized maximum likelihood)
Log-likelihood:    122 on 5 Df
Pseudo R-squared: 0.4162
Upsilon statistic: 0.06446
AIC: -233.9
Number of iterations in BFGS optimization: 15

```

A função `summary()` retorna as principais informações a respeito do ajuste. Por *default*, essa função apresenta as principais estatísticas descritivas do resíduo padronizado para o ajuste feito. O tipo de resíduo utilizado pode ser mudado através do argumento `type`. As estimativas dos parâmetros do modelo são apresentadas de forma análoga às principais funções para ajuste de modelos de regressão no R. Essa função também apresenta a distribuição PL utilizada (`family`), o valor da log-verossimilhança, da medida Υ_ζ e do AIC. O método de estimação utilizado também é informado. Para ajustar o mesmo modelo considerando o estimador de máxima verossimilhança usual, basta seguir o comando:

```
R> PLreg(FIRMCOST ~ SIZELOG + INDCOST, data = Firm, family = "SLASH",
+       zeta = 2.33, type = "ML")
```

A função `envelope()` fornece o gráfico de probabilidade normal com envelopes simulados para o modelo ajustado. O argumento `type` especifica o tipo de resíduo utilizado. Através da Figura 4.7, notamos que a observação 15 aparece destacada no gráfico de envelope. Essa observação, no entanto, é a que recebe menor peso no processo de estimação dos parâmetros do submodelo da mediana; veja o gráfico de $v(z)$ versus o resíduo padronizado na Figura 4.7.

```
R> set.seed(234)
R> envelope(fitPL_slash, type = "standardized")

R> plot(fitPL_slash, which = 7)
R> identify(residuals(fitPL_slash, type = "standardized"), fitPL_slash$v)
```

4.4.3 Dados `bodyfat_Aeolus`: Ajustando o modelo limite

A seguir, consideramos o conjunto de dados `bodyfat_Aeolus` no contexto de regressão. A variável resposta é `percentfat` e as covariáveis são `sex`, `days` e `year`. Consideramos o ajuste do modelo de regressão PL-PE com dispersão constante. O valor escolhido para ζ , através da função `extra.parameter()`, foi $\zeta = 1.67$.

```
R> fitPL_PE2 <- PLreg(percentfat ~ days + sex + year,
+                   data = bodyfat_Aeolus, family = "PE",
+                   zeta = 1.67)
R> fitPL_PE2
```

Call:

```
PLreg(formula = percent.fat ~ days + sex + year, data = bodyfat_Aeolus,
      family = "PE", zeta = 1.67)
```

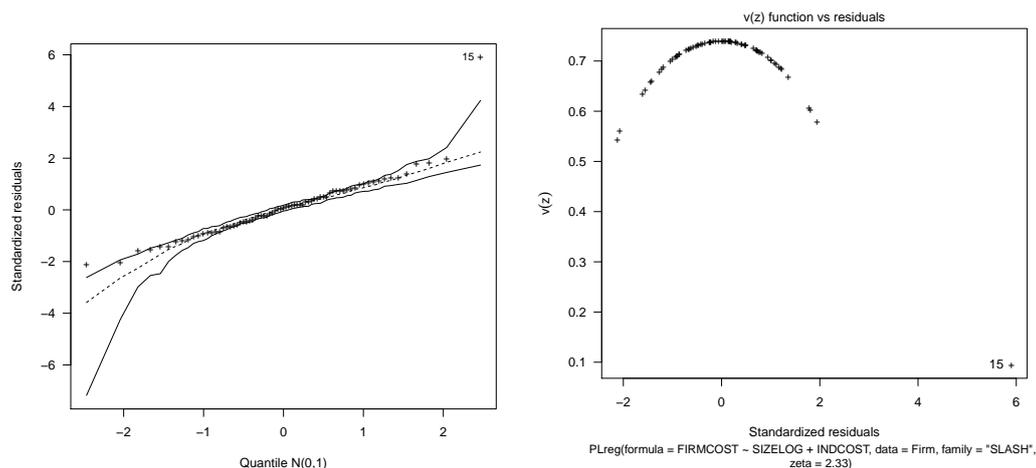


Figura 4.7: Gráfico de probabilidade normal com envelopes simulados (lado esquerdo) e gráfico de $v(z)$ versus o resíduo padronizado para o ajuste `fitPL_slash`.

Coefficients (median model with logit link):

(Intercept)	days	sexM	year2016
-1.161295	-0.009258	-0.051481	0.518582

Sigma coefficients (dispersion model with log link):

(sigma)
-1.943

Lambda coefficients:

(lambda)
0.000448

Conforme observamos, a estimativa do parâmetro de assimetria λ é bem próxima de zero, indicando que o modelo limite pode ser mais adequado. O ajuste do modelo limite pode ser feito através do seguinte comando:

```
R> fitloglog_PE <- PLreg(percentfat ~ days + sex + year,
+                          data = bodyfat_Aeolus, family = "PE",
+                          zeta = 1.67, control = PLreg.control(lambda = 0))
R> summary(fitloglog_PE)
```

Call:

```
PLreg(formula = percent.fat ~ days + sex + year, data = bodyfat_Aeolus,
      family = "PE", zeta = 1.67, control = PLreg.control(lambda = 0))
```

Standardized residuals:

Min	1Q	Median	3Q	Max
-2.9997	-0.6021	0.0208	0.6340	2.5296

Coefficients (median model with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.1612704	0.0643371	-18.050	<2e-16 ***
days	-0.0092578	0.0005356	-17.284	<2e-16 ***
sexM	-0.0514785	0.0572194	-0.900	0.368
year2016	0.5185855	0.0608399	8.524	<2e-16 ***

```

Sigma coefficients (dispersion model with log link):
      Estimate Std. Error z value Pr(>|z|)
(sigma) -1.94396    0.06137  -31.68  <2e-16 ***

Fixed skewness parameter (limiting case lambda -> 0).
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Family: log-log - PE ( 1.67 ) (log-log power exponential)
Estimation method: ML (maximum likelihood)
Log-likelihood: 319.9 on 5 Df
Pseudo R-squared: 0.6756
Upsilon statistic: 0.0482
AIC: -629.9
Number of iterations in BFGS optimization: 19

```

Observe que, para este modelo, a função `PLreg` utiliza o método de máxima verossimilhança usual para a estimação dos parâmetros do modelo. Esse método também é utilizado quando o parâmetro de assimetria é fixado. Nestes casos, o algoritmo de otimização BFGS utiliza derivadas analíticas no processo iterativo. Finalmente, as covariáveis `days` e `year` são significantes para explicar a mediana da variável resposta (ao nível de significância de 5%).

Gráficos de diagnóstico podem ser obtidos através da função `plot()`. O argumento `which` assume valores de um a sete (atualmente, sete gráficos de diagnóstico são disponíveis). Por *default*, o tipo de resíduo retornado por essa função é o padronizado (3.11), mas isso pode ser mudado com o argumento `type` dessa função. Os resíduos utilizados nesses gráficos podem ser obtidos através da função `residuals()`. Por exemplo, para obter os resíduos quantílicos associados ao ajuste `fitloglog_PE`, pode-se utilizar o comando: `residuals(fitloglog_PE, type = "quantile")`; veja a Figura 4.8.

```

R> par(mfrow = c(3, 2))
R> plot(fitloglog_PE, which = 1:6)

```

Medidas de influência podem ser obtidas através da função `influence()`. Esta função retorna duas medidas de influência (influência local sob o esquema de ponderação de casos e a medida de influência total definida em (3.13)) e a medida de alavancagem generalizada (elementos diagonais de $GL(\hat{\theta})$). Se o argumento `graph = TRUE`, esta função também fornece gráficos de índices dessas medidas; veja a Figura 4.9.

```

R> influence(fitloglog_PE, graph = TRUE)

```

4.4.4 Dados PeruVotes: Ajustando modelos com dispersão variável e λ fixo

Nesta aplicação, consideramos os dados `PeruVotes` para exemplificar o uso do pacote `PLreg` nos ajustes de modelos com dispersão variável e λ fixo. A variável resposta é `votes`, a proporção de votos brancos nas eleições gerais do Peru em 2006, e a covariável é `HDI`, índice de desenvolvimento humano. Vamos considerar como gerador de densidades a distribuição t-Student e fixemos $\zeta = 5$. Considere os seguintes ajustes:

```

R> fitPL1 <- PLreg(votes ~ HDI | HDI, data = PeruVotes,
+               family = "TF", zeta = 5,
+               control = PLreg.control(lambda = 1))

```

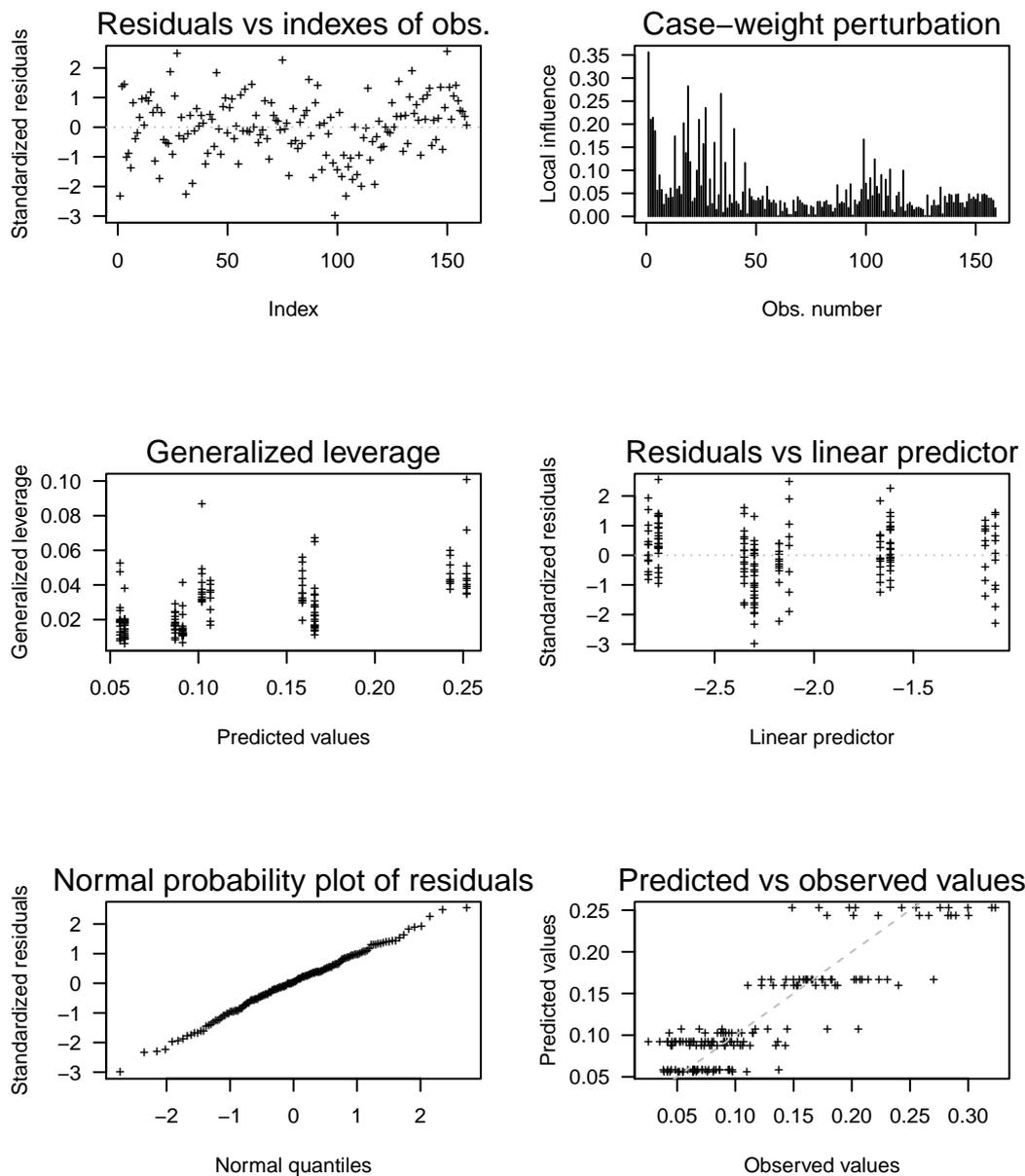


Figura 4.8: Gráficos fornecidos pela função `plot()` para o ajuste `fitloglog_PE`.

```
R> fitPL05 <- PLreg(votes ~ HDI | HDI, data = PeruVotes,
+                 family = "TF", zeta = 5,
+                 control = PLreg.control(lambda = 0.5))

R> fitPL0 <- PLreg(votes ~ HDI | HDI, data = PeruVotes,
+                 family = "TF", zeta = 5,
+                 control = PLreg.control(lambda = 0))
```

Para λ fixado, o valor deve ser especificado no argumento `control`, através da função `PLreg.control()`. Os ajustes `fitPL1` e `fitPL0` são, respectivamente, os ajustes dos modelos de regressão GJS t-Student e log-log t-Student. Abaixo, apresentamos o resumo do ajuste `fitPL1`.

```
R> summary(fitPL1)
```

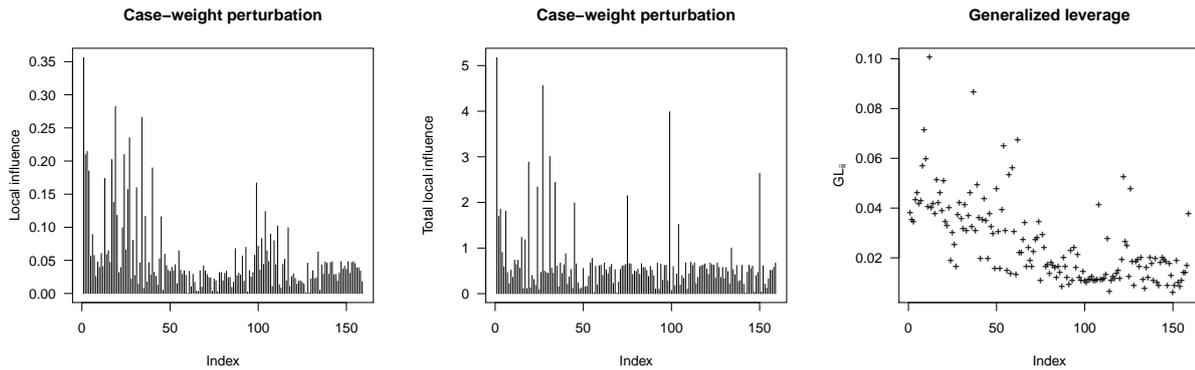


Figura 4.9: Gráficos fornecidos pela função *influence()* para o ajuste *fitloglog_PE*.

Call:

```
PLreg(formula = votes ~ HDI | HDI, data = PeruVotes, family = "TF",
      zeta = 5, control = PLreg.control(lambda = 1))
```

Standardized residuals:

Min	1Q	Median	3Q	Max
-5.4228	-0.6212	-0.0012	0.6125	4.4180

Coefficients (median model with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.3004	0.2089	11.01	<2e-16 ***
HDI	-6.7965	0.3784	-17.96	<2e-16 ***

Sigma coefficients (dispersion model with log link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.7069	0.6787	-3.988	6.65e-05 ***
HDI	2.1401	1.2035	1.778	0.0754 .

Fixed skewness parameter (lambda = 1).

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Family: PL - TF (5) (Power logit Student-t)

Estimation method: ML (maximum likelihood)

Log-likelihood: 353.2 on 4 Df

Pseudo R-squared: 0.5822

Upsilon statistic: 0.06286

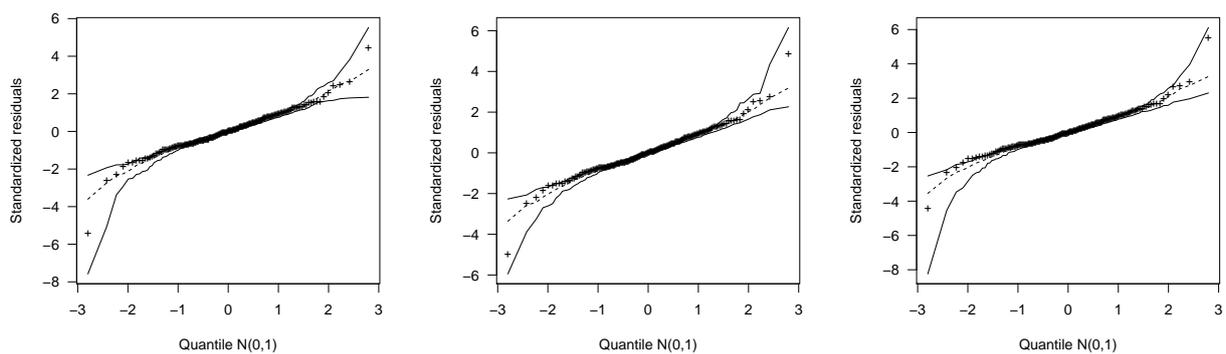
AIC: -698.5

Number of iterations in BFGS optimization: 9

Os resultados dos ajustes são apresentados na Tabela 4.3. Note que as estimativas dos parâmetros associados ao submodelo da mediana não mudam de forma substancial quando comparamos os quatro ajustes. No entanto, observamos mudança nos parâmetros associados ao submodelo da dispersão. Os gráficos de probabilidade normal com envelopes simulados (Figura 4.10) indicam que todos os quatro modelos apresentam ajustes satisfatórios.

Tabela 4.3: Estimativas e erros-padrão dos ajustes *fitPL1*, *fitPL05* e *fitPLO*, respectivamente.

	$\lambda = 1$		$\lambda = 0.5$		$\lambda = 0$	
	Est.	e.p.	Est.	e.p.	Est.	e.p.
<i>Submodelo da mediana</i>						
intercepto	2.300	0.209	2.298	0.209	2.294	0.209
HDI	-6.796	0.378	-6.790	0.378	-6.780	0.378
<i>Submodelo da dispersão</i>						
intercepto	-2.707	0.679	-2.591	0.680	-2.321	0.683
HDI	2.140	1.203	1.335	1.206	0.146	1.212

**Figura 4.10:** Gráficos de probabilidade normal com envelopes simulados fornecidos pela função *envelope()* para os ajustes *fitPL1*, *fitPL05* e *fitPLO*, respectivamente.

Capítulo 5

O modelo de regressão *power logit* inflacionado em zero ou um

Existem diferentes situações práticas em que o interesse é modelar uma variável aleatória que assume valores no intervalo unitário, mas que pode apresentar zeros ou uns (mas não ambos). Por exemplo, a proporção da renda familiar gasta em telefonia celular. Há famílias que não possuem celular e o gasto com esse item é nulo. Por outro lado, dificilmente alguma família gasta a renda inteira neste item. Assim, neste capítulo, estendemos o modelo de regressão *power logit* para o caso em que a variável resposta pode assumir os valores 0 ou 1. A ideia principal é assumir que a distribuição dos dados é uma mistura entre a distribuição *power logit* e uma distribuição degenerada em 0 (zero) ou 1 (um), dependendo do caso. Este modelo será denominado *modelo de regressão power logit inflacionado em zero ou um*. O capítulo encontra-se organizado como segue. Na Seção 5.1 definimos a classe de distribuições *power logit* inflacionada em zero ou um. O modelo de regressão *power logit* inflacionado é estudado na Seção 5.2 e medidas de diagnóstico e influência são propostas na Seção 5.3. Finalmente, na Seção 5.4 apresentamos uma função que pode ser útil para ajustar o modelo proposto e uma aplicação em dados reais é conduzida na Seção 5.5.

5.1 A classe *power logit* inflacionada em zero ou um

Para definirmos a classe de distribuições *power logit* inflacionada em zero ou um, vamos assumir que a distribuição dos dados é uma mistura, em que o componente contínuo dos dados é modelado por uma distribuição *power logit* (Definição 2.2) e o componente discreto será modelado por uma distribuição degenerada em c , com $c = 0$ ou $c = 1$, dependendo do caso. Dessa forma, dizemos que Y tem distribuição *power logit* inflacionada em c se sua função de distribuição acumulada é dada por

$$\text{IPL}^{(c)}(y; \alpha, \mu, \sigma, \lambda) = \alpha \mathbb{I}_{[c, \infty)}(y) + (1 - \alpha) F_Y(y; \mu, \sigma, \lambda), \quad y \in \mathbb{R}, \quad (5.1)$$

em que $\mathbb{I}_A(y)$ é a função indicadora, com valor 1 se $y \in A$ e 0 se $y \notin A$ e a função $F(y; \mu, \sigma, \lambda)$ é a função de distribuição acumulada da distribuição $\text{PL}(\mu, \sigma, \lambda; r)$ cuja função densidade de probabilidade é dada em (2.4). Escrevemos $Y \sim \text{IPL}_c(\alpha, \mu, \sigma, \lambda; r)$. Note que a função $\text{IPL}^{(c)}(\cdot; \alpha, \mu, \sigma, \lambda)$ não é absolutamente contínua, pois tem um ponto de massa em $y = c$. Também, com probabilidade α , a variável Y é selecionada de uma distribuição degenerada no ponto c ($c = 0$ ou $c = 1$) e, com probabilidade $1 - \alpha$, a variável Y é selecionada de uma distribuição *power logit*, que é absolutamente contínua com respeito à medida de Lebesgue¹. Assim, a variável aleatória Y é do tipo mista.

¹A medida de probabilidade P correspondente a $\text{IPL}^{(c)}(y; \cdot)$ definida sob o espaço mensurável $((0, 1) \cup \{c\}, \mathcal{B})$, em que \mathcal{B} a classe de todos os subconjuntos de Borel de $(0, 1) \cup \{c\}$, é tal que $P \ll \zeta + \delta_c$, sendo ζ a medida de Lebesgue e δ_c o ponto de massa em c ; isto é, $\delta_c(A) = 1$, se $c \in A$ e $\delta_c(A) = 0$, se $c \notin A$, $A \in \mathcal{B}$.

A função densidade de probabilidade de Y é da forma

$$f_Y^{(c)}(y; \alpha, \mu, \sigma, \lambda) = \begin{cases} \alpha, & y = c, \\ (1 - \alpha)f_Y(y; \mu, \sigma, \lambda), & y \in (0, 1), \end{cases} \quad (5.2)$$

em que $\alpha \in (0, 1)$, $\mu \in (0, 1)$, $\sigma > 0$, $\lambda > 0$ e $f_Y(y; \mu, \sigma, \lambda)$ é a função densidade apresentada em (2.4). Note que $\alpha = \mathbb{P}(Y = c)$ representa a probabilidade de se observar zero ($c = 0$) ou um ($c = 1$), dependendo do valor de c . Já μ , σ e λ são parâmetros da distribuição *power logit* e representam, respectivamente, a mediana, dispersão e assimetria de $Y|Y \in (0, 1)$. A função densidade em (5.2) caracteriza uma distribuição *power logit* inflacionada no ponto c , já que, se $\alpha > 0$, a massa de probabilidade da distribuição *power logit* em $y = c$ é excedida. Podemos expressar $f_Y^{(c)}(y; \alpha, \mu, \sigma, \lambda)$ como

$$f_Y^{(c)}(y; \alpha, \mu, \sigma, \lambda) = \left\{ \alpha^{\mathbb{I}_{\{c\}}(y)} (1 - \alpha)^{1 - \mathbb{I}_{\{c\}}(y)} \right\} \left\{ f(y; \mu, \sigma, \lambda)^{1 - \mathbb{I}_{\{c\}}(y)} \right\};$$

isto é, $f_Y^{(c)}(y; \alpha, \mu, \sigma, \lambda)$ pode ser fatorada em dois termos, sendo que o primeiro depende somente de α e o segundo depende unicamente de μ , σ e λ . Essa fatoração será útil no processo de estimação.

Temos a seguinte definição.

Definição 5.1 *Seja $Y \sim PLI_c(\alpha, \mu, \sigma, \lambda; g)$. Então:*

1. *Se $c = 0$, temos a distribuição *power logit* inflacionada no ponto zero (ZIPL), e denotada por $Y \sim ZIPL(\alpha, \mu, \sigma, \lambda; g)$;*
2. *Se $c = 1$, temos a distribuição *power logit* inflacionada no ponto um (OIPL), e denotada por $Y \sim OIPL(\alpha, \mu, \sigma, \lambda; g)$.*

A Figura 5.1 apresenta gráficos das densidades ZIPL- $t_{(2)}$ e OIPL-slash $_{(2,5)}$ para diferentes valores dos parâmetros. Nestes gráficos, a barra vertical com o ponto acima representa $\alpha = \mathbb{P}(Y = 0)$ para a ZIPL e $\alpha = \mathbb{P}(Y = 1)$ para a OIPL. Conforme vimos no Capítulo 2, essas densidades podem apresentar diferentes formas, a depender dos valores dos parâmetros e do gerador de densidades.

5.2 O modelo de regressão *power logit* inflacionado em zero ou um

Considere Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes cada uma com função densidade *power logit* inflacionada no ponto c ($c = 0$ ou $c = 1$), dada por (5.2), isto é, $Y_i \sim IPL_c(\alpha_i, \mu_i, \sigma_i, \lambda; r)$, para $i = 1, \dots, n$. O modelo de regressão *power logit* inflacionado é definido por (5.2) e pelos componentes sistemáticos

$$\begin{aligned} d_0(\alpha_i) &= \mathbf{z}_i^\top \boldsymbol{\kappa} = \eta_{0i}, \\ d_1(\mu_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} = \eta_{1i}, \\ d_2(\sigma_i) &= \mathbf{s}_i^\top \boldsymbol{\tau} = \eta_{2i}, \end{aligned} \quad (5.3)$$

em que $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_m)^\top \in \mathbb{R}^m$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_q)^\top \in \mathbb{R}^q$ e $\lambda > 0$ são os parâmetros desconhecidos, $\mathbf{z}_i = (z_{i1}, \dots, z_{im})^\top$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ e $\mathbf{s}_i = (s_{i1}, \dots, s_{iq})^\top$ são observações das covariáveis com $m + p + q + 1 < n$, e η_{0i} , η_{1i} e η_{2i} são os preditores lineares. Assumimos que as funções de ligação $d_0 : (0, 1) \rightarrow \mathbb{R}$, $d_1 : (0, 1) \rightarrow \mathbb{R}$ e $d_2 : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. Diversas funções de ligação podem ser utilizadas.

O modelo definido por (5.2) e (5.3) engloba algumas classes de modelos. O modelo de regressão *power logit* inflacionado em zero ($c = 0$) e o modelo de regressão *power logit* inflacionado em um ($c = 1$). O modelo de regressão GJS inflacionado em zero ou um (Queiroz e Lemonte, 2021) é um caso particular quando $\lambda = 1$. Além disso, o modelo de regressão *power logit* é um caso limite quando $\alpha_i = \alpha \rightarrow 0$, com $i = 1, \dots, n$. Adicionalmente, quando $\lambda \rightarrow 0$, temos o modelo de regressão

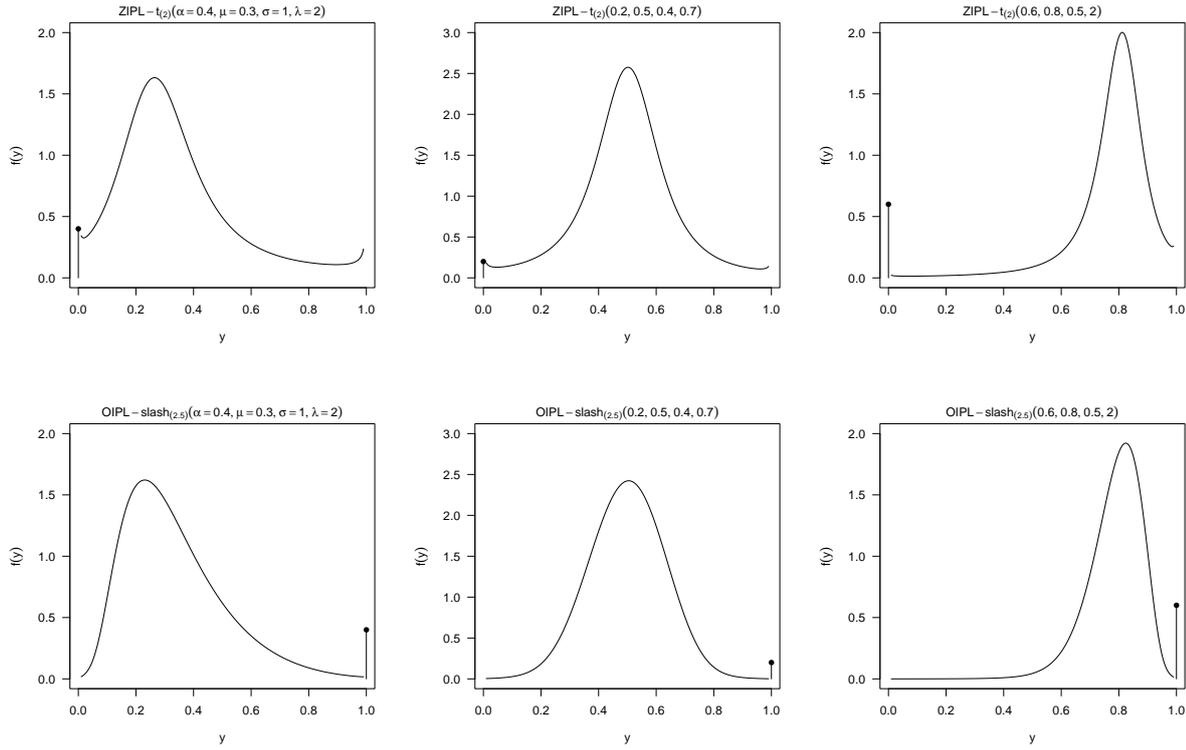


Figura 5.1: Densidades $\text{ZIPL-}t_{(2)}$ (primeira linha) e $\text{OIPL-slash}_{(2.5)}$ (segunda linha) para diferentes valores dos parâmetros.

log-log inflacionado. Finalmente, quando σ_i é constante, isto é, $\sigma_i = \sigma$, para $i = 1, \dots, n$, temos o modelo de regressão *power logit* inflacionado com dispersão constante.

Considerando $\theta = (\boldsymbol{\kappa}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$ o vetor de parâmetros do modelo, a função de verossimilhança de θ para o modelo de regressão *power logit* inflacionado é da forma

$$L(\theta) = \prod_{i=1}^n f^{(c)}(y_i; \alpha_i, \mu_i, \sigma_i, \lambda) = L_1(\boldsymbol{\kappa})L_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda),$$

em que

$$L_1(\boldsymbol{\kappa}) = \prod_{i=1}^n \alpha_i^{\mathbb{I}_{\{c\}}(y_i)} (1 - \alpha_i)^{1 - \mathbb{I}_{\{c\}}(y_i)},$$

$$L_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \prod_{i: y_i \in (0,1)} f(y_i; \mu_i, \sigma_i, \lambda).$$

Note que a função de verossimilhança $L(\theta)$ pode ser fatorada em dois termos, um que depende do vetor de parâmetros $\boldsymbol{\kappa}$, e outro que depende somente de $(\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$. Assim, dizemos que os parâmetros $\boldsymbol{\kappa}$ e $(\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$ são separáveis (Pace e Salvani, 1997) e a inferência por máxima verossimilhança sobre $(\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$ pode ser feita de forma independente de $\boldsymbol{\kappa}$. Note ainda que $L_1(\boldsymbol{\kappa})$ envolve somente os parâmetros utilizados para modelar a probabilidade de ocorrência de zero ou de um, enquanto que $L_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ envolve apenas os parâmetros utilizados para modelar a distribuição condicional da variável resposta no intervalo $(0, 1)$.

A função de log-verossimilhança para θ é dada por

$$\ell(\theta) = \ell_1(\boldsymbol{\kappa}) + \ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda), \quad (5.4)$$

em que

$$\begin{aligned}\ell_1(\boldsymbol{\kappa}) &= \sum_{i=1}^n \ell_i(\alpha_i), \\ \ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) &= \sum_{i: y_i \in (0,1)} \ell_i(\mu_i, \sigma_i, \lambda),\end{aligned}$$

sendo

$$\begin{aligned}\ell_i(\alpha_i) &= \mathbb{I}_{\{c\}}(y_i) \log(\alpha_i) + [1 - \mathbb{I}_{\{c\}}(y_i)] \log(1 - \alpha_i), \\ \ell_i(\mu_i, \sigma_i, \lambda) &= \log \lambda - \log \sigma_i - \log\{1 - y_i^\lambda\} + \log\{r(z_i^2)\},\end{aligned}\tag{5.5}$$

com $c = 0$ ou $c = 1$, dependendo do caso. Note que, para $i = 1, \dots, n$, $\mathbb{I}_{\{c\}}(Y_i)$ é uma variável aleatória Bernoulli, com probabilidade de sucesso $\mathbb{P}(\mathbb{I}_{\{c\}}(Y_i) = 1) = \alpha_i$, que é associada às covariáveis através do preditor linear η_{0i} e uma função de ligação $d_0(\cdot)$ da forma apresentada em (5.3). Assim, $\ell_1(\boldsymbol{\kappa})$ é a função de log-verossimilhança de um modelo linear generalizado com resposta binária (ver, por exemplo, McCullagh e Nelder (1989)). Adicionalmente, note que $\ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ é a função de log-verossimilhança de um modelo de regressão *power logit* em que a variável resposta é restrita às observações em $(0, 1)$. Dessa forma, a estimação dos parâmetros do modelo pode ser feita seguindo dois passos:

1. Ajuste um modelo de regressão binomial considerando a variável resposta $\mathbb{I}_{\{c\}}(Y_i)$, para $i = 1, \dots, n$, com função de ligação $d_0(\cdot)$ e preditor linear η_{0i} . Neste passo, obtemos as estimativas e erros padrão associados ao vetor $\boldsymbol{\kappa}$.
2. Considere o conjunto $\wp = \{i : y_i \in (0, 1)\}$ de tamanho n^\dagger . De (5.4), para estimar $(\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$, consideramos apenas as observações cujo valor da variável resposta está no intervalo $(0, 1)$. Assim, neste segundo passo, ajustamos o modelo de regressão *power logit* considerando as variáveis Y_i , para $i \in \wp$, funções de ligação $d_1(\cdot)$ e $d_2(\cdot)$ e vetores de covariáveis $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ e $\mathbf{s}_i = (s_{i1}, \dots, s_{iq})^\top$, para $i \in \wp$. Neste passo, também é estimado o parâmetro extra (ou vetor de parâmetros extras), se existir.

No que segue, definimos o vetor escore e a matriz da informação observada para $\boldsymbol{\theta}$. Inicialmente, notemos que

$$\mathbf{U}_r = \frac{\partial \ell_1(\boldsymbol{\kappa})}{\partial \kappa_r} = \sum_{i=1}^n \frac{\partial \ell_i(\alpha_i)}{\partial \alpha_i} \frac{d\alpha_i}{d\eta_{0i}} \frac{\partial \eta_{0i}}{\partial \kappa_r}, \quad r = 1, \dots, m,$$

em que

$$\frac{\partial \ell_i(\alpha_i)}{\partial \alpha_i} = \frac{\mathbb{I}_{\{c\}}(y_i) - \alpha_i}{\alpha_i(1 - \alpha_i)}, \quad \frac{d\alpha_i}{d\eta_{0i}} = \frac{1}{\dot{d}_0(\alpha_i)}.$$

Assim, como $\partial \eta_{0i} / \partial \kappa_r = z_{ir}$, temos que

$$\mathbf{U}_r = \sum_{i=1}^n \frac{\mathbb{I}_{\{c\}}(y_i) - \alpha_i}{\alpha_i(1 - \alpha_i)} \frac{1}{\dot{d}_0(\alpha_i)} z_{ir}, \quad r = 1, \dots, m.$$

Logo $\mathbf{U}_\kappa = \mathbf{Z}^\top \mathbf{A} \mathbf{T}_0 \boldsymbol{\alpha}^*$, em que $\mathbf{Z} = (z_1, \dots, z_n)^\top$ é uma matriz $n \times m$ de posto completo, cuja i -ésima linha ($i = 1, \dots, n$) é dada por $\mathbf{z}_i^\top = (z_{i1}, \dots, z_{im})$, $\mathbf{A} = \text{diag}\{\mathbb{I}_{\{c\}}(y_1) - \alpha_1, \dots, \mathbb{I}_{\{c\}}(y_n) - \alpha_n\}$, $\boldsymbol{\alpha}^* = (1/[\alpha_1(1 - \alpha_1)], \dots, 1/[\alpha_n(1 - \alpha_n)])^\top$ e $\mathbf{T}_0 = \text{diag}\{1/\dot{d}_0(\alpha_1), \dots, 1/\dot{d}_0(\alpha_n)\}$. Para obtermos \mathbf{U}_β , \mathbf{U}_τ e \mathbf{U}_λ , vamos considerar o desenvolvimento feito na Seção 3.2. Para tanto, considere a seguinte notação $\mathbf{V}^\dagger = \text{diag}\{v_1^*, \dots, v_n^*\}$ e $\mathbf{v}^\dagger = (v_1^*, \dots, v_n^*)$, em que $v_i^* = v_i$, se $i \in \wp$, e $v_i^* = 0$, caso contrário. Assim, por exemplo, \mathbf{W}^\dagger é uma matriz diagonal de dimensão $n \times n$, cujas colunas \mathbf{w}_i são iguais a zero para $i \notin \wp$. Analogamente, $\boldsymbol{\mu}^{*\dagger}$ é um vetor de tamanho n , com elementos μ_i^* ,

para $i \in \wp$, e elementos nulos para $i \notin \wp$. Dessa forma, o vetor escore para $\boldsymbol{\theta}$ é dado por

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{Z}^\top \mathbf{A} \mathbf{T}_0 \boldsymbol{\alpha}^* \\ \mathbf{X}^\top \mathbf{W}^\dagger \mathbf{T}_1^\dagger \boldsymbol{\mu}^{*\dagger} \\ \mathbf{S}^\top \mathbf{T}_2^\dagger \boldsymbol{\sigma}^{*\dagger} \\ \mathbf{1}_n^\top \boldsymbol{\lambda}^{*\dagger} \end{bmatrix}.$$

A segunda derivada de $\ell(\boldsymbol{\theta})$ com respeito a κ_r e κ_c , $R, c = 1, \dots, n$, é dada por

$$\begin{aligned} \frac{\partial^2 \ell_1(\boldsymbol{\kappa})}{\partial \kappa_r \partial \kappa_c} &= \sum_{i=1}^n \frac{\partial}{\partial \alpha_i} \left(\frac{\partial \ell_i(\alpha_i)}{\partial \alpha_i} \frac{d\alpha_i}{d\eta_{0i}} \right) \frac{d\alpha_i}{d\eta_{0i}} \frac{\partial \eta_{0i}}{\partial \kappa_c} z_{ir} \\ &= \sum_{i=1}^n \left(\frac{\partial \ell_i^2(\alpha_i)}{\partial \alpha_i^2} \frac{d\alpha_i}{d\eta_{0i}} + \frac{\partial \ell_i(\alpha_i)}{\partial \alpha_i} \frac{\partial}{\partial \alpha_i} \frac{d\alpha_i}{d\eta_{0i}} \right) \frac{d\alpha_i}{d\eta_{0i}} z_{ic} z_{ir}, \end{aligned}$$

com

$$\frac{\partial \ell_i^2(\alpha_i)}{\partial \alpha_i^2} = \frac{1 - \mathbb{I}_{\{c\}}(y_i)}{(1 - \alpha_i)^2} - \frac{\mathbb{I}_{\{c\}}(y_i)}{\alpha_i^2} \quad \text{e} \quad \frac{\partial}{\partial \alpha_i} \frac{d\alpha_i}{d\eta_{0i}} = -\frac{\ddot{d}_0(\alpha_i)}{\dot{d}_0(\alpha_i)^2}.$$

Assim, a matriz da informação observada para $\boldsymbol{\theta}$ é dada por

$$\mathbf{J}_n(\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}) = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\kappa}\boldsymbol{\kappa}} & \mathbf{0}_{m,p+q+1} \\ \mathbf{0}_{p+q+1,m} & \mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}) \end{bmatrix},$$

em que

$$\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}) = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\tau}} & \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\lambda}} \\ \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\tau}}^\top & \mathbf{J}_{\boldsymbol{\tau}\boldsymbol{\tau}} & \mathbf{J}_{\boldsymbol{\tau}\boldsymbol{\lambda}} \\ \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\lambda}}^\top & \mathbf{J}_{\boldsymbol{\tau}\boldsymbol{\lambda}}^\top & \mathbf{J}_{\boldsymbol{\lambda}\boldsymbol{\lambda}} \end{bmatrix},$$

sendo $\mathbf{J}_{\boldsymbol{\kappa}\boldsymbol{\kappa}} = \mathbf{Z}^\top \mathbf{W}_0 \mathbf{T}_0 \mathbf{Z}$, $\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{W}_1^\dagger \mathbf{T}_1^\dagger \mathbf{X}$, $\mathbf{J}_{\boldsymbol{\tau}\boldsymbol{\tau}} = \mathbf{S}^\top \mathbf{W}_2^\dagger \mathbf{T}_2^\dagger \mathbf{S}$, $\mathbf{J}_{\boldsymbol{\lambda}\boldsymbol{\lambda}} = \mathbf{1}_n^\top \mathbf{W}_3^\dagger \mathbf{1}_n$, $\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\tau}} = \mathbf{X}^\top \mathbf{W}_4^\dagger \mathbf{T}_1^\dagger \mathbf{T}_2^\dagger \mathbf{S}$, $\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\lambda}} = \mathbf{X}^\top \mathbf{W}_5^\dagger \mathbf{T}_1^\dagger \mathbf{1}_n$, $\mathbf{J}_{\boldsymbol{\tau}\boldsymbol{\lambda}} = \mathbf{S}^\top \mathbf{W}_6^\dagger \mathbf{T}_1^\dagger \mathbf{1}_n$, em que $\mathbf{W}_0 = \text{diag}\{w_1^{(0)}, \dots, w_n^{(0)}\}$, com

$$w_i^{(0)} = - \left(-\frac{(1 - \mathbb{I}_{\{c\}}(y_i))}{(1 - \alpha_i)^2} - \frac{\mathbb{I}_{\{c\}}(y_i)}{\alpha_i^2} \right) \dot{d}_0(\alpha_i)^{-1} + \left(\frac{\mathbb{I}_{\{c\}}(y_i)}{\alpha_i} - \frac{(1 - \mathbb{I}_{\{c\}}(y_i))}{(1 - \alpha_i)} \right) \frac{\ddot{d}_0(\alpha_i)}{\dot{d}_0(\alpha_i)^2}.$$

Note que, como os parâmetros são separáveis, a matriz $\mathbf{J}_n(\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda})$ é bloco diagonal. O estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ pode ser obtido através da maximização da função de log-verossimilhança dada em (5.4). Sob condições gerais de regularidade, esse estimador é consistente para $\boldsymbol{\theta}$ e

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_{m+p+q+1}(\mathbf{0}_{m+p+q+1}, \mathbf{K}(\boldsymbol{\theta})^{-1}), \quad \text{quando } n \rightarrow \infty,$$

em que $\mathbf{K}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda})$ é a matriz da informação de Fisher de $\boldsymbol{\theta}$ para uma única observação. Não existe uma expressão analítica fechada para $\mathbf{K}(\boldsymbol{\theta})$, mas resultados assintóticos mostram que podemos utilizar a matriz $\mathbf{J}_n(\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda})$ para realizar inferência sobre $\boldsymbol{\theta}$.

Finalmente, a log-verossimilhança $\ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda})$ pode apresentar os mesmos problemas de concavidade observados na estimação dos parâmetros dos modelos *power logit*, em geral, quando o tamanho da amostra é pequeno². Nestes casos, pode-se obter o estimador de máxima verossimilhança penalizada de forma análoga ao apresentado nas Seções 2.3.1 e 3.2.

²Note que, neste caso, nos referimos ao tamanho amostral n^\dagger .

5.3 Diagnóstico e influência no modelo de regressão *power logit* inflacionado em zero ou um

A análise de resíduos nos modelos de regressão *power logit* inflacionados em zero ou um pode ser feita de duas formas:

- (i) fazer a análise separadamente para as partes discreta e contínua do modelo, utilizando dois tipos de resíduos;
- (ii) utilizar um único resíduo para os dois componentes do modelo (discreto e contínuo).

Para realizar a análise de forma separada, pode-se utilizar diferentes tipos de resíduos. Para a parte contínua, sugerimos utilizar os resíduos propostos na Seção 3.3, para os modelos de regressão *power logit*. Já para a parte discreta, sugerimos o resíduo deviance ou o resíduo padronizado, definido como

$$r_i^p = \frac{\mathbb{I}_{\{c\}}(y_i) - \hat{\alpha}_i}{\sqrt{\hat{\alpha}_i(1 - \hat{\alpha}_i)(1 - \hat{h}_{ii}^*)}}, \quad i = 1, \dots, n,$$

em que \hat{h}_{ii}^* é o i -ésimo elemento diagonal de $\mathbf{H}^* = \mathbf{Q}^{1/2} \mathbf{Z}(\mathbf{Z}^\top \mathbf{Q} \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{Q}^{1/2}$ avaliado na estimativa de máxima verossimilhança, sendo $\mathbf{Q} = \text{diag}\{q_1, \dots, q_n\}$, com $q_i = 1/[\alpha_i(1 - \alpha_i)d_0(\alpha_i)^2]$, para $i = 1, \dots, n$. No entanto, outros resíduos existentes na literatura podem ser utilizados.

Por outro lado, se o objetivo é utilizar um único resíduo para os dois componentes do modelo, um candidato natural é o resíduo quantílico aleatorizado (Dunn e Smyth, 1996). No caso dos modelos de regressão *power logit*, esse resíduo é definido por

$$r_i^q = \begin{cases} \Phi^{-1}(u_i), & y_i = c, \\ \Phi^{-1}(F^{(c)}(y_i; \tilde{\alpha}_i, \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda})), & y_i \in (0, 1) \end{cases}$$

em que u_i é uma variável aleatória com distribuição uniforme no intervalo $(a_i, b_i]$, com $a_i = \lim_{y \uparrow y_i} \text{IPL}^{(c)}(y_i; \tilde{\alpha}_i, \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda})$ e $b_i = \text{IPL}^{(c)}(y_i; \tilde{\alpha}_i, \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda})$, para $i = 1, \dots, n$.

No modelo de regressão PL inflacionado em zero, temos que o ponto de massa é no zero, assim, $y_i = 0$ e

$$a_i = \lim_{y \uparrow 0} \text{IPL}^{(0)}(y; \tilde{\alpha}_i, \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda}) = \alpha_i \lim_{y \uparrow 0} \mathbb{I}_{[0, \infty)}(y) + (1 - \hat{\alpha}_i) \lim_{y \uparrow 0} F(y; \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda}) = 0,$$

$$b_i = \text{IPL}^{(0)}(0; \tilde{\alpha}_i, \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda}) = \hat{\alpha}_i,$$

e, assim, $u_i \sim U(0, \hat{\alpha}_i]$, para $i = 1, \dots, n$. No modelo de regressão PL inflacionado em um, temos que

$$a_i = \lim_{y \uparrow 1} \text{IPL}^{(1)}(y; \hat{\alpha}_i, \hat{\xi}_i, \hat{\delta}_i) = \alpha_i \lim_{y \uparrow 1} \mathbb{I}_{[1, \infty)}(y) + (1 - \hat{\alpha}_i) \lim_{y \uparrow 1} F(y; \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda}) = (1 - \hat{\alpha}_i),$$

$$b_i = \text{IPL}^{(1)}(1; \tilde{\alpha}_i, \tilde{\mu}_i, \tilde{\sigma}_i, \tilde{\lambda}) = 1,$$

e, assim, $u_i \sim U(1 - \hat{\alpha}_i, 1]$. O gráfico de r_i^q pode revelar observações atípicas ou alguma tendência sistemática. O resíduo r_i^q tem distribuição assintótica normal com média zero e variância um. Esse resíduo pode variar de uma realização para outra em situações práticas, uma vez que o resíduo referente à parte discreta é obtido através de um processo de aleatorização.

No que segue, desenvolvemos o método de influência local sob dois esquemas de perturbação (perturbação de casos e perturbação nas variáveis explicativas) para o modelo de regressão *power logit* inflacionado em zero ou um. Assim como na Seção 3.4, utilizamos a abordagem de Cook (1986). A seguir, apresentamos a matriz Δ para cada um dos esquemas de perturbação. Os detalhes técnicos podem ser encontrados na Seção C.2.

- Ponderação de casos

$$\Delta = \begin{bmatrix} \mathbf{Z}^\top \widehat{\mathbf{A}} \widehat{\mathbf{T}}_0 \widehat{\mathbf{D}}_\kappa \\ \mathbf{X}^\top \widehat{\mathbf{W}}^\dagger \widehat{\mathbf{T}}_1 \widehat{\mathbf{D}}_\beta^\dagger \\ \mathbf{S}^\top \widehat{\mathbf{T}}_2 \widehat{\mathbf{D}}_\tau^\dagger \end{bmatrix}.$$

- Perturbação individual das covariáveis do componente discreto

$$\Delta = \begin{bmatrix} -\sigma_z \widehat{\kappa}_u \mathbf{Z}^\top \widehat{\mathbf{W}}_0 \widehat{\mathbf{T}}_0 + \sigma_z \mathbf{c}_\kappa^u \widehat{\boldsymbol{\alpha}}^{*\top} \widehat{\mathbf{A}} \widehat{\mathbf{T}}_0 \\ \mathbf{0}_{p,n} \\ \mathbf{0}_{q,n} \end{bmatrix}.$$

- Perturbação individual das covariáveis da mediana condicional

$$\Delta = \begin{bmatrix} \mathbf{0}_{m,n} \\ -\sigma_x \widehat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1^\dagger \widehat{\mathbf{T}}_1^\dagger + \sigma_x \mathbf{c}_\beta^j \widehat{\boldsymbol{\mu}}^{*\dagger\top} \widehat{\mathbf{W}}^\dagger \widehat{\mathbf{T}}_1^\dagger \\ -\sigma_x \widehat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger \end{bmatrix}.$$

- Perturbação individual das covariáveis da dispersão condicional

$$\Delta = \begin{bmatrix} \mathbf{0}_{m,n} \\ -\sigma_s \widehat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger, \\ -\sigma_s \widehat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2^\dagger \widehat{\mathbf{T}}_2^\dagger + \sigma_s \mathbf{c}_\tau^k \widehat{\boldsymbol{\sigma}}^{*\dagger\top} \widehat{\mathbf{T}}_2^\dagger \end{bmatrix}.$$

- Perturbação simultânea das covariáveis do componente contínuo

$$\Delta = \begin{bmatrix} \mathbf{0}_{m,n} \\ -\sigma_x \widehat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1^\dagger \widehat{\mathbf{T}}_1^\dagger + \sigma_x \mathbf{c}_\beta^j \widehat{\boldsymbol{\mu}}^{*\dagger\top} \widehat{\mathbf{W}}^\dagger \widehat{\mathbf{T}}_1^\dagger - \sigma_s \widehat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger \\ -\sigma_s \widehat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2^\dagger \widehat{\mathbf{T}}_2^\dagger + \sigma_s \mathbf{c}_\tau^k \widehat{\boldsymbol{\sigma}}^{*\dagger\top} \widehat{\mathbf{T}}_2^\dagger - \sigma_x \widehat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger \end{bmatrix}.$$

- Perturbação simultânea do componente discreto e da mediana condicional

$$\Delta = \begin{bmatrix} -\sigma_z \widehat{\kappa}_u \mathbf{Z}^\top \widehat{\mathbf{W}}_0 \widehat{\mathbf{T}}_0 + \sigma_z \mathbf{c}_\kappa^u \widehat{\boldsymbol{\alpha}}^{*\top} \widehat{\mathbf{A}} \widehat{\mathbf{T}}_0 \\ -\sigma_x \widehat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1^\dagger \widehat{\mathbf{T}}_1^\dagger + \sigma_x \mathbf{c}_\beta^j \widehat{\boldsymbol{\mu}}^{*\dagger\top} \widehat{\mathbf{W}}^\dagger \widehat{\mathbf{T}}_1^\dagger \\ -\sigma_x \widehat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger \end{bmatrix}.$$

- Perturbação simultânea do componente discreto e da dispersão condicional

$$\Delta = \begin{bmatrix} -\sigma_z \widehat{\kappa}_u \mathbf{Z}^\top \widehat{\mathbf{W}}_0 \widehat{\mathbf{T}}_0 + \sigma_z \mathbf{c}_\kappa^u \widehat{\boldsymbol{\alpha}}^{*\top} \widehat{\mathbf{A}} \widehat{\mathbf{T}}_0 \\ -\sigma_s \widehat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger, \\ -\sigma_s \widehat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2^\dagger \widehat{\mathbf{T}}_2^\dagger + \sigma_s \mathbf{c}_\tau^k \widehat{\boldsymbol{\sigma}}^{*\dagger\top} \widehat{\mathbf{T}}_2^\dagger \end{bmatrix}.$$

- Perturbação simultânea das covariáveis do componente discreto e contínuo

$$\Delta = \begin{bmatrix} -\sigma_z \widehat{\kappa}_u \mathbf{Z}^\top \widehat{\mathbf{W}}_0 \widehat{\mathbf{T}}_0 + \sigma_z \mathbf{c}_\kappa^u \widehat{\boldsymbol{\alpha}}^{*\top} \widehat{\mathbf{A}} \widehat{\mathbf{T}}_0 \\ -\sigma_x \widehat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1^\dagger \widehat{\mathbf{T}}_1^\dagger + \sigma_x \mathbf{c}_\beta^j \widehat{\boldsymbol{\mu}}^{*\dagger\top} \widehat{\mathbf{W}}^\dagger \widehat{\mathbf{T}}_1^\dagger - \sigma_s \widehat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger \\ -\sigma_s \widehat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2^\dagger \widehat{\mathbf{T}}_2^\dagger + \sigma_s \mathbf{c}_\tau^k \widehat{\boldsymbol{\sigma}}^{*\dagger\top} \widehat{\mathbf{T}}_2^\dagger - \sigma_x \widehat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger \end{bmatrix}.$$

5.4 Implementação computacional

O ajuste do modelo de regressão *power logit* inflacionado em zero ou um pode ser feito utilizando as funções `glm` e `PLreg` disponíveis no *software* R através dos pacotes `stats` e `PLreg`, respectivamente. Esse ajuste é feito seguindo os passos apresentados na Seção 5.2. Dessa forma, utilizamos a função `glm` para estimar os parâmetros do componente discreto, e a função `PLreg` para estimar os parâmetros do componente contínuo. Nesse processo, também é possível realizar a análise de diagnóstico separada. Esse processo está implementado na função `IPLreg`, disponível em <https://github.com/ffqueiroz/IPLreg>:

```
IPLreg(formula, data, family, zeta = 2, type = "pML", c)
```

O argumento `formula` deve conter quatro partes (separadas pelos símbolos “ \sim ”, “ $|$ ” e “ $|$ ”), a saber: variável resposta observada no intervalo $[0, 1)$ ou $(0, 1]$, preditor linear do submodelo da mediana, da dispersão e da probabilidade de ocorrer o valor $c = 0$ ou $c = 1$. Por exemplo, se `formula = y ~ x1 + x2 | s1 + s2 + s3 | z1` é fornecido, ele descreve y para a variável resposta, $x1$ e $x2$ para o submodelo da mediana, $s1$, $s2$ e $s3$ para o submodelo da dispersão e $z1$ para o submodelo da probabilidade de observar c do modelo de regressão *power logit* inflacionado em c . Se algum submodelo não tiver covariável, deve-se colocar o valor 1 na respectiva parte da fórmula. Assim, um modelo com dispersão constante pode ter, por exemplo, a seguinte fórmula: $y \sim x1 + x2 | 1 | z1$. O gerador de densidade da distribuição *power logit* deve ser fornecido no argumento `family` e o parâmetro extra no argumento `zeta`. O argumento `c` identifica se o modelo é inflacionado em zero ($c=0$) ou em um ($c=1$). As funções de ligação padrão são logito para a mediana e a probabilidade de ocorrer c e logarítmica para a dispersão. Configurações extras podem ser adicionadas através da modificação dessa função pelo usuário.

A função `IPLreg` retorna uma lista com três objetos: “`fitPL`”, “`fitglm`” e “`residual`”, que representam os ajustes feitos para a parte contínua (através do pacote `PLreg`), para a parte discreta (através da função `glm`) e o resíduo quantílico aleatorizado para o modelo ajustado. Com os objetos “`fitPL`” e “`fitglm`”, o usuário pode realizar a análise de diagnóstico separada.

5.5 Aplicação

Muscarella *et al.* (2020) estudam dados referentes à abundância de palmeiras arbóreas em biomas de floresta de folhas largas úmidas tropicais e subtropicais. Um dos objetivos dos autores é estudar a relação entre a abundância de palmeiras arbóreas e variáveis climáticas. Entende-se por palmeiras arbóreas aquelas que atingem 10 cm de diâmetro à altura do peito, e assim, estão incluídas em parcelas típicas de inventário florestal. Os dados utilizados são referentes a 786 locais situados ao redor do mundo (ver Muscarella *et al.* (2020) para detalhes adicionais)³. A variável resposta de interesse é Y , a área basal relativa de palmeiras arbóreas, definida como a soma da área basal das palmeiras dividida pela área basal total. Seguindo a análise feita pelos autores, consideramos as seguintes covariáveis:

- pma: precipitação média anual (em mm por ano);
- pts: precipitação no trimestre mais seco (em mm por trimestre);
- etc: capacidade de troca de cátions extraídos (cmol+ por kg) – pode ser vista como a fertilidade do solo;
- plf: profundidade do lençol freático (em m);
- ac: altura da copa

³Os dados são referentes a 842 locais, mas após excluir os locais com observações faltantes para alguma covariável chegamos a 786 locais.

- dm: densidade média da madeira;
- area: área do local amostrado;
- reino: reino biogeográfico (*Neotropics*, *Afrotropics*, *IndoMalay*, *Australasia/Oceania*).

Ajustaremos o modelo de regressão *power logit* normal inflacionado em zero aos dados. Assim, assumimos que $Y_i \sim \text{IPL-NO}_0(\alpha_i, \mu_i, \sigma_i, \lambda; r)$, para $i = 1, \dots, 786$, são variáveis aleatórias independentes, tais que

$$\begin{aligned} \log\left(\frac{\alpha_i}{1-\alpha_i}\right) &= \kappa_0 + \kappa_1 \text{area}_i + \kappa_2 \text{pma}_i + \kappa_3 \text{pts}_i + \kappa_4 \text{ctc}_i + \kappa_5 \text{plf}_i + \kappa_6 \text{ac}_i + \kappa_7 \text{dm}_i \\ &\quad + \kappa_8 \text{reino_AustraOcean}_i + \kappa_9 \text{reino_IndoMalay}_i + \kappa_{10} \text{reino_Neotropics}_i, \\ \log\left(\frac{\mu_i}{1-\mu_i}\right) &= \beta_0 + \beta_1 \text{area}_i + \beta_2 \text{pma}_i + \beta_3 \text{pts}_i + \beta_4 \text{ctc}_i + \beta_5 \text{plf}_i + \beta_6 \text{ac}_i + \beta_7 \text{dm}_i \\ &\quad + \beta_8 \text{reino_AustraOcean}_i + \beta_9 \text{reino_IndoMalay}_i + \beta_{10} \text{reino_Neotropics}_i, \\ \log \sigma_i &= \tau_0 + \tau_1 \text{area}_i + \tau_2 \text{pma}_i + \tau_3 \text{pts}_i + \tau_4 \text{ctc}_i + \tau_5 \text{plf}_i + \tau_6 \text{ac}_i + \tau_7 \text{dm}_i \\ &\quad + \tau_8 \text{reino_AustraOcean}_i + \tau_9 \text{reino_IndoMalay}_i + \tau_{10} \text{reino_Neotropics}_i. \end{aligned}$$

Os resultados do ajuste são apresentados na Tabela 5.1. Para fins de comparação, ajustamos o modelo de regressão beta inflacionado em zero aos dados, considerando as mesmas covariáveis para os submodelos da probabilidade da variável assumir valor zero (α), da média condicional (μ) e da precisão condicional (σ) (Ospina e Ferrari, 2012). Os resultados do ajuste são apresentados na Tabela 5.2. Os modelos *power logit* normal inflacionado e beta inflacionado diferem apenas no componente contínuo; para os dois modelos, o componente discreto é modelado através de um modelo de regressão logística. Na Figura 5.2 apresentamos os gráficos de resíduos com envelope simulado para a parte discreta e contínua, considerando os ajustes dos modelos PL normal inflacionado e beta inflacionado. O resíduo utilizado para a parte discreta é o resíduo deviance e, conforme observamos na Figura 5.2(a), o modelo de regressão logística apresenta um ajuste satisfatório. Para o modelo de regressão beta inflacionado, utilizamos o resíduo padronizado ponderado 2 (Espinheira *et al.*, 2008) para a parte contínua e a Figura 5.2(b) sugere que esse modelo não é adequado. Finalmente, o gráfico do resíduo padronizado para a parte contínua do modelo PL-NO inflacionado indica que o ajuste desse modelo é razoável; veja a Figura 5.2(c).

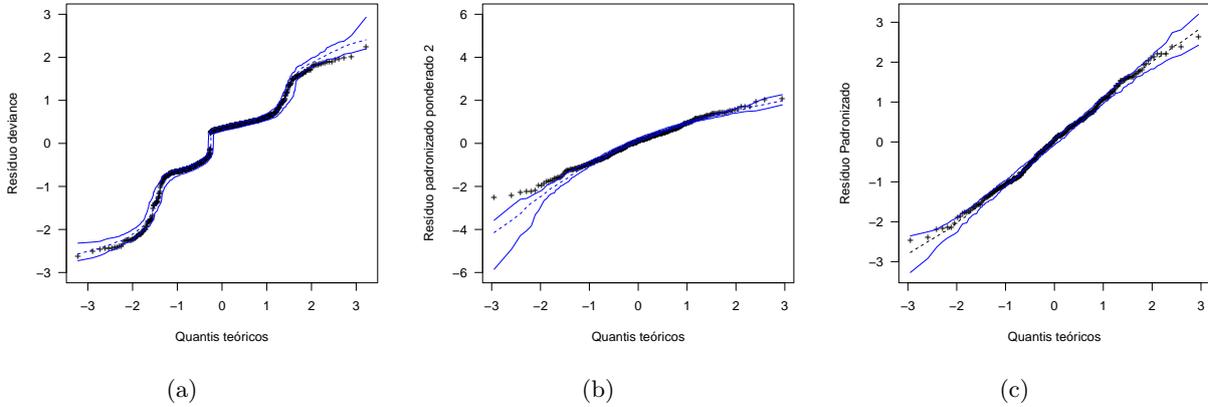
Tabela 5.1: Estimativas (*Est.*), erros padrão assintóticos (*e.p.*) e valores *p* para o modelo de regressão IPL-NO com dispersão variável.

	$\log[\alpha/(1-\alpha)]$			$\log[\mu/(1-\mu)]$			$\log \sigma$		
	Est.	e.p.	valor <i>p</i>	Est.	e.p.	valor <i>p</i>	Est.	e.p.	valor <i>p</i>
intercepto	2.081	0.290	< 0.01	-4.878	0.315	< 0.01	-0.337	0.285	0.236
area	-0.217	0.098	0.026	-0.250	0.127	0.049	0.062	0.045	0.164
pma	-0.434	0.144	< 0.01	0.324	0.102	< 0.01	-0.108	0.069	0.114
pts	-0.237	0.123	0.054	-0.091	0.092	0.323	-0.036	0.060	0.552
ctc	0.155	0.126	0.219	-0.320	0.091	< 0.01	-0.005	0.053	0.926
plf	0.013	0.106	0.905	-0.122	0.097	0.206	0.037	0.051	0.477
ac	0.022	0.135	0.868	-0.443	0.124	< 0.01	-0.242	0.071	< 0.01
dm	-0.133	0.121	0.273	-0.718	0.085	< 0.01	0.009	0.055	0.872
reino_AustraOcean	-1.926	0.475	< 0.01	-1.006	0.463	0.030	-0.028	0.251	0.913
reino_IndoMalay	0.283	0.396	0.475	-0.271	0.414	0.514	0.100	0.224	0.656
reino_Neotropics	-3.856	0.347	< 0.01	1.606	0.326	< 0.01	-0.062	0.190	0.744
λ	0.420	0.138							
AIC	-778.0								

A Tabela 5.1 revela que apenas as covariáveis area, pma, pts e reino são marginalmente significantes para explicar a probabilidade de não haver palmeiras arbóreas; area, pma, etc, ac, dm e reino são marginalmente significantes para explicar a mediana condicional da variável resposta e apenas a covariável ac é marginalmente significante para o submodelo da dispersão condicional. Considere o modelo reduzido que tem apenas essas covariáveis, isto é,

Tabela 5.2: Estimativas (*Est.*), erros padrão assintóticos (*e.p.*) e valores *p* para o modelo de regressão beta inflacionado com precisão variável.

	$\log[\alpha/(1-\alpha)]$			$\log[\mu/(1-\mu)]$			$\log \sigma$		
	Est.	e.p.	valor <i>p</i>	Est.	e.p.	valor <i>p</i>	Est.	e.p.	valor <i>p</i>
intercepto	2.081	0.290	< 0.01	-4.055	0.260	< 0.01	3.684	0.373	< 0.01
area	-0.217	0.098	0.026	-0.127	0.065	0.052	0.055	0.087	0.529
pma	-0.434	0.144	< 0.01	0.160	0.087	0.066	0.104	0.137	0.447
pts	-0.237	0.123	0.054	-0.066	0.079	0.406	0.078	0.122	0.525
ctc	0.155	0.126	0.219	-0.238	0.073	< 0.01	0.153	0.105	0.145
plf	0.013	0.106	0.905	-0.064	0.073	0.385	-0.032	0.097	0.740
ac	0.022	0.135	0.868	-0.599	0.097	< 0.01	0.815	0.139	< 0.01
dm	-0.133	0.121	0.273	-0.597	0.070	< 0.01	0.412	0.104	< 0.01
reino_AustraOcean	-1.926	0.475	< 0.01	-0.723	0.404	0.073	0.468	0.557	0.401
reino_IndoMalay	0.283	0.396	0.475	-0.105	0.347	0.761	-0.138	0.486	0.776
reino_Neotropics	-3.856	0.347	< 0.01	1.316	0.271	< 0.01	-0.892	0.392	0.023
AIC	-754.0								

**Figura 5.2:** Gráficos de resíduos com envelopes simulados para a parte discreta (a) e para a parte contínua considerando os modelos beta inflacionado (b) e PL normal inflacionado (c).

$$\begin{aligned} \log\left(\frac{\alpha_i}{1-\alpha_i}\right) &= \kappa_0 + \kappa_1 \text{area}_i + \kappa_2 \text{pma}_i + \kappa_3 \text{pts}_i + \kappa_4 \text{reino_AustraOcean}_i \\ &\quad + \kappa_5 \text{reino_IndoMalay}_i + \kappa_6 \text{reino_Neotropics}_i, \\ \log\left(\frac{\mu_i}{1-\mu_i}\right) &= \beta_0 + \beta_1 \text{area}_i + \beta_2 \text{pma}_i + \beta_3 \text{ctc}_i + \beta_4 \text{ac}_i + \beta_5 \text{dm}_i + \beta_6 \text{reino_AustraOcean}_i \\ &\quad + \beta_7 \text{reino_IndoMalay}_i + \beta_8 \text{reino_Neotropics}_i, \\ \log \sigma_i &= \tau_0 + \tau_1 \text{ac}_i. \end{aligned}$$

A estatística do teste da razão de verossimilhanças para testar o modelo reduzido contra o modelo completo é 19.89 (valor $p = 0.866$). Isto é, o modelo reduzido não é rejeitado considerando os níveis de significância usuais e, portanto, aquelas covariáveis podem ser excluídas do modelo. O resultado do ajuste do modelo reduzido é apresentado na Tabela 5.3. A estimativa de λ é 0.307 e o intervalo de 95% de confiança é (0.127, 0.487), indicando que o modelo de regressão GJS normal inflacionado em zero ($\lambda = 1$) não é indicado.

As estimativas dos parâmetros revelam algumas considerações interessantes, muitas delas observadas em Muscarella *et al.* (2020). A probabilidade de não ocorrência de palmeiras arbóreas é negativamente relacionada com as covariáveis area, pma e pts; assim, palmeiras arbóreas tendem a não estar presentes em locais com baixa precipitação (tanto anual, quanto em estações secas), como esperado. Além disso, quanto maior a área amostrada, menor é a probabilidade de não se

observar palmeiras arbóreas. Essas palmeiras tendem a estar mais presentes no reino Neotropics do que nos demais reinos. A mediana condicional da área relativa basal de palmeiras arbóreas está negativamente relacionada com as variáveis *area*, *ctc*, *ac* e *dm*, e positivamente relacionada com a variável *pma*; isto é, uma relação negativa com a área do local amostrado, fertilidade do solo, altura da copa e densidade de madeira, e uma relação positiva com a precipitação média anual. Assim, as palmeiras arbóreas representam uma maior proporção da área basal total em locais com maior volume de chuvas e menor fertilidade do solo.

Tabela 5.3: Estimativas (*Est.*), erros padrão assintóticos (*e.p.*) e valores *p* para o modelo de regressão reduzido *PL-NO* com precisão variável.

	Est.	e.p.	valor <i>p</i>
$\log[\alpha/(1 - \alpha)]$			
intercepto	1.956	0.240	< 0.01
area	-0.235	0.097	0.016
pma	-0.430	0.136	< 0.01
pts	-0.214	0.120	0.074
reino_AustraOcean	-1.696	0.406	< 0.01
reino_IndoMalay	0.464	0.335	0.166
reino_Neotropics	-3.741	0.293	< 0.01
$\log[\mu/(1 - \mu)]$			
intercepto	-4.890	0.305	< 0.01
area	-0.164	0.071	0.020
pma	0.242	0.081	< 0.01
ctc	-0.346	0.087	< 0.01
ac	-0.423	0.123	< 0.01
dm	-0.706	0.086	< 0.01
reino_AustraOcean	-1.087	0.483	0.024
reino_IndoMalay	-0.304	0.399	0.447
reino_Neotropics	1.647	0.316	< 0.01
$\log \sigma$			
intercepto	-0.527	0.146	< 0.01
ac	-0.288	0.057	< 0.01
λ	0.307	0.092	
AIC	-788.1		

Gráficos de diagnóstico para o modelo ajustado são apresentados na Figura 5.3. Os gráficos de resíduo foram feitos para a parte discreta, contínua e para ambas as partes (considerando o resíduo quantílico aleatorizado; veja as Figuras 5.3(a)-5.3(f). Esses gráficos evidenciam que o ajuste do modelo de regressão *power logit* normal inflacionado em zero apresentou ajuste satisfatório aos dados em estudo. Gráficos de influência são apresentados nas Figuras 5.3(g)-5.3(i). Algumas observações aparecem levemente destacadas nos gráficos de índice de $|h_{\max}|$ para o esquema de ponderação de casos (parte contínua) e no gráfico da distância de Cook para a parte discreta. A exclusão dessas observações, no entanto, não altera o resultado do ajuste de maneira significativa.

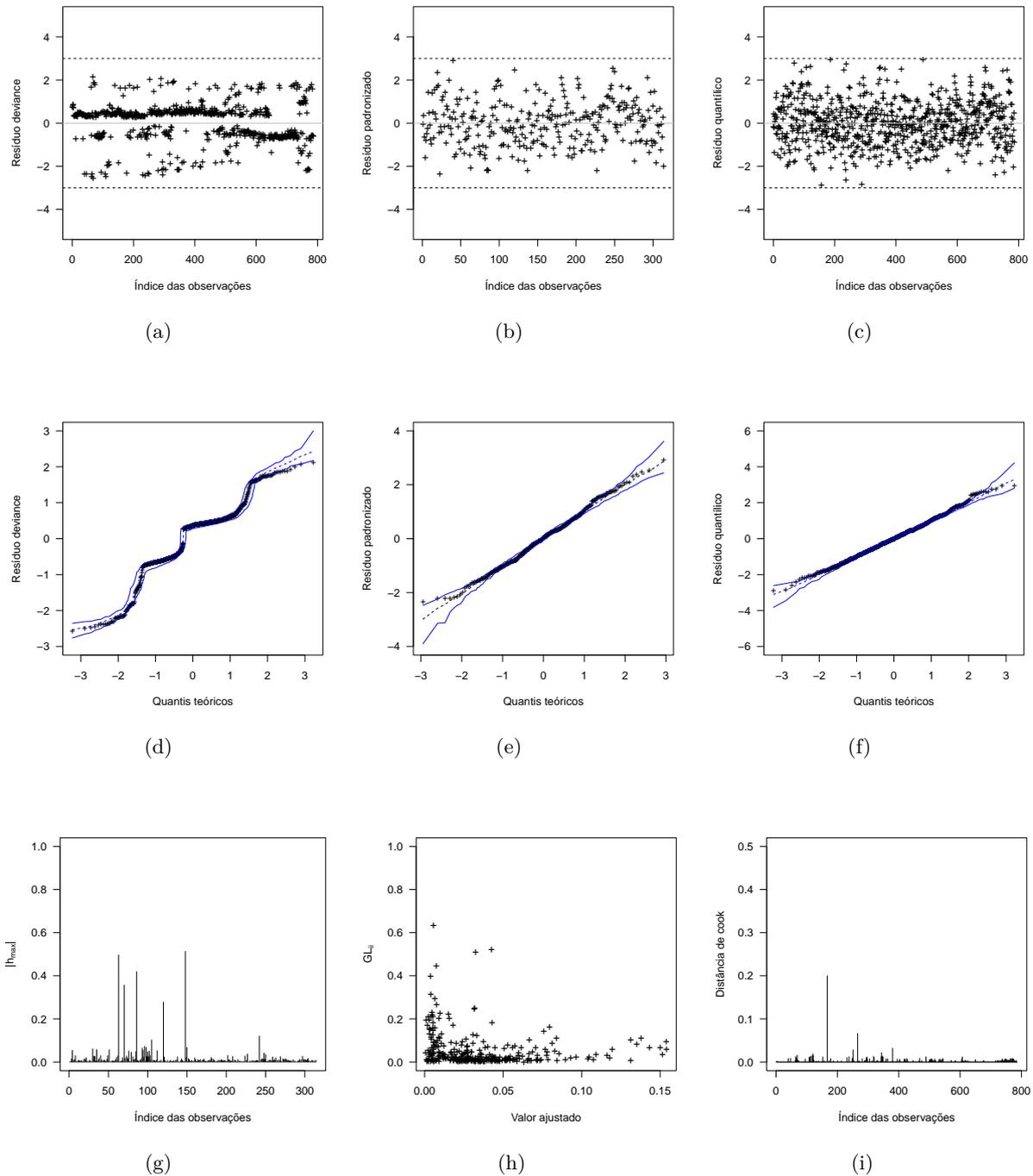


Figura 5.3: Gráficos de índice do resíduo deviance para a parte discreta (a), padronizado para a parte contínua (b) e quantílico para o modelo completo (c), respectivos gráficos de resíduo com envelopes simulados (d), (e) e (f), gráfico de índices de $|h_{max}|$ para o esquema de ponderação de casos para a parte contínua (g), diagrama de dispersão dos valores ajustados versus GL_{ii} para a parte contínua (h) e gráfico de índices da distância de Cook para a parte discreta (i).

Capítulo 6

Inferência robusta no modelo de regressão beta inflacionado

O modelo de regressão beta inflacionado foi proposto por [Ospina e Ferrari \(2012\)](#) e tem sido bastante utilizado para a modelagem de dados no intervalo unitário que apresentam os valores zero e/ou um, por exemplo, a taxa de mortalidade ou infecção para uma determinada doença, a proporção da renda familiar gasta com educação dos filhos, a proporção de pessoas que apresentam uma determinada característica, etc.. Esse modelo é uma extensão natural do modelo de regressão beta ([Ferrari e Cribari-Neto, 2004](#)). Embora o modelo de regressão beta inflacionado apresente um bom ajuste em várias situações, a inferência via máxima verossimilhança neste modelo é fortemente influenciada por observações atípicas. Neste capítulo, estudamos a falta de robustez da inferência via máxima verossimilhança neste modelo e propomos métodos de estimação robustos para os parâmetros do modelo.

O capítulo é organizado como segue. Na Seção 6.1 apresentamos uma breve revisão sobre conceitos associados à inferência robusta, bem como os problemas de inferência via máxima verossimilhança nos modelos de regressão para dados binários e beta na presença de contaminação. O desenvolvimento de métodos de estimação robustos para o modelo de regressão beta inflacionado em zero ou um é apresentado na Seção 6.2. Na Seção 6.3 são feitas aplicações em dados simulados e, por fim, estudos de simulação são conduzidos na Seção 6.4.

6.1 Preliminares

Nesta seção abordaremos os principais conceitos e resultados necessários para o desenvolvimento do método de estimação robusto no modelo de regressão beta inflacionado em zero ou um. Iniciamos esta seção abordando alguns conceitos relacionados à inferência robusta. Em seguida, estudamos o efeito de contaminações nos modelos de regressão para dados binários e em regressão beta. Faremos isso de forma separada pois, conforme será apresentado na Seção 6.2, o modelo de regressão beta inflacionado pode ser visto como uma mistura desses dois modelos. A notação utilizada nas Seções 6.1.2 e 6.1.3 deve ser lida de forma independente.

6.1.1 Conceitos em inferência robusta

Essa seção é baseada no Capítulo 2 de [Ribeiro \(2020\)](#). Para mais detalhes, pode-se consultar o referido trabalho.

Sejam Y_1, \dots, Y_n uma amostra aleatória de tamanho n , em que a função de distribuição de Y_1 pertence à família de distribuições paramétrica $\mathcal{F}_{\Theta} = \{F_{\theta}, \theta \in \Theta \subset \mathbb{R}^p\}$, $p \geq 1$. Seja f_{θ} a função densidade de Y_1 . Consideremos estimadores funcionais. Tais estimadores dependem dos dados apenas através da função de distribuição empírica, $F_n(\cdot)$, e satisfazem

$$\mathbf{T}_n(Y_1, \dots, Y_n) = \mathbf{T}(F_n).$$

Neste caso, dizemos que $\mathbf{T}(F_n)$ é um estimador funcional. Uma propriedade importante relacionada a estimadores funcionais é a de Fisher-consistência (Kallianpur e Rao, 1955).

Definição 6.1 (Estimador Fisher-consistente.) Dizemos que $\mathbf{T}(F_n)$ é um estimador Fisher-consistente para $\boldsymbol{\theta}$ se é funcional e satisfaz

$$\mathbf{T}(F_{\boldsymbol{\theta}}) = \boldsymbol{\theta}, \forall \boldsymbol{\theta} \in \Theta.$$

A propriedade de Fisher-consistência garante que o estimador atinge o verdadeiro valor de $\boldsymbol{\theta}$ quando é calculado sob a distribuição populacional dos dados.

A seguir, definimos a função de influência, proposta por Hampel (1974).

Definição 6.2 (Função de influência.) Seja $F_{h,y} = (1-h)F_{\boldsymbol{\theta}} + h\Delta_y$ a função de distribuição contaminada, obtida após a introdução da contaminação infinitesimal h no ponto y . A função de influência do estimador $\mathbf{T}(\cdot)$ em $F_{\boldsymbol{\theta}}$ é dada por

$$IF(y; \mathbf{T}, F_{\boldsymbol{\theta}}) = \lim_{h \rightarrow 0} \frac{\mathbf{T}((1-h)F_{\boldsymbol{\theta}} + h\Delta_y) - \mathbf{T}(F_{\boldsymbol{\theta}})}{h},$$

em que Δ_y é a medida de probabilidade que coloca toda a massa em y .

A função de influência pode ser interpretada como sendo o efeito causado no estimador $\mathbf{T}(\cdot)$ após uma contaminação infinitesimal h no ponto y . Também, segundo Hampel et al. (2011), tal função mede o viés assintótico causado pela contaminação nos dados.

Uma medida de robustez bastante conhecida e baseada na função de influência é a sensibilidade a erro grosseiro não-padronizada, dada por

$$\gamma^* = \sup_y \|IF(y; \mathbf{T}, F_{\boldsymbol{\theta}})\|,$$

em que $\|\cdot\|$ denota a norma euclidiana. A quantidade γ^* mede o viés máximo causado na estimativa de $\mathbf{T}(F_n)$ devido à contaminação infinitesimal introduzida. Dessa forma, uma característica desejável é que γ^* seja finito. Note que, se o vetor de estimadores $\mathbf{T}(\cdot)$ contiver pelo menos uma de suas entradas cuja respectiva função de influência diverja, então γ^* será infinito. Estimadores que possuem γ^* finito são chamados de B-robustos.

Um dos principais estimadores no contexto de inferência robusta foi proposto por Huber (1964): o M-estimador. A ideia dele é generalizar o procedimento de estimação via máxima verossimilhança com o objetivo de desenvolver um método de estimação robusto.

Definição 6.3 (M-estimador.) Dizemos que o estimador $\mathbf{T}(F_n)$ é um M-estimador para $\boldsymbol{\theta}$ se

$$\mathbf{T}(F_n) = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^n \rho(y_i, \boldsymbol{\theta}),$$

em que $\rho: \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ é uma função diferenciável e \mathcal{X} denota o conjunto suporte.

Não é difícil verificar que o estimador de máxima verossimilhança é um M-estimador. Para tal, basta considerar $\rho(y_i, \boldsymbol{\theta}) = -\log(f_{\boldsymbol{\theta}}(y_i))$. A equação de estimação associada ao M-estimador é

$$\sum_{i=1}^n \psi(y_i, \mathbf{T}(F_n)) = \mathbf{0}, \quad (6.1)$$

em que $\psi(\cdot, \mathbf{T}(F_n)) = (\partial/\partial\boldsymbol{\theta})\rho(\cdot, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\mathbf{T}(F_n)}$. O estimador $\mathbf{T}(F_n)$ será Fisher-consistente se, e somente se, $\mathbb{E} \left[\sum_{i=1}^n \psi(Y_i, \boldsymbol{\theta}) \right] = \mathbf{0}, \forall \boldsymbol{\theta} \in \Theta$; isto é, se a função de estimação for não viesada.

A função de influência para o M-estimador é dada por

$$\text{IF}(y; \mathbf{T}, F_{\boldsymbol{\theta}}) = \mathbf{M}(\boldsymbol{\psi}, F_{\boldsymbol{\theta}})^{-1} \boldsymbol{\psi}(y, \mathbf{T}(\boldsymbol{\theta})),$$

sendo

$$\mathbf{M}(\boldsymbol{\psi}, F_{\boldsymbol{\theta}}) = - \int \frac{\partial}{\partial \boldsymbol{\theta}} [\boldsymbol{\psi}(y, \boldsymbol{\theta})] |_{\boldsymbol{\theta}=\mathbf{T}(F_n)} dF_{\boldsymbol{\theta}}(y).$$

Observe que se algum componente do vetor $\boldsymbol{\psi}(y, \boldsymbol{\theta})$ for ilimitado o estimador não será B-robusto. Adicionalmente, pode-se mostrar que a matriz de variâncias e covariâncias assintótica de um M-estimador $\mathbf{T}(\cdot)$ é dada por

$$\mathbf{V}(\mathbf{T}, F_{\boldsymbol{\theta}}) = \mathbf{M}(\boldsymbol{\psi}, F_{\boldsymbol{\theta}})^{-1} \mathbf{Q}(\boldsymbol{\psi}, F_{\boldsymbol{\theta}}) \mathbf{M}(\boldsymbol{\psi}, F_{\boldsymbol{\theta}})^{-1},$$

em que

$$\mathbf{Q}(\boldsymbol{\psi}, F_{\boldsymbol{\theta}}) = - \int \boldsymbol{\psi}(y, \mathbf{T}(\boldsymbol{\theta})) \boldsymbol{\psi}(y, \mathbf{T}(\boldsymbol{\theta}))^{\top} dF_{\boldsymbol{\theta}}(y).$$

Finalmente,

$$\sqrt{n}(\mathbf{T}(F_n) - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{V}(\mathbf{T}, F_{\boldsymbol{\theta}})), \quad \text{quando } n \rightarrow \infty.$$

Veja [Huber \(1964\)](#) para mais detalhes.

6.1.2 Regressão para dados binários na presença de contaminação

O modelo de regressão para dados binários é definido como segue. Sejam Y_1, \dots, Y_n variáveis aleatórias independentes com $Y_i \sim \text{Bernoulli}(\vartheta_i)$, em que

$$d_0(\vartheta_i) = \eta_{0i} = \mathbf{s}_i^{\top} \boldsymbol{\kappa},$$

sendo $\mathbf{s}_i = (s_{i1}, \dots, s_{ip_0})^{\top}$ o vetor de variáveis explicativas e $\boldsymbol{\kappa} \in \mathbb{R}^{p_0}$ o vetor de parâmetros desconhecidos. A função de probabilidade de Y_i é

$$f(y_i; \vartheta_i) = \vartheta_i^{y_i} (1 - \vartheta_i)^{1-y_i}, \quad y_i = 0, 1.$$

A estimativa de máxima verossimilhança de $\boldsymbol{\kappa}$ é

$$\hat{\boldsymbol{\kappa}}_{MV} = \arg \min_{\boldsymbol{\kappa} \in \mathbb{R}^{p_0}} \sum_{i=1}^n D(y_i, \eta_{0i}),$$

em que $D(y_i, \eta_{0i}) = -\log f(y_i; \vartheta_i) = -y_i \log \vartheta_i - (1 - y_i) \log(1 - \vartheta_i)$ é a função desvio.

A estimação via máxima verossimilhança no modelo de regressão para dados binários é fortemente influenciada por observações atípicas. Sem perda de generalidade, consideremos o modelo de regressão logística, isto é, $d_0(\vartheta_i) = \log(\vartheta_i/(1 - \vartheta_i))$. Neste caso, a equação de estimação para a obtenção do estimador de máxima verossimilhança é

$$\sum_{i=1}^n (y_i - \vartheta_i) \mathbf{s}_i = \mathbf{0}.$$

Dessa forma, a função de influência é dada por

$$\text{IF}(y_i; \hat{\boldsymbol{\kappa}}) = \mathbf{M}^{-1}(y_i - \vartheta_i) \mathbf{s}_i,$$

em que $\mathbf{M} = -\mathbb{E}[(\partial/\partial \boldsymbol{\kappa})(y_i - \vartheta_i) \mathbf{s}_i]$. Observe que $(y - \vartheta)$ é limitado. Assim, os únicos *outliers* que terão influência são aqueles que possuem $\|\mathbf{s}_i\|$ grande. Mas o que são *outliers* no contexto de regressão logística? Conforme discutido em [Copas \(1988\)](#), a ideia de *outlier*, em geral, está associada com uma interpretação geométrica, na qual o ponto está longe da massa dos dados, e uma probabilística, na qual, se o modelo ajustado for verdadeiro, um valor “ofensivo” de y seria

aquele que ocorre com probabilidade muito pequena. Em regressão logística, há pouco espaço para discrepâncias geométricas na direção de y . De fato, todos os y são 0 ou 1 e, assim, um erro na direção de y só pode ocorrer quando $0 \rightarrow 1$ ou $1 \rightarrow 0$. Portanto, nestes modelos, um *outlier* é definido apenas no sentido probabilístico. Dizemos que uma observação y é um outlier se $y = 1$ e ϑ é próximo de zero ou $y = 0$ e ϑ é próximo de um. É possível também haver *outlier* no espaço das covariáveis, chamados de pontos de alavanca. Um ponto de alavanca pode ser bom ou ruim. Um ponto de alavanca bom é aquele que é um *outlier* no espaço de \mathbf{s} , mas que o valor de $(y - \vartheta)$ é pequeno. Podemos dizer que esse tipo de *outlier* é inofensivo para o estimador de máxima verossimilhança. Já o ponto de alavanca ruim é aquele que é um *outlier* no espaço dos \mathbf{s} com $(y - \vartheta)$ grande. Esse tipo de ponto de alavanca produz valores grandes para a função de influência do estimador de máxima verossimilhança (para mais detalhes, ver [Croux e Haesbroeck \(2003\)](#)).

O efeito de *outliers* em regressão logística tem sido discutido em diversos trabalhos. Em particular, [Croux et al. \(2002\)](#) mostram que os estimadores de máxima verossimilhança, exceto o do intercepto, tendem para zero quando muitos *outliers* são adicionados aos dados. Nas últimas décadas, diversos trabalhos têm aparecido na literatura no sentido de desenvolver estimadores robustos para os parâmetros do modelo de regressão logística. [Pregibon \(1982\)](#) foi o pioneiro. O autor sugere um M-estimador baseado em uma função de perda do tipo Huber ([Huber, 1964](#)). No entanto, o estimador resultante não é Fisher-consistente e é assintoticamente viesado. [Carroll e Pederson \(1993\)](#) propuseram uma versão do estimador de [Pregibon \(1982\)](#) que é menos viesada em amostras de tamanho pequeno. A ideia é ponderar a equação de estimação utilizando uma matriz de alavanca. Uma versão Fisher-consistente e assintoticamente normal do estimador proposto por [Pregibon \(1982\)](#) foi apresentada por [Bianco e Yohai \(1996\)](#). No entanto, em alguns casos, o estimador proposto pode não existir. Recentemente, [Croux e Haesbroeck \(2003\)](#) propuseram uma modificação no estimador de [Bianco e Yohai \(1996\)](#) a fim de garantir sua existência, dadas algumas condições. Outros trabalhos são encontrados em [Cantoni e Ronchetti \(2001\)](#), [Bondell \(2005\)](#) e [Ghosh e Basu \(2016\)](#). Nesta tese, daremos atenção ao estimador proposto por [Ghosh e Basu \(2016\)](#), que é obtido através da minimização da versão empírica da divergência potência entre densidades introduzida por [Basu et al. \(1998\)](#). Este estimador será estudado em detalhes nas próximas seções.

6.1.3 Regressão beta na presença de contaminação

O modelo de regressão beta, na parametrização proposta por [Ferrari e Cribari-Neto \(2004\)](#), é definido a seguir. Considere Y_1, \dots, Y_n variáveis aleatórias independentes, com $Y_i \sim \text{beta}(\mu_i, \phi_i)$, para $i = 1, \dots, n$, isto é, a função densidade de probabilidade de Y_i é dada por

$$f_{\theta}(y_i; \mu_i, \phi_i) = \frac{1}{B(\mu_i \phi_i, (1 - \mu_i) \phi_i)} y_i^{\mu_i \phi_i - 1} (1 - y_i)^{(1 - \mu_i) \phi_i - 1}, \quad 0 < y_i < 1, \quad (6.2)$$

em que $B(\cdot, \cdot)$ representa a função beta e μ_i e ϕ_i representam a média e a precisão de Y_i , respectivamente, com

$$\begin{aligned} g_{\mu}(\mu_i) &= \mathbf{x}_i^{\top} \boldsymbol{\beta} = \eta_{1i}, \\ g_{\phi}(\phi_i) &= \mathbf{z}_i^{\top} \boldsymbol{\gamma} = \eta_{2i}, \end{aligned} \quad (6.3)$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p_1})^{\top} \in \mathbb{R}^{p_1}$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{p_2})^{\top} \in \mathbb{R}^{p_2}$ são os vetores de parâmetros desconhecidos, $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \boldsymbol{\gamma}^{\top})^{\top}$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip_1})^{\top}$ e $\mathbf{z}_i = (z_{i1}, \dots, z_{ip_2})^{\top}$ são observações das covariáveis com $p_1 + p_2 < n$, e η_{1i} e η_{2i} são os preditores lineares. Assumimos que as funções de ligação $g_{\mu} : (0, 1) \rightarrow \mathbb{R}$ e $g_{\phi} : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis.

Inferência sobre os parâmetros do modelo de regressão beta é feita, em geral, através do método de máxima verossimilhança. O logaritmo da função de verossimilhança para $\boldsymbol{\theta}$ é dado por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\mu_i, \phi_i), \quad (6.4)$$

em que

$$\ell_i(\mu_i, \phi_i) = \log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1 - \mu_i) \phi_i) + (\mu_i \phi_i - 1) y_i^* + (\phi_i - 2) y_i^\dagger,$$

sendo $y_i^* = \log\{y_i/(1 - y_i)\}$ e $y_i^\dagger = \log(1 - y_i)$. O estimador de máxima verossimilhança de θ é obtido maximizando (6.4). A equação de estimação associada a esse estimador é

$$\sum_{i=1}^n U(y_i, \theta) = \mathbf{0},$$

em que $U(y_i, \theta) = (\partial/\partial\theta) \log[f_\theta(y_i; \mu_i, \phi_i)]$ é o vetor escore de θ correspondente à i -ésima observação, dado por

$$U(y_i, \theta) = \left(\phi_i \frac{(y_i^* - \mu_i^*)}{g'_\mu(\mu_i)} \mathbf{x}_i^\top, \frac{\mu_i(y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)}{g'_\phi(\phi_i)} \mathbf{z}_i^\top \right)^\top,$$

em que $\mu_i^* = \mathbb{E}(Y_i^*) = \psi(\mu_i \phi_i) - \psi((1 - \mu_i) \phi_i)$ e $\mu_i^\dagger = \mathbb{E}(Y_i^\dagger) = \psi((1 - \mu_i) \phi_i) - \psi(\phi_i)$, e $\psi(\cdot)$ é a função digamma.

A função de influência do estimador de máxima verossimilhança sob o modelo de regressão beta é dada por

$$\text{IF}(y; \hat{\theta}) = -K^{-1}(\theta)U(y, \theta),$$

em que $K(\theta)$ é a matriz da informação de Fisher para uma única observação. Conforme observado em Ribeiro e Ferrari (2022), essa função de influência não é limitada. De fato, basta verificar que os limites dos componentes do vetor escore divergem quando y tende para os limites do suporte (isto é, $y \rightarrow 0$ ou $y \rightarrow 1$). Dessa forma, o estimador de máxima verossimilhança não é B-robusto. Muitos trabalhos citam a falta de robustez do estimador de máxima verossimilhança em regressão beta. A seguir, mencionamos três trabalhos na literatura que estudam inferência robusta neste modelo.

O primeiro deles é o trabalho de Ghosh (2019), em que é proposto um estimador robusto baseado na minimização da versão empírica da divergência potência entre densidades, o MDPDE. A segunda proposta é apresentada em Ribeiro e Ferrari (2022). As autoras definem um estimador robusto que é baseado na maximização de uma função de L_q -verossimilhança reparametrizada, denominado estimador de máxima verossimilhança substituto (*surrogate maximum likelihood estimator*; SMLE). Em geral, métodos robustos dependem de uma constante, denominada aqui constante de afinação (*tunning constant*). A escolha dessa constante é crucial nesses procedimentos pois representa o equilíbrio entre robustez e eficiência. Ribeiro e Ferrari (2022) desenvolveram um algoritmo para a escolha ótima dessa constante através dos dados. As autoras verificaram, através de extensos estudos de simulação e de aplicações em dados reais, que o procedimento proposto é eficaz para a escolha da constante de afinação tanto para o SMLE quanto para o MDPDE. Embora os estimadores propostos por Ghosh (2019) e Ribeiro e Ferrari (2022) mostrem-se razoáveis em diversos cenários, esses estimadores não são bem definidos para todos os valores dos parâmetros e da constante de afinação. De fato, Ribeiro e Ferrari (2022) verificam que uma condição suficiente para que o SMLE e o MDPDE estejam bem definidos é que a densidade beta seja limitada, isto é, $\mu\phi > 1$ e $(1 - \mu)\phi > 1$. Nesse sentido, Maluf *et al.* (2022) propuseram uma extensão do MDPDE e do SMLE, o LMDPDE e o LSMLE, que tem a vantagem de estar bem definido para todos os valores dos parâmetros e da constante de afinação. A ideia dos autores é desenvolver os estimadores considerando a distribuição do logito de Y , ao invés da distribuição beta diretamente. Estes estimadores serão discutidos em detalhe nas próximas seções.

6.2 Inferência robusta no modelo de regressão beta inflacionado

Nas seções que seguem, definimos o modelo de regressão beta inflacionado e propomos estimadores robustos para os parâmetros do modelo. A notação utilizada em cada um dos estimadores propostos pode coincidir e deverá ser diferenciada através do contexto.

6.2.1 O modelo de regressão beta inflacionado

A distribuição beta inflacionada em zero ou um é construída a partir da mistura entre uma variável aleatória com distribuição beta e uma variável degenerada em c , em que $c = 0$ ou $c = 1$, dependendo do caso. Nesse caso, o componente contínuo dos dados é modelado por uma distribuição beta e o componente discreto, isto é, o ponto de massa, será modelado por uma distribuição degenerada no valor conhecido c . A função de distribuição acumulada da mistura é dada por

$$\text{BI}_c(y; \vartheta, \mu, \phi) = \vartheta \mathbb{I}_{[c, \infty)}(y) + (1 - \vartheta)F(y; \mu, \phi),$$

sendo $\mathbb{I}_A(y)$ a função indicadora, com valor 1 se $y \in A$ e 0 se $y \notin A$ e a função $F(y; \mu, \phi)$ é a função de distribuição acumulada da distribuição beta (μ, ϕ) . Escrevemos $\mathcal{Y} \sim \text{BI}_c(\vartheta, \mu, \phi)$. A função densidade de probabilidade de \mathcal{Y} é dada por¹

$$\text{bi}_c(y; \vartheta, \mu, \phi) = \begin{cases} \vartheta, & y = c, \\ (1 - \vartheta)f(y; \mu, \phi), & y \in (0, 1), \end{cases} \quad (6.5)$$

em que $0 < \vartheta < 1$, $\mu \in (0, 1)$, $\sigma > 0$, $\lambda > 0$ e $f(y; \mu, \phi)$ é a função densidade de probabilidade de uma distribuição beta (μ, ϕ) . Note que $\vartheta = \mathbb{P}(\mathcal{Y} = c)$ representa a probabilidade de se observar zero ou um e μ e ϕ são os parâmetros da distribuição beta, representando a média e precisão da distribuição condicional de $\mathcal{Y}|\mathcal{Y} \in (0, 1)$. A média e a variância de \mathcal{Y} podem ser escritas como

$$\begin{aligned} \mathbb{E}(\mathcal{Y}) &= \vartheta c + (1 - \vartheta)\mu, \\ \text{Var}(\mathcal{Y}) &= (1 - \vartheta)\frac{\mu(1 - \mu)}{1 + \phi} + \vartheta(1 - \vartheta)(c - \mu)^2. \end{aligned}$$

Além disso, $\mathbb{E}(\mathcal{Y}|\mathcal{Y} \in (0, 1)) = \mu$ e $\text{Var}(\mathcal{Y}|\mathcal{Y} \in (0, 1)) = \mu(1 - \mu)/(1 + \phi)$. Para mais detalhes a respeito da distribuição beta inflacionada, sugerimos a leitura de [Ospina e Ferrari \(2010\)](#).

Sejam $\mathcal{Y}_1, \dots, \mathcal{Y}_n$ variáveis aleatórias independentes, com $\mathcal{Y}_i \sim \text{BI}_c(\vartheta_i, \mu_i, \phi_i)$, para $i = 1, \dots, n$. O modelo de regressão beta inflacionado no ponto c ($c = 0$ ou $c = 1$) tem como componentes sistemáticos

$$\begin{aligned} g_\vartheta(\vartheta_i) &= \mathbf{S}_i^\top \boldsymbol{\kappa} = \eta_{0i}, \\ g_\mu(\mu_i) &= \mathbf{X}_i^\top \boldsymbol{\beta} = \eta_{1i}, \\ g_\phi(\phi_i) &= \mathbf{Z}_i^\top \boldsymbol{\gamma} = \eta_{2i}, \end{aligned} \quad (6.6)$$

em que $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_{p_0})^\top \in \mathbb{R}^{p_0}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p_1})^\top \in \mathbb{R}^{p_1}$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{p_2})^\top \in \mathbb{R}^{p_2}$ são os vetores de parâmetros desconhecidos, $\mathbf{S}_i = (\mathcal{S}_{i1}, \dots, \mathcal{S}_{ip_0})^\top$, $\mathbf{X}_i = (\mathcal{X}_{i1}, \dots, \mathcal{X}_{ip_1})^\top$ e $\mathbf{Z}_i = (\mathcal{Z}_{i1}, \dots, \mathcal{Z}_{ip_2})^\top$ são observações das covariáveis com $p_0 + p_1 + p_2 < n$, e η_{0i} , η_{1i} e η_{2i} são os preditores lineares. Assumimos que as funções de ligação $g_\vartheta : (0, 1) \rightarrow \mathbb{R}$, $g_\mu : (0, 1) \rightarrow \mathbb{R}$ e $g_\phi : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. A escolha dessas funções pode variar, por exemplo, podem-se escolher as funções logito, probito e log-log para g_ϑ e g_μ , e a função logarítmica para g_ϕ . Os parâmetros do modelo de regressão beta inflacionado em zero ou um são interpretados em termos de ϑ_i , μ_i e ϕ_i , que representam, respectivamente, a probabilidade da variável \mathcal{Y}_i assumir o valor c (fixo), a média e a precisão de \mathcal{Y}_i condicional a $\mathcal{Y}_i \in (0, 1)$.

Seja $\boldsymbol{\Upsilon} = (\boldsymbol{\kappa}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ o vetor de parâmetros do modelo. A função de verossimilhança de $\boldsymbol{\Upsilon}$ para o modelo de regressão beta inflacionado em c é da forma

$$L(\boldsymbol{\Upsilon}) = \prod_{i=1}^n \text{bi}_c(y_i; \vartheta_i, \mu_i, \phi_i) = L_1(\boldsymbol{\kappa})L_2(\boldsymbol{\beta}, \boldsymbol{\gamma}),$$

¹A medida de probabilidade P correspondente a $\text{BI}^{(c)}(y; \cdot)$ definida sob o espaço mensurável $((0, 1) \cup \{c\}, \mathcal{B})$, sendo \mathcal{B} a classe de todos os subconjuntos de Borel de $(0, 1) \cup \{c\}$, é tal que $P \ll \zeta + \delta_c$, em que ζ é a medida de Lebesgue e δ_c o ponto de massa em c ; isto é, $\delta_c(A) = 1$, se $c \in A$ e $\delta_c(A) = 0$, se $c \notin A$, $A \in \mathcal{B}$.

em que

$$L_1(\boldsymbol{\kappa}) = \prod_{i=1}^n \vartheta_i^{\mathbb{I}_{\{c\}}(y_i)} (1 - \vartheta_i)^{1 - \mathbb{I}_{\{c\}}(y_i)},$$

$$L_2(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i: y_i \in (0,1)} f(y_i; \mu_i, \phi_i).$$

Observe que a função de verossimilhança $L(\boldsymbol{\Upsilon})$ é separada em dois termos, um que depende de $\boldsymbol{\kappa}$ e o outro que depende de $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Dessa forma, dizemos que esses vetores de parâmetros são separáveis (Pace e Salvan, 1997). Assim, a inferência por máxima verossimilhança pode ser realizada de forma individual para $\boldsymbol{\kappa}$ e $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Além disso, $L_1(\boldsymbol{\kappa})$ é a função de verossimilhança de um modelo linear generalizado com resposta binária (McCullagh e Nelder, 1989) e $L_2(\boldsymbol{\beta}, \boldsymbol{\gamma})$ é a função de verossimilhança de um modelo de regressão beta em que a variável resposta é restrita às observações no intervalo $(0, 1)$.

A função de log-verossimilhança para $\boldsymbol{\Upsilon}$ é dada por

$$\ell(\boldsymbol{\Upsilon}) = \ell_1(\boldsymbol{\kappa}) + \ell_2(\boldsymbol{\beta}, \boldsymbol{\gamma}), \quad (6.7)$$

em que

$$\ell_1(\boldsymbol{\kappa}) = \sum_{i=1}^n \ell_{1i}(\vartheta_i),$$

$$\ell_2(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i: y_i \in (0,1)} \ell_{2i}(\mu_i, \phi_i),$$

sendo

$$\begin{aligned} \ell_{1i}(\vartheta_i) &= \mathbb{I}_{\{c\}}(y_i) \log(\vartheta_i) + [1 - \mathbb{I}_{\{c\}}(y_i)] \log(1 - \vartheta_i), \\ \ell_{2i}(\mu_i, \phi_i) &= \log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1 - \mu_i) \phi_i) + (\mu_i \phi_i - 1) \log(y_i / (1 - y_i)) \\ &\quad + (\phi_i - 2) \log(1 - y_i). \end{aligned}$$

Adicionalmente, observe que, para $i = 1, \dots, n$, a variável aleatória $\mathbb{I}_{\{c\}}(\mathcal{Y}_i)$ é do tipo Bernoulli com probabilidade de sucesso $\mathbb{P}(\mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 1) = \vartheta_i$. O estimador de máxima verossimilhança (MLE) de $\boldsymbol{\Upsilon}$ é obtido através da maximização de (6.7).

Considere o conjunto aleatório

$$\wp = \wp(\mathbb{I}_{\{c\}}(\mathcal{Y}_1), \dots, \mathbb{I}_{\{c\}}(\mathcal{Y}_n)) = \{i \in \{1, \dots, n\} : \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0\}, \quad (6.8)$$

de tamanho n^\dagger , que representa o conjunto dos índices das observações em $(0, 1)$. No que segue, a depender do contexto, \wp denotará o conjunto aleatório $\wp(\mathbb{I}_{\{c\}}(\mathcal{Y}_1), \dots, \mathbb{I}_{\{c\}}(\mathcal{Y}_n))$ ou o conjunto observado $\wp(\mathbb{I}_{\{c\}}(y_1), \dots, \mathbb{I}_{\{c\}}(y_n))$.

Sejam

$$\mathcal{Y}_i^* = \begin{cases} \log\left(\frac{\mathcal{Y}_i}{1 - \mathcal{Y}_i}\right), & i \in \wp, \\ 0, & i \notin \wp, \end{cases} \quad \mathcal{Y}_i^\dagger = \begin{cases} \log(1 - \mathcal{Y}_i), & i \in \wp, \\ 0, & i \notin \wp, \end{cases}$$

$$\mu_i^* = \mathbb{E}(\mathcal{Y}_i^* | \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0) = \psi(\mu_i \phi_i) - \psi((1 - \mu_i) \phi_i)$$

e

$$\mu_i^\dagger = \mathbb{E}(\mathcal{Y}_i^\dagger | \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0) = \psi((1 - \mu_i) \phi_i) - \psi(\phi_i).$$

Os valores observados de \mathcal{Y}_i^* e \mathcal{Y}_i^\dagger são denotados por y_i^* e y_i^\dagger , respectivamente. Dessa forma, o sistema de equações de estimação associado ao estimador de máxima verossimilhança pode ser

escrito como

$$\begin{aligned} \sum_{i=1}^n \mathbf{U}(y_i; \boldsymbol{\kappa}) &= \mathbf{0}, \\ \sum_{i \in \wp} \mathbf{U}(y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \mathbf{0}, \end{aligned} \quad (6.9)$$

em que

$$\mathbf{U}(y_i; \boldsymbol{\kappa}) = \frac{\mathbb{I}_{\{c\}}(y_i) - \vartheta_i}{\vartheta_i(1 - \vartheta_i)} \frac{1}{g'_{\vartheta}(\vartheta_i)} \boldsymbol{\mathcal{S}}_i, \quad i = 1, \dots, n$$

e

$$\mathbf{U}(y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}) = \left(\phi_i \frac{(y_i^* - \mu_i^*)}{g'_{\mu}(\mu_i)} \boldsymbol{\mathcal{X}}_i^{\top}, \frac{\mu_i(y_i^* - \mu_i^*) + (y_i^{\dagger} - \mu_i^{\dagger})}{g'_{\phi}(\phi_i)} \boldsymbol{\mathcal{Z}}_i^{\top} \right)^{\top}, \quad i \in \wp.$$

O sistema de equações de estimação (6.9) pode ser resolvido de forma separada, sendo uma equação para a estimação dos parâmetros associados à parte discreta e a outra à contínua. A equação relativa à parte discreta equivale à de um modelo de regressão para dados binários, em que a variável dependente é $\mathbb{I}_{\{c\}}(\mathcal{Y}_i) \sim \text{Bernoulli}(\vartheta_i)$ e as covariáveis $\boldsymbol{\mathcal{S}}_i$ são relacionadas com o vetor de parâmetros $\boldsymbol{\kappa}$ através da função de ligação $g_{\vartheta}(\vartheta_i)$, conforme (6.6). Já a equação referente à parte contínua equivale à de uma regressão beta, em que a variável resposta é $\mathcal{Y}_i \sim \text{beta}(\mu_i, \phi_i)$, para $i \in \wp$, e as covariáveis $\boldsymbol{\mathcal{X}}_i$ e $\boldsymbol{\mathcal{Z}}_i$ são relacionadas com os vetores de parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ através das funções de ligação $g_{\mu}(\mu_i)$ e $g_{\phi}(\phi_i)$, respectivamente (ver (6.6)). Dessa forma, na presença de observações atípicas, a inferência via máxima verossimilhança nos modelos de regressão beta inflacionados herda a falta de robustez observada nos modelos de regressão para dados binários e em regressão beta (ver Seções 6.1.2 e 6.1.3, respectivamente)

Neste capítulo, desenvolvemos novos estimadores robustos para os parâmetros do modelo de regressão beta inflacionado. Nossa proposta é baseada nos trabalhos de Ghosh e Basu (2016), Ghosh (2019), Ribeiro e Ferrari (2022) e Maluf *et al.* (2022). A ideia é tratar o processo de estimação de forma separada e considerar os estimadores robustos para a parte discreta e contínua. Para a parte discreta, utilizaremos o MDPDE e, para a parte contínua, o SMLE, MDPDE, LSMLE e LMDPDE. Feito isso, unificaremos as duas abordagens, obtendo quatro classes de estimadores para os parâmetros do modelo de regressão beta inflacionado em zero ou um.

6.2.2 Estimação via MDPD para κ

Ghosh e Basu (2016) desenvolveram um estimador robusto, chamado de estimador de mínima divergência potência entre densidades (*minimum density power divergence estimator*; MDPDE) no contexto de modelos lineares generalizados. O MDPDE é obtido através da minimização da versão empírica da divergência potência entre duas densidades, que foi desenvolvida por Basu *et al.* (1998) em termos de uma constante de afinação (*tuning constant*) $\alpha \geq 0$. Essa classe de estimadores será definida a seguir para os modelos de regressão para dados binários.

A família de divergências potência de f_{δ} com respeito a g é definida por

$$d_{\alpha}(f_{\delta}, g) = \int f_{\delta}(y)^{1+\alpha} dy - \frac{1+\alpha}{\alpha} \int f_{\delta}(y)^{\alpha} g(y) dy + \frac{1}{\alpha} \int g(y)^{1+\alpha} dy,$$

para $\alpha > 0$ e

$$d_0(f_{\delta}, g) = \lim_{\alpha \rightarrow 0} d_{\alpha}(f_{\delta}, g) = \int \log \left(\frac{g(y)}{f_{\delta}(y)} \right) g(y) dy.$$

Na prática, g representa a densidade dos dados e f_{δ} a densidade do modelo (que depende de parâmetros desconhecidos). As densidades podem ser substituídas por funções de probabilidade, no contexto de variáveis aleatórias discretas; neste caso, as integrais são substituídas por somatórios. Finalmente, o MDPDE é obtido minimizando a versão empírica de $d_{\alpha}(f_{\delta}, g)$ que, a menos de um

termo que não depende de $\boldsymbol{\delta}$, é dada por

$$\mathcal{H}_n(\boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^n \mathcal{V}_i(Y_i; \boldsymbol{\delta}),$$

em que

$$\mathcal{V}_i(y_i; \boldsymbol{\delta}) = \int f_{\boldsymbol{\delta}}(y)^{1+\alpha} dy - \frac{1+\alpha}{\alpha} f_{\boldsymbol{\delta}}(y_i)^\alpha,$$

desde que a integral convirja. Observe que se $\alpha = 0$, a divergência potência coincide com a de Kullback-Leibler e o estimador MDPDE é o de máxima verossimilhança. Adicionalmente, para $0 < \alpha < 1$ a família de divergências potência fornece um caminho suave entre a divergência de Kullback-Leibler e a distância L_2 entre densidades (Basu *et al.*, 1998). Neste trabalho, consideramos $0 \leq \alpha < 1$.

Sejam $(\mathbb{I}_{\{c\}}(\mathcal{Y}_i), \boldsymbol{S}_i)$, para $i = 1, \dots, n$, dados provenientes de um modelo de regressão para dados binários. Defina $Y_i^c = \mathbb{I}_{\{c\}}(\mathcal{Y}_i)$. Dessa forma, assumimos que $Y_i^c \sim \text{Bernoulli}(\vartheta_i)$, em que $g_{\vartheta}(\vartheta_i) = \boldsymbol{S}_i^\top \boldsymbol{\kappa} = \eta_{0i}$ e Y_i^c são variáveis aleatórias independentes para $i = 1, \dots, n$. Desta forma, o MDPDE para $\boldsymbol{\kappa}$, denotado aqui por $\hat{\boldsymbol{\kappa}}_\alpha$ é obtido minimizando

$$\mathcal{H}_n(\boldsymbol{\kappa}) = \frac{1}{n} \sum_{i=1}^n \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}),$$

em que

$$\mathcal{V}_i(y_i^c; \boldsymbol{\kappa}) = \sum_{y^c=0}^1 f_{\boldsymbol{\kappa}}(y^c; \vartheta_i)^{1+\alpha} - \frac{1+\alpha}{\alpha} f_{\boldsymbol{\kappa}}(y_i^c; \vartheta_i)^\alpha,$$

com $f_{\boldsymbol{\kappa}}(y^c; \vartheta_i) = \vartheta_i^{y^c} (1 - \vartheta_i)^{1-y^c}$ sendo a função de probabilidade de Y^c . Assim, temos que

$$\mathcal{V}_i(y_i^c; \boldsymbol{\kappa}) = \vartheta_i^{1+\alpha} + (1 - \vartheta_i)^{1+\alpha} - \frac{1+\alpha}{\alpha} f_{\boldsymbol{\kappa}}(y_i^c; \vartheta_i)^\alpha.$$

Para obtermos a equação de estimação associada a $\hat{\boldsymbol{\kappa}}_\alpha$, devemos observar que

$$\frac{\partial}{\partial \vartheta_i} \mathcal{V}_i(y_i^c; \boldsymbol{\kappa}) = \frac{(1+\alpha)}{\vartheta_i(1-\vartheta_i)} \left[(1-\vartheta_i)\vartheta_i^{1+\alpha} - \vartheta_i(1-\vartheta_i)^{1+\alpha} - (y_i^c - \vartheta_i)f_{\boldsymbol{\kappa}}(y_i^c; \vartheta_i)^\alpha \right].$$

Assim, a equação de estimação fica dada por

$$\sum_{i=1}^n \mathbf{U}^*(y_i^c; \boldsymbol{\kappa}) = \mathbf{0}, \quad (6.10)$$

em que

$$\mathbf{U}^*(y_i^c; \boldsymbol{\kappa}) = \frac{(1+\alpha)}{\vartheta_i(1-\vartheta_i)} \left[(1-\vartheta_i)\vartheta_i^{1+\alpha} - \vartheta_i(1-\vartheta_i)^{1+\alpha} - (y_i^c - \vartheta_i)f_{\boldsymbol{\kappa}}(y_i^c; \vartheta_i)^\alpha \right] \frac{1}{g'_{\vartheta}(\vartheta_i)} \boldsymbol{S}_i.$$

Note que

$$\mathbb{E}[(Y_i^c - \vartheta_i)f_{\boldsymbol{\kappa}}(Y_i^c; \vartheta_i)^\alpha] = \sum_{y^c=0}^1 (y_i^c - \vartheta_i)f_{\boldsymbol{\kappa}}(y_i^c; \vartheta_i)^{1+\alpha} = \vartheta_i(1-\vartheta_i)^{1+\alpha} - (1-\vartheta_i)\vartheta_i^{1+\alpha}, \quad \forall \boldsymbol{\kappa} \in \mathbb{R}^{p_0},$$

o que implica que $\mathbb{E}[\mathbf{U}^*(Y_i^c; \boldsymbol{\kappa})] = \mathbf{0}$, $\forall \boldsymbol{\kappa} \in \mathbb{R}^{p_0}$. Ou seja, a função de estimação em (6.10) é não viesada. Este resultado será utilizado para mostrar a Fisher-consistência dos estimadores propostos para a regressão beta inflacionada.

Note que pontos de alavanca ruins, que têm grande influência na estimação via máxima verossimilhança, têm pouca influência para o MDPDE. De fato, se α não for muito próximo de zero, observações com valores grandes de $(y^c - \vartheta)$ recebem peso pequeno, dado por $f_{\boldsymbol{\kappa}}(y^c; \vartheta)^\alpha$. Dessa

forma, o peso inibe o efeito do ponto de alavanca ruim, dado pelo valor da covariável. A constante de afinação α faz o balanço entre eficiência ($\alpha = 0$ leva à estimação via máxima verossimilhança) e robustez.

6.2.3 Estimação via maximização da L_q -verossimilhança reparametrizada para (β, γ)

No que segue, trataremos de estimação robusta sobre os parâmetros associados à parte contínua do modelo de regressão beta inflacionado considerando a maximização de uma L_q -verossimilhança reparametrizada.

Por enquanto, assumimos que o conjunto \wp é fixado, isto é, não aleatório. Para a amostra $y_i, i \in \wp$, observada do modelo de regressão beta (6.2) e (6.3), a função de L_q -verossimilhança (Ferrari e Yang, 2010) de θ é dada por

$$\ell_q(\theta) = \sum_{i \in \wp} L_q(f_{\theta}(y_i; \mu_i, \phi_i)),$$

em que

$$L_q(u) = \begin{cases} (u^{1-q} - 1)/(1 - q), & q \neq 1, \\ \log(u), & q = 1, \end{cases}$$

$0 < q \leq 1$ é uma constante, que chamaremos de constante de afinação. A fim de mantermos a notação utilizada nas seções anteriores, definimos $\alpha = 1 - q$. Neste caso, temos $0 \leq \alpha < 1$. Note que $L_{1-\alpha}(u)$ é a transformação Box-Cox (Box e Cox, 1964), também chamada de logaritmo distorcido. O estimador de máxima L_q -verossimilhança (ML_qE) é obtido maximizando $\ell_{1-\alpha}(\theta)$. Se $\alpha = 0$, temos o estimador de máxima verossimilhança usual. A equação de estimação associada ao ML_qE é

$$\sum_{i \in \wp} \mathbf{U}(y_i; \theta) f_{\theta}(y_i; \mu_i, \phi_i)^{\alpha} = \mathbf{0}, \quad (6.11)$$

em que $\mathbf{U}(y_i; \theta)$ foi definido em (6.9). Para $\alpha \neq 0$, a função de estimação é viesada, de modo que o ML_qE não é Fisher-consistente. Ribeiro e Ferrari (2022) se baseiam nas ideias de Ferrari e La Vecchia (2012) para a construção de um estimador Fisher-consistente usando uma função L_q -verossimilhança reparametrizada. Esse procedimento é descrito a seguir; para detalhes adicionais, ver Ribeiro e Ferrari (2022). Ferrari e La Vecchia (2012) verificaram que é possível utilizar uma função de calibração para reescalonar as estimativas de θ de forma a garantir Fisher-consistência, desde que a família de distribuições seja fechada sob a transformação potência.

Definição 6.4 (Transformação potência.) Dada uma densidade v e uma constante $\omega > 0$, a transformação potência é definida como

$$v^{(\omega)}(y) = \frac{v(y)^{\omega}}{\int v(y)^{\omega} dy} \propto v(y)^{\omega}, \forall y \text{ no suporte},$$

desde que $\int v(y)^{\omega} dy < \infty$, para todo $\omega > 0$.

Para a família de densidades $\{v_{\theta}, \theta \in \Theta\}$, fechada sob a transformação potência para todo $\omega > 0$, seja $\tau_{\omega}(\theta) : \Theta \mapsto \Theta$ uma função contínua tal que

$$v_{\tau_{\omega}(\theta)}(y) = v_{\theta}^{(\omega)}(y), \quad (6.12)$$

para todo y no suporte (que não depende de θ). Isto é, aplicando a transformação potência em v_{θ} , obtemos uma densidade $v_{\theta}^{(\omega)}(\cdot)$ pertencente à mesma família de distribuições, mas com parametrização em termos de $\tau_{\omega}(\theta)$. Ribeiro e Ferrari (2022) mostraram que a densidade beta (6.2) é

fechada sob transformação potência se $\mu_i \phi_i > 1$ e $(1 - \mu_i) \phi_i > 1$, isto é, se a densidade $f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)$ é limitada. Finalmente, [Ferrari e La Vecchia \(2012\)](#) mostraram que o estimador $\hat{\boldsymbol{\theta}}_{\alpha}$, obtido através da maximização da função L_q -verossimilhança na parametrização $\tau_{1/(1-\alpha)}(\boldsymbol{\theta})$, é Fisher-consistente para $\boldsymbol{\theta}$.

No modelo de regressão beta, $\hat{\boldsymbol{\theta}}_{\alpha}$ é obtido maximizando

$$\ell_{1-\alpha}^*(\boldsymbol{\theta}) = \sum_{i \in \wp} L_{1-\alpha} \left(f_{\tau_{1/(1-\alpha)}(\boldsymbol{\theta})}(y_i; \mu_i, \phi_i) \right) = \sum_{i \in \wp} L_{1-\alpha} \left(f_{\boldsymbol{\theta}}^{(\frac{1}{1-\alpha})}(y_i; \mu_i, \phi_i) \right),$$

em que $f_{\boldsymbol{\theta}}^{(\frac{1}{1-\alpha})}(y_i; \mu_i, \phi_i) = f_{\boldsymbol{\theta}}(y_i; \mu_{i,(1-\alpha)^{-1}}, \phi_{i,(1-\alpha)^{-1}})$, com $0 \leq \alpha < 1$ e

$$\mu_{i,1-\alpha} = \phi_{i,1-\alpha}^{-1}[(1-\alpha)(\mu_i \phi_i - 1) + 1], \quad \phi_{i,1-\alpha} = (1-\alpha)(\phi_i - 2) + 2,$$

desde que $0 < \mu_{i,(1-\alpha)^{-1}} < 1$ e $\phi_{i,(1-\alpha)^{-1}} > 0$. Se a densidade $f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)$ é limitada essas desigualdades são satisfeitas para todo $0 \leq \alpha < 1$.

[Ribeiro e Ferrari \(2022\)](#) observam que $f_{\boldsymbol{\theta}}(y_i; \mu_{i,(1-\alpha)^{-1}}, \phi_{i,(1-\alpha)^{-1}})$ pode ser vista como a função densidade associada a um modelo de regressão beta modificado. Este modelo é definido através da densidade beta $f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)$, com submodelos da média e da precisão dados, respectivamente, por

$$g_{\mu}^*(\mu_i) = g_{\mu}(\mu_{i,1-\alpha}) = \boldsymbol{\mathcal{X}}_i^{\top} \boldsymbol{\beta}, \quad g_{\phi}^*(\phi_i) = g_{\phi}(\phi_{i,1-\alpha}) = \boldsymbol{\mathcal{Z}}_i^{\top} \boldsymbol{\gamma},$$

que será denotada por $f_{\boldsymbol{\theta}}^*(y_i; \mu_i, \phi_i)$. Dessa forma, o estimador $\hat{\boldsymbol{\theta}}_{\alpha}$ pode ser obtido através da maximização de

$$\ell_{1-\alpha}^*(\boldsymbol{\theta}) = \sum_{i \in \wp} L_{1-\alpha} (f_{\boldsymbol{\theta}}^*(y_i; \mu_i, \phi_i)),$$

cujas equações de estimação é dada por

$$\sum_{i \in \wp} \mathbf{U}^*(y_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(y_i; \mu_i, \phi_i)^{\alpha} = \mathbf{0}, \quad (6.13)$$

e $\mathbf{U}^*(y_i; \boldsymbol{\theta}) = (\partial/\partial \boldsymbol{\theta}) \log f_{\boldsymbol{\theta}}^*(y_i; \mu_i, \phi_i)$ é o vetor escore modificado para a i -ésima observação, $i \in \wp$, que é dado por

$$\mathbf{U}^*(y_i; \boldsymbol{\theta}) = \left((1-\alpha)^{-1} \phi_{i,1-\alpha} \frac{(y_i^* - \mu_i^*)}{g'_{\mu}(\mu_{i,1-\alpha})} \boldsymbol{\mathcal{X}}_i^{\top}, \quad (1-\alpha)^{-1} \frac{\mu_{i,1-\alpha} (y_i^* - \mu_i^*) + (y_i^{\dagger} - \mu_i^{\dagger})}{g'_{\phi}(\phi_{i,1-\alpha})} \boldsymbol{\mathcal{Z}}_i^{\top} \right)^{\top}.$$

O estimador final é chamado de estimador de máxima verossimilhança substituto (*surrogate maximum likelihood estimator*; SMLE). Note que o fator $f_{\boldsymbol{\theta}}^*(y_i; \mu_i, \phi_i)^{\alpha}$ na equação de estimação em (6.13) funciona como um peso no processo de estimação. Se $\alpha = 0$, todas as observações têm o mesmo peso (estimação por máxima verossimilhança). Por outro lado, para $\alpha > 0$, observações inconsistentes com o modelo postulado, isto é, que têm baixa densidade, recebem peso menor do aquelas que têm alta densidade.

O método de estimação proposto [Ribeiro e Ferrari \(2022\)](#) resulta em estimadores que apresentam propriedades interessantes e são aplicáveis em uma vasta gama de situações. No entanto, limita-se a densidades beta limitadas. [Maluf et al. \(2022\)](#) propuseram um estimador alternativo, que está bem definido e preserva propriedades de robustez mesmo para densidades beta que tendem a infinito em um ou ambos os extremos do intervalo unitário. A ideia principal na proposta dos autores é utilizar o método de [Ribeiro e Ferrari \(2022\)](#), mas considerando a distribuição do logito da variável resposta. Essa abordagem será descrita a seguir.

Definição 6.5 (Distribuição EGB.) *Seja $Y^* = \log(Y/(1-Y))$, em que $Y \sim \text{beta}(\mu, \phi)$. A*

função densidade de Y^* é dada por

$$h(y; \mu, \phi) = \frac{1}{B(\mu\phi, (1-\mu)\phi)} \frac{e^{-y(1-\mu)\phi}}{(1+e^{-y})^\phi}, \quad y \in \mathbb{R}, \quad (6.14)$$

com $0 < \mu < 1$ e $\phi > 0$. A distribuição de Y^* é chamada de beta exponencial generalizada de segundo tipo e escrevemos $Y^* \sim EGB(\mu, \phi)$.

Observe que $h(y; \mu, \phi)^\omega$ é proporcional ao núcleo de uma densidade $EGB(\mu, \omega\phi)$, para todo $\omega > 0$. Logo, a classe de distribuições EGB é fechada sob transformação potência.

O estimador proposto por Maluf *et al.* (2022) para θ , denominado LSMLE (*logit surrogate maximum likelihood estimator*), é obtido através da maximização de

$$\ell_{1-\alpha}^*(\theta) = \sum_{i \in \wp} L_{1-\alpha} \left(h_{\theta}^{\left(\frac{1}{1-\alpha}\right)}(y_i^*; \mu_i, \phi_i) \right),$$

em que $h_{\theta}^{\left(\frac{1}{1-\alpha}\right)}(y_i^*; \mu_i, \phi_i) = h_{\theta} \left(y_i^*; \mu_i, \phi'_{i,(1-\alpha)^{-1}} \right)$, com $0 \leq \alpha < 1$ e $\phi'_{i,(1-\alpha)^{-1}} = \phi_i(1-\alpha)$, $y_i^* = \log[y_i/(1-y_i)]$, para $i \in \wp$, e $h_{\theta}(y_i^*; \mu_i, \phi_i) = h(y_i^*; \mu_i, \phi_i)$ dado em (6.14) com μ_i e ϕ_i satisfazendo (6.3). Note que não é necessário impor nenhuma restrição a respeito dos valores de μ_i e ϕ_i . Note ainda que $h_{\theta} \left(y_i^*; \mu_i, \phi'_{i,(1-\alpha)^{-1}} \right)$ pode ser vista como a função densidade associada a um modelo de regressão modificado, definido através da densidade $h_{\theta}^*(y_i^*; \mu_i, \phi_i)$, com

$$g_{\mu}^*(\mu_i) = g_{\mu}(\mu_i) = \mathbf{X}_i^{\top} \boldsymbol{\beta}, \quad g_{\phi}^*(\phi_i) = g_{\phi}(\phi'_{i,(1-\alpha)^{-1}}) = \mathbf{Z}_i^{\top} \boldsymbol{\gamma},$$

que será denotada por $h_{\theta}^*(y_i^*; \mu_i, \phi_i)$. Assim, o LSMLE pode ser obtido através da maximização de

$$\ell_{1-\alpha}^*(\theta) = \sum_{i \in \wp} L_{1-\alpha} \left(h_{\theta}^*(y_i^*; \mu_i, \phi_i) \right),$$

cujas equações de estimação é dada por

$$\sum_{i \in \wp} \mathbf{U}^*(y_i^*; \theta) h_{\theta}^*(y_i^*; \mu_i, \phi_i)^{\alpha} = \mathbf{0}, \quad (6.15)$$

e $\mathbf{U}^*(y_i^*; \theta) = (\partial/\partial\theta) \log h_{\theta}^*(y_i^*; \mu_i, \phi_i)$ é o vetor escore modificado para a i -ésima observação, $i \in \wp$, que é dado por

$$\mathbf{U}^*(y_i^*; \theta) = \left(\phi_i \frac{(y_i^* - \mu_i^*)}{g'_{\mu}(\mu_i)} \mathbf{X}_i^{\top}, \quad (1-\alpha)^{-1} \frac{\mu_i(y_i^* - \mu_i^*) + (y_i^{\dagger} - \mu_i^{\dagger})}{g'_{\phi}(\phi'_{i,(1-\alpha)^{-1}})} \mathbf{Z}_i^{\top} \right)^{\top}.$$

6.2.4 Estimação via MDPD para (β, γ)

Ghosh (2019) propôs o MDPDE para os modelos de regressão beta. Esse estimador, assim como no caso de regressão para dados binários, tem a vantagem de ser Fisher-consistente por construção (uma vez que a função de estimação é não viesada). O MDPDE para θ é obtido minimizando

$$\mathcal{H}_n(\theta) = \frac{1}{n} \sum_{i \in \wp} \mathcal{V}_i(y_i; \theta),$$

em que

$$\mathcal{V}_i(y_i; \theta) = \mathcal{K}_{i,1+\alpha}(\theta) - \frac{1+\alpha}{\alpha} f_{\theta}(y_i; \mu_i, \phi_i)^{\alpha},$$

e

$$\mathcal{K}_{i,1+\alpha}(\theta) = \int_0^1 f_{\theta}(y; \mu_i, \phi_i)^{1+\alpha} dy = \frac{B((1+\alpha)(\mu_i\phi_i - 1) + 1, (1+\alpha)[(1-\mu_i)\phi_i - 1] + 1)}{B(\mu_i\phi_i, (1-\mu_i)\phi_i)^{1+\alpha}},$$

para $0 \leq \alpha < 1$, desde que a integral seja finita.

Ribeiro e Ferrari (2022) observaram que a minimização da divergência potência empírica $\mathcal{H}_n(\boldsymbol{\theta})$ corresponde à maximização de uma versão centralizada da função de L_q -verossimilhança $\ell_{1-\alpha}(\boldsymbol{\theta})$, dada por

$$\ell_{1-\alpha}^{**}(\boldsymbol{\theta}) = \sum_{i \in \varphi} [(1 + \alpha)L_{1-\alpha}(f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)) - \mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta})],$$

cujas equação de estimação é dada por

$$\sum_{i \in \varphi} [\mathbf{U}(y_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)^\alpha - \mathcal{E}_{i,1-\alpha}(\boldsymbol{\theta})] = \mathbf{0}, \quad (6.16)$$

em que $\mathcal{E}_{i,1-\alpha}(\boldsymbol{\theta}) = \mathbb{E} [\mathbf{U}(\mathcal{Y}_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\mathcal{Y}_i; \mu_i, \phi_i)^\alpha | \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0] = (\gamma_{1,i}^{(1+\alpha)} \boldsymbol{\mathcal{X}}_i^\top, \gamma_{2,i}^{(1+\alpha)} \boldsymbol{\mathcal{Z}}_i^\top)^\top$, sendo

$$\gamma_{1,i}^{(1+\alpha)} = \frac{\phi_i \mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta})}{g'_\mu(\mu_i)} (\mu_{i,1+\alpha}^* - \mu_i^*) \quad \text{e} \quad \gamma_{2,i}^{(1+\alpha)} = \frac{\mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta})}{g'_\phi(\phi_i)} [\mu_i (\mu_{i,1+\alpha}^* - \mu_i^*) + (\mu_{i,1+\alpha}^\dagger - \mu_i^\dagger)],$$

com $\mu_{i,1+\alpha}^* = \psi(\mu_i, 1+\alpha, \phi_i, 1+\alpha) - \psi((1 - \mu_i, 1+\alpha)\phi_i, 1+\alpha)$ e $\mu_{i,1+\alpha}^\dagger = \psi((1 - \mu_i, 1+\alpha)\phi_i, 1+\alpha) - \psi(\phi_i, 1+\alpha)$.

As quantidades $\mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta})$ e $\mathcal{E}_{i,1-\alpha}(\boldsymbol{\theta})$ estão bem definidas para $\mu_i \phi_i > \alpha/(1 + \alpha)$ e $(1 - \mu_i)\phi_i > \alpha/(1 + \alpha)$. Logo, assim como no caso do SMLE, as quantidades estão bem definidas para todo $0 \leq \alpha < 1$ para as densidades beta limitadas.

Finalmente, os estimadores SMLE e MDPDE são equivalentes e apresentam resultados similares. Conforme mencionado, assim como o SMLE, o MDPDE tem a desvantagem de não estar bem definido para todas as densidades beta. Dessa forma, Maluf *et al.* (2022) propuseram um método alternativo para obter o MDPDE para $\boldsymbol{\theta}$. De maneira análoga ao que foi feito para o caso do SMLE, os autores propõem calcular o estimador utilizando como densidade alvo a EGB ao invés da densidade beta. O estimador obtido será chamado de LMDPDE.

O LMDPDE é obtido através da minimização de

$$\mathcal{H}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i \in \varphi} \mathcal{V}_i(y_i^*; \boldsymbol{\theta}), \quad (6.17)$$

em que

$$\mathcal{V}_i(y_i^*; \boldsymbol{\theta}) = \mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta}) - \frac{1 + \alpha}{\alpha} h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^\alpha,$$

e

$$\mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} h_{\boldsymbol{\theta}}(y^*; \mu_i, \phi_i)^{1+\alpha} dy^* = \frac{B(\mu_i \phi_i (1 + \alpha), (1 - \mu_i)\phi_i (1 + \alpha))}{B(\mu_i \phi_i, (1 - \mu_i)\phi_i)^{1+\alpha}},$$

para $0 \leq \alpha < 1$. Note que a integral é finita para todo $0 \leq \alpha < 1$.

A equação de estimação fica dada por

$$\sum_{i \in \varphi} [\mathbf{U}(y_i; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^\alpha - \mathcal{E}'_{i,1-\alpha}(\boldsymbol{\theta})] = \mathbf{0}, \quad (6.18)$$

em que $\mathcal{E}'_{i,1-\alpha}(\boldsymbol{\theta}) = \mathbb{E} [\mathbf{U}(\mathcal{Y}_i; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha | \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0] = (\Lambda_{1,i}^{(1+\alpha)} \boldsymbol{\mathcal{X}}_i^\top, \Lambda_{2,i}^{(1+\alpha)} \boldsymbol{\mathcal{Z}}_i^\top)^\top$, sendo

$$\Lambda_{1,i}^{(\alpha)} = \frac{\phi_i \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{g'_\mu(\mu_i)} (\mu_{i,\alpha}^* - \mu_i^*), \quad \text{e} \quad \Lambda_{2,i}^{(\alpha)} = \frac{\mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{g'_\phi(\phi_i)} [\mu_i (\mu_{i,\alpha}^* - \mu_i^*) + (\mu_{i,\alpha}^\dagger - \mu_i^\dagger)],$$

em que $\mu_{i,\alpha}^* = \psi(\mu_i \phi'_{i,\alpha}) - \psi((1 - \mu_i)\phi'_{i,\alpha})$ e $\mu_{i,\alpha}^\dagger = \psi((1 - \mu_i)\phi'_{i,\alpha}) - \psi(\phi'_{i,\alpha})$.

Adicionalmente, também é possível encontrar uma correspondência entre a minimização de (6.17) e a maximização da versão empírica da função de L_q -verossimilhança centralizada.

6.2.5 Estimadores robustos para o modelo de regressão beta inflacionado em zero ou um e suas propriedades

Baseando-se no desenvolvimento feito nas Seções 6.2.2 - 6.2.4, propomos estimadores robustos para os parâmetros do modelo de regressão beta inflacionado em zero ou um. Esses estimadores serão chamados de M-SE, M-LSE, M-ME e M-LME e são obtidos através da solução da equação de estimação

$$\mathbf{U}(\mathbf{y}; \mathbf{\Upsilon}) = \mathbf{0}, \quad (6.19)$$

em que $\mathbf{U}(\mathbf{y}; \mathbf{\Upsilon}) = (\mathbf{U}(\boldsymbol{\kappa})^\top, \mathbf{U}(\boldsymbol{\theta})^\top)^\top$, $\mathbf{\Upsilon} = (\boldsymbol{\kappa}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$ e $\mathbf{U}(\boldsymbol{\kappa})$ e $\mathbf{U}(\boldsymbol{\theta})$ serão dados a seguir e dependerão do estimador considerado.

Definição 6.6 (M-SE.) *O estimador M-SE é obtido através da solução de (6.19) com $\mathbf{U}(\boldsymbol{\kappa})$ e $\mathbf{U}(\boldsymbol{\theta})$ dados em (6.10) e (6.13), respectivamente. Este estimador considera o MDPDE para a parte discreta e o SMLE para a parte contínua.*

Definição 6.7 (M-LSE.) *O estimador M-LSE é obtido através da solução de (6.19) com $\mathbf{U}(\boldsymbol{\kappa})$ e $\mathbf{U}(\boldsymbol{\theta})$ dados em (6.10) e (6.15), respectivamente. Este estimador considera o MDPDE para a parte discreta e o LSMLE para a parte contínua.*

Definição 6.8 (M-ME.) *O estimador M-ME é obtido através da solução de (6.19) com $\mathbf{U}(\boldsymbol{\kappa})$ e $\mathbf{U}(\boldsymbol{\theta})$ dados em (6.10) e (6.16), respectivamente. Este estimador considera o MDPDE para as partes discreta e contínua.*

Definição 6.9 (M-LME.) *O estimador M-LME é obtido através da solução de (6.19) com $\mathbf{U}(\boldsymbol{\kappa})$ e $\mathbf{U}(\boldsymbol{\theta})$ dados em (6.10) e (6.18), respectivamente. Este estimador considera o MDPDE para a parte discreta e o LMDPDE para a parte contínua.*

A seguir, apresentamos algumas das propriedades dos estimadores robustos. Para todas as propriedades referentes ao M-SE e ao M-ME, estamos considerando que $\mu_i \phi_i \geq 1$ e $(1 - \mu_i) \phi_i \geq 1$, para $i = 1, \dots, n$.

Fisher-consistência

Para verificar a propriedade de Fisher-consistência, mostraremos que as funções de estimação referentes a cada um dos parâmetros são não viesadas. Todos os quatro estimadores têm a mesma função de estimação para a parte discreta, dada em (6.10). Na proposição a seguir, mostramos que essa função de estimação é não viesada.

Proposição 6.1 *A função de estimação em (6.10) é não viesada.*

Demonstração. Na Seção 6.2.2, mostramos que $\mathbb{E}[\mathbf{U}^*(Y_i^c; \boldsymbol{\kappa})] = \mathbf{0}$, $\forall \boldsymbol{\kappa} \in \mathbb{R}^{p_0}$. Como $\mathbf{U}^*(Y_i^c; \boldsymbol{\kappa})$ não depende de $\boldsymbol{\theta}$, a igualdade vale $\forall \mathbf{\Upsilon} \in \mathbb{R}^{p_0+p_1+p_2}$. Logo, a função de estimação referente a $\boldsymbol{\kappa}$ é não viesada. ■

As proposições que seguem enunciam a Fisher-consistência dos estimadores propostos.

Proposição 6.2 *O M-SE e o M-LSE são Fisher-consistentes.*

Demonstração. Mostremos que as equações de estimação em (6.13) e (6.15) são não viesadas. Esse processo é feito de forma análoga e, portanto, mostraremos apenas para o caso do M-SE. Neste caso, precisamos verificar que

$$\mathbb{E} \left[\sum_{i \in \mathcal{D}} \mathbf{U}^*(\mathcal{Y}_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(\mathcal{Y}_i; \mu_i, \phi_i)^\alpha \right] = \mathbf{0}, \quad \forall \mathbf{\Upsilon} \in \mathbb{R}^{p_0+p_1+p_2},$$

com $\mathbf{U}^*(y_i; \boldsymbol{\theta})$ dado em (6.13). Usando a Proposição G.1 do Apêndice G.2, segue que

$$\begin{aligned} & \mathbb{E} \left[\sum_{i \in \mathcal{C}} \mathbf{U}^*(\mathcal{Y}_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(\mathcal{Y}_i; \mu_i, \phi_i)^\alpha \right] = \mathbf{0}, \quad \forall \boldsymbol{\Upsilon} \in \mathbb{R}^{p_0+p_1+p_2} \\ \iff & \sum_{i=1}^n (1 - \vartheta_i) \mathbb{E} \left[\mathbf{U}^*(\mathcal{Y}_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(\mathcal{Y}_i; \mu_i, \phi_i)^\alpha \mathbb{I}_{\{\mathcal{C}\}}(\mathcal{Y}_i) = 0 \right] = \mathbf{0}, \quad \forall \boldsymbol{\Upsilon} \in \mathbb{R}^{p_0+p_1+p_2}. \end{aligned}$$

Assim, é suficiente verificar que $\mathbb{E} \left[\mathbf{U}^*(\mathcal{Y}; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(\mathcal{Y}; \mu, \phi)^\alpha \mathbb{I}_{\{\mathcal{C}\}}(\mathcal{Y}) = 0 \right] = \mathbf{0}, \forall \boldsymbol{\Upsilon} \in \mathbb{R}^{p_0+p_1+p_2}$. De fato, temos que

$$\begin{aligned} \mathbb{E} \left[\mathbf{U}^*(\mathcal{Y}; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(\mathcal{Y}; \mu, \phi)^\alpha \mathbb{I}_{\{\mathcal{C}\}}(\mathcal{Y}) = 0 \right] &= \int_0^1 \mathbf{U}^*(y; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(y; \mu, \phi)^\alpha f_{\boldsymbol{\theta}}(y; \mu, \phi) dy \\ &= \int_0^1 \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log(f_{\boldsymbol{\theta}}^*(y; \mu, \phi)) \right] f_{\boldsymbol{\theta}}^*(y; \mu, \phi)^\alpha f_{\boldsymbol{\theta}}(y; \mu, \phi) dy \\ &= \int_0^1 \frac{(\partial/\partial \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(y; \mu, \phi)}{f_{\boldsymbol{\theta}}^*(y; \mu, \phi)} f_{\boldsymbol{\theta}}^*(y; \mu, \phi)^\alpha f_{\boldsymbol{\theta}}(y; \mu, \phi) dy \\ &= \int_0^1 \frac{(\partial/\partial \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(y; \mu, \phi)}{f_{\boldsymbol{\theta}}^*(y; \mu, \phi)^{1-\alpha}} f_{\boldsymbol{\theta}}(y; \mu, \phi) dy. \end{aligned}$$

Como $f_{\boldsymbol{\theta}}(y; \mu, \phi) = c_\alpha(\boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(y; \mu, \phi)^{1-\alpha}$, com $c_\alpha(\boldsymbol{\theta}) = \left[\int_0^1 f_{\boldsymbol{\theta}}(y; \mu, \phi)^{1/(1-\alpha)} dy \right]^{1-\alpha}$, temos que

$$\begin{aligned} \mathbb{E} \left[\mathbf{U}^*(\mathcal{Y}; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(\mathcal{Y}; \mu, \phi)^\alpha \mathbb{I}_{\{\mathcal{C}\}}(\mathcal{Y}) = 0 \right] &= c_\alpha(\boldsymbol{\theta}) \int_0^1 \frac{\partial}{\partial \boldsymbol{\theta}} f_{\boldsymbol{\theta}}^*(y; \mu, \phi) dy \\ &= c_\alpha(\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} \int_0^1 f_{\boldsymbol{\theta}}^*(y; \mu, \phi) dy \\ &= 0, \quad \forall \boldsymbol{\Upsilon} \in \mathbb{R}^{p_0+p_1+p_2}. \end{aligned}$$

Com o resultado da Proposição 6.1, segue a Fisher-consistência do M-SE. A Fisher-consistência do M-LSE pode ser demonstrada de forma análoga. ■

Proposição 6.3 *O M-ME e o M-LME são Fisher-consistentes.*

Demonstração. Da Proposição 6.1 e de (6.16) e (6.18), segue que as funções de estimação do M-ME e do M-LME são não viesadas. Assim, esses estimadores são Fisher-consistentes. ■

B-robustez

Proposição 6.4 *Os estimadores M-SE, M-LSE, M-ME e M-LME são B-robustos.*

Para verificar essa proposição, basta verificar que os componentes das funções de estimação são limitados. Os componentes da função de estimação referentes à parte discreta (6.10) são claramente limitados, uma vez que $y^c \in \{0, 1\}$. As demonstrações de que os componentes das funções de estimação referentes à parte contínua são limitados podem ser encontradas em Ribeiro e Ferrari (2022) para o SMLE e MDPDE e em Maluf *et al.* (2022) para o LSMLE e LMDPDE.

Normalidade assintótica

A normalidade assintótica dos estimadores propostos vem do fato de serem M-estimadores. Assim, os estimadores M-SE, M-LSE, M-ME e M-LME têm distribuição assintoticamente normal

com média Υ e matriz de variância e covariância $V_\alpha(\Upsilon)$. Uma vez que a função de estimação da parte discreta não depende dos parâmetros da parte contínua (e vice-versa), a matriz $V_\alpha(\Upsilon)$ será bloco diagonal. Isto é, podemos escrever

$$V_\alpha(\Upsilon) = \begin{bmatrix} V_\alpha(\boldsymbol{\kappa}) & \mathbf{0} \\ \mathbf{0} & V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma}) \end{bmatrix}.$$

Como os quatro estimadores compartilham do mesmo método para estimar os parâmetros associados à parte discreta, a matriz $V_\alpha(\boldsymbol{\kappa})$ é a mesma para todos eles. Essa matriz é dada por

$$V_\alpha(\boldsymbol{\kappa}) = \mathbf{A}_\alpha^{-1}(\boldsymbol{\kappa})\mathbf{B}_\alpha(\boldsymbol{\kappa})\mathbf{A}_\alpha^{-1}(\boldsymbol{\kappa}),$$

com

$$\begin{aligned} \mathbf{A}_\alpha(\boldsymbol{\kappa}) &= \sum_{i=1}^n \mathbb{E} \left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \right] = \mathbf{S}^\top \mathbf{M} \boldsymbol{\Lambda} \mathbf{T}_\vartheta^2 \mathbf{S}, \\ \mathbf{B}_\alpha(\boldsymbol{\kappa}) &= \sum_{i=1}^n \mathbb{E} \left[\left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right)^\top \right] = \mathbf{S}^\top \mathbf{M}^2 \boldsymbol{\Lambda} \mathbf{T}_\vartheta^2 \mathbf{S}, \end{aligned} \quad (6.20)$$

em que $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_{p_0})$, $\mathbf{M} = \text{diag}\{M_i; i = 1, \dots, n\}$, $\boldsymbol{\Lambda} = \text{diag}\{[\vartheta_i(1 - \vartheta_i)]^{-1}; i = 1, \dots, n\}$, $\mathbf{T}_\vartheta = \text{diag}\{[g'_\vartheta(\vartheta_i)]^{-1}; i = 1, \dots, n\}$ e

$$M_i = M_i(\boldsymbol{\kappa}, \alpha) = (1 + \alpha) [(1 - \vartheta_i)\vartheta_i^\alpha + \vartheta_i(1 - \vartheta_i)^\alpha].$$

No Apêndice G.1 apresentamos detalhes dos cálculos para a obtenção de $V_\alpha(\boldsymbol{\kappa})$.

A matriz $V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma})$ depende do estimador robusto considerado. A seguir, apresentamos as expressões para essa matriz para cada um dos estimadores em estudo. A obtenção dessa matriz para cada um dos estimadores é muito similar. No Apêndice G.2 apresentamos detalhes referentes à obtenção de $V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma})$ para o M-LSE.

M-SE. A matriz $V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma})$ para o M-SE é dada por

$$V_\alpha(\boldsymbol{\theta}) = \mathbf{J}_\alpha^{-1}(\boldsymbol{\theta})\mathbf{K}_\alpha(\boldsymbol{\theta})\mathbf{J}_\alpha^{-1}(\boldsymbol{\theta}),$$

em que

$$\mathbf{J}_\alpha(\boldsymbol{\theta}) = \mathbb{E} \left\{ \sum_{i \in \varnothing} \frac{\partial}{\partial \boldsymbol{\theta}^\top} [\mathbf{U}^*(\mathcal{Y}_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(\mathcal{Y}_i; \mu_i, \phi_i)^\alpha] \right\}$$

e

$$\mathbf{K}_\alpha(\boldsymbol{\theta}) = \mathbb{E} \left\{ \sum_{i \in \varnothing} \mathbf{U}^*(\mathcal{Y}_i; \boldsymbol{\theta}) \mathbf{U}^*(\mathcal{Y}_i; \boldsymbol{\theta})^\top f_{\boldsymbol{\theta}}^*(\mathcal{Y}_i; \mu_i, \phi_i)^{2\alpha} \right\},$$

sendo $\mathbf{U}^*(y_i; \boldsymbol{\theta})$ dado em (6.13) e as matrizes $\mathbf{J}_\alpha(\boldsymbol{\theta})$ e $\mathbf{K}_\alpha(\boldsymbol{\theta})$ são dadas, respectivamente, por

$$\mathbf{J}_\alpha(\boldsymbol{\theta}) = -(1 - \alpha)^{-1} \begin{bmatrix} \boldsymbol{\mathcal{X}}^\top \mathbf{B}_1 \mathbf{A} \mathbf{T}_\mu^{*2} \boldsymbol{\Phi}_\alpha \mathbf{V} \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^\top \mathbf{B}_1 \mathbf{A} \mathbf{T}_\mu^* \mathbf{T}_\phi^* \mathbf{C}^* \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\mathcal{Z}}^\top \mathbf{B}_1 \mathbf{A} \mathbf{T}_\mu^* \mathbf{A} \mathbf{T}_\phi^* \mathbf{C}^* \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{Z}}^\top \mathbf{B}_1 \mathbf{A} \mathbf{T}_\phi^{*2} \mathbf{D}^* \boldsymbol{\mathcal{Z}} \end{bmatrix}$$

e

$$\mathbf{K}_\alpha(\boldsymbol{\theta}) = (1 - \alpha)^2 \begin{bmatrix} \boldsymbol{\mathcal{X}}^\top \mathbf{B}_2 \mathbf{A} \mathbf{T}_\mu^{*2} \boldsymbol{\Phi}_\alpha^2 (\mathbf{V}_{1+\alpha} + \mathbf{M}_1^2) \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^\top \mathbf{B}_2 \mathbf{A} \mathbf{T}_\mu^* \mathbf{T}_\phi^* (\mathbf{C}_{1+\alpha}^* + \mathbf{M}_3) \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\mathcal{Z}}^\top \mathbf{B}_2 \mathbf{A} \mathbf{T}_\mu^* \mathbf{T}_\phi^* (\mathbf{C}_{1+\alpha}^* + \mathbf{M}_3) \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{Z}}^\top \mathbf{B}_2 \mathbf{A} \mathbf{T}_\phi^{*2} (\mathbf{D}_{1+\alpha}^* + \mathbf{M}_2^2) \boldsymbol{\mathcal{Z}} \end{bmatrix},$$

sendo $\mathbf{B}_j = \text{diag}\{b_{i,j}; i = 1, \dots, n\}$, $j = 1, 2$,

$$b_{i,1} = \frac{B(\mu_i \phi_i, (1 - \mu_i) \phi_i)^{1-\alpha}}{B(\mu_{i,1-\alpha} \phi_{i,1-\alpha}, (1 - \mu_{i,1-\alpha}) \phi_{i,1-\alpha})},$$

$$b_{i,2} = \frac{B(\mu_{i,1+\alpha} \phi_{i,1+\alpha}, (1 - \mu_{i,1+\alpha}) \phi_{i,1+\alpha})}{B(\mu_i \phi_i, (1 - \mu_i) \phi_i)^{2\alpha} B(\mu_{i,1-\alpha} \phi_{i,1-\alpha}, (1 - \mu_{i,1-\alpha}) \phi_{i,1-\alpha})},$$

$\mathbf{T}_\mu^* = \text{diag}\{t_{i,\mu}; i = 1, \dots, n\}$, $\mathbf{T}_\phi^* = \text{diag}\{t_{i,\phi}; i = 1, \dots, n\}$,

$$t_{i,\mu} = [g'_\mu(\mu_{i,1-\alpha})]^{-1}, \quad t_{i,\phi} = [g'_\phi(\phi_{i,1-\alpha})]^{-1},$$

$\mathbf{A} = \text{diag}\{1 - \vartheta_i; i = 1, \dots, n\}$, $\mathbf{\Phi}_\alpha = \text{diag}\{\phi_{i,1-\alpha}; i = 1, \dots, n\}$, $\mathbf{V} = \text{diag}\{v_i; i = 1, \dots, n\}$,
 $\mathbf{V}_{1+\alpha} = \text{diag}\{v_{i,1+\alpha}; i = 1, \dots, n\}$

$$v_i = \text{Var}(\mathcal{Y}_i^* | \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0) = \psi'(\mu_i \phi_i) + \psi'((1 - \mu_i) \phi_i),$$

$$v_{i,1+\alpha} = \psi'(\mu_{i,1+\alpha} \phi_{i,1+\alpha}) + \psi'((1 - \mu_{i,1+\alpha}) \phi_{i,1+\alpha}),$$

$\mathbf{C}^* = \text{diag}\{c_i^*; i = 1, \dots, n\}$, $\mathbf{C}_{1+\alpha}^* = \text{diag}\{c_{i,1+\alpha}^*; i = 1, \dots, n\}$

$$c_i^* = \phi_{i,1-\alpha} [\mu_{i,1-\alpha} \psi'(\mu_i \phi_i) - (1 - \mu_{i,1-\alpha}) \psi'((1 - \mu_i) \phi_i)],$$

$$c_{i,1+\alpha}^* = \phi_{i,1-\alpha} [\mu_{i,1-\alpha} \psi'(\mu_{i,1+\alpha} \phi_{i,1+\alpha}) - (1 - \mu_{i,1-\alpha}) \psi'((1 - \mu_{i,1+\alpha}) \phi_{i,1+\alpha})],$$

$\mathbf{D}^* = \text{diag}\{d_i^*; i = 1, \dots, n\}$, $\mathbf{D}_{1+\alpha}^* = \text{diag}\{d_{i,1+\alpha}^*; i = 1, \dots, n\}$

$$d_i^* = \mu_{i,1-\alpha}^2 \psi'(\mu_i \phi_i) + (1 - \mu_{i,1-\alpha})^2 \psi'((1 - \mu_i) \phi_i) - \psi'(\phi_i),$$

$$d_{i,1+\alpha}^* = \mu_{i,1-\alpha}^2 \psi'(\mu_{i,1+\alpha} \phi_{i,1+\alpha}) + (1 - \mu_{i,1-\alpha})^2 \psi'((1 - \mu_{i,1+\alpha}) \phi_{i,1+\alpha}) - \psi'(\phi_{i,1+\alpha}),$$

$\mathbf{M}_1 = \text{diag}\{(\mu_{i,1+\alpha}^* - \mu_i^*); i = 1, \dots, n\}$, $\mathbf{M}_2 = \text{diag}\{\mu_{i,1+\alpha}^d; i = 1, \dots, n\}$,

$\mathbf{M}_3 = \text{diag}\{\mu_{i,1+\alpha}^d \phi_{i,1-\alpha} (\mu_{i,1+\alpha}^* - \mu_i^*); i = 1, \dots, n\}$ e $\mu_{i,1+\alpha}^d = \mu_{i,1-\alpha}^* (\mu_{i,1+\alpha}^* - \mu_i^*) + \mu_{i,1+\alpha}^\dagger - \mu_i^\dagger$.

A matriz $\mathbf{V}_\alpha(\boldsymbol{\theta})$ está bem definida para $\mu_i \phi_i > 2\alpha/(1 + \alpha)$ e $(1 - \mu_i) \phi_i > 2\alpha/(1 + \alpha)$. Uma condição suficiente para que isso valha para todo $0 \leq \alpha < 1$ é que $\mu_i \phi_i \geq 1$ e $(1 - \mu_i) \phi_i \geq 1$.

M-LSE. A matriz $V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma})$ para o M-LSE é dada por

$$\mathbf{V}_\alpha(\boldsymbol{\theta}) = \mathbf{J}_\alpha^{-1}(\boldsymbol{\theta}) \mathbf{K}_\alpha(\boldsymbol{\theta}) \mathbf{J}_\alpha^{-1}(\boldsymbol{\theta}),$$

em que

$$\mathbf{J}_\alpha(\boldsymbol{\theta}) = \mathbb{E} \left\{ \sum_{i \in \varphi} \frac{\partial}{\partial \boldsymbol{\theta}^\top} [\mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) h_\theta^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha] \right\}$$

e

$$\mathbf{K}_\alpha(\boldsymbol{\theta}) = \mathbb{E} \left\{ \sum_{i \in \varphi} \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top h_\theta^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha} \right\},$$

sendo $\mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})$ dado em (6.15) e as matrizes $\mathbf{J}_\alpha(\boldsymbol{\theta})$ e $\mathbf{K}_\alpha(\boldsymbol{\theta})$ dadas por

$$\mathbf{J}_\alpha(\boldsymbol{\theta}) = - \begin{bmatrix} (1 - \alpha) \boldsymbol{\mathcal{X}}^\top \mathbf{A} \mathbf{B}'_1 \mathbf{T}_\mu^2 \boldsymbol{\Phi}^2 \mathbf{V} \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^\top \mathbf{A} \mathbf{B}'_1 \mathbf{T}_\mu \mathbf{T}_\phi^* \mathbf{C} \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \mathbf{B}'_1 \mathbf{T}_\mu \mathbf{T}_\phi^* \mathbf{C} \boldsymbol{\mathcal{X}} & (1 - \alpha)^{-1} \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \mathbf{B}'_1 \mathbf{T}_\phi^{*2} \mathbf{D} \boldsymbol{\mathcal{Z}} \end{bmatrix} \quad (6.21)$$

e

$$\mathbf{K}_\alpha(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}^\top \mathbf{A} \mathbf{B}'_2 \mathbf{T}_\mu^2 \boldsymbol{\Phi}^2 \mathbf{V}'_{1+\alpha} \boldsymbol{\mathcal{X}} & (1-\alpha)^{-1} \boldsymbol{\mathcal{X}}^\top \mathbf{A} \mathbf{B}'_2 \mathbf{T}_\mu \mathbf{T}'_\phi^* \mathbf{C}'_{1+\alpha} \boldsymbol{\mathcal{Z}} \\ (1-\alpha)^{-1} \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \mathbf{B}'_2 \mathbf{T}_\mu \mathbf{T}'_\phi^* \mathbf{C}'_{1+\alpha} \boldsymbol{\mathcal{X}} & (1-\alpha)^{-2} \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \mathbf{B}'_2 \mathbf{T}_\phi^{*2} \mathbf{D}'_{1+\alpha} \boldsymbol{\mathcal{Z}} \end{bmatrix}, \quad (6.22)$$

sendo $\mathbf{B}'_j = \text{diag}\{b'_{i,j}; i = 1, \dots, n\}$, $j = 1, 2$,

$$b'_{i,1} = \frac{B(\mu_i \phi_i, (1-\mu_i) \phi_i)^{1-\alpha}}{B(\mu_i \phi'_{i,1-\alpha}, (1-\mu_i) \phi'_{i,1-\alpha})},$$

$$b'_{i,2} = \frac{B(\mu_i \phi'_{i,1+\alpha}, (1-\mu_i) \phi'_{i,1+\alpha})}{B(\mu_i \phi_i, (1-\mu_i) \phi_i)^{2\alpha} B(\mu_i \phi'_{i,1-\alpha}, (1-\mu_i) \phi'_{i,1-\alpha})},$$

 $\mathbf{T}_\mu = \text{diag}\{t_{i,\mu}; i = 1, \dots, n\}$, $\mathbf{T}'_\phi^* = \text{diag}\{t'_{i,\phi}; i = 1, \dots, n\}$,

$$t_{i,\mu} = [g'_\mu(\mu_i)]^{-1}, \quad t'_{i,\phi} = [g'_\phi(\phi'_{i,1-\alpha})]^{-1},$$

 $\boldsymbol{\Phi} = \text{diag}\{\phi_i; i = 1, \dots, n\}$, $\mathbf{V} = \text{diag}\{v_i; i = 1, \dots, n\}$, $\mathbf{V}'_{1+\alpha} = \text{diag}\{v'_{i,1+\alpha}; i = 1, \dots, n\}$

$$v'_{i,1+\alpha} = \psi'(\mu_i \phi'_{i,1+\alpha}) + \psi'((1-\mu_i) \phi'_{i,1+\alpha}),$$

 $\mathbf{C} = \text{diag}\{c_i; i = 1, \dots, n\}$, $\mathbf{C}'_{1+\alpha} = \text{diag}\{c'_{i,1+\alpha}; i = 1, \dots, n\}$

$$c_i = \phi_i [\mu_i \psi'(\mu_i \phi_i) - (1-\mu_i) \psi'((1-\mu_i) \phi_i)],$$

$$c'_{i,1+\alpha} = \phi_i [\mu_i \psi'(\mu_i \phi'_{i,1+\alpha}) - (1-\mu_i) \psi'((1-\mu_i) \phi'_{i,1+\alpha})],$$

 $\mathbf{D} = \text{diag}\{d_i; i = 1, \dots, n\}$, $\mathbf{D}'_{1+\alpha} = \text{diag}\{d'_{i,1+\alpha}; i = 1, \dots, n\}$

$$d_i = \mu_i^2 \psi'(\mu_i \phi_i) + (1-\mu_i)^2 \psi'((1-\mu_i) \phi_i) - \psi'(\phi_i),$$

$$d'_{i,1+\alpha} = \mu_i^2 \psi'(\mu_i \phi'_{i,1+\alpha}) + (1-\mu_i)^2 \psi'((1-\mu_i) \phi'_{i,1+\alpha}) - \psi'(\phi'_{i,1+\alpha}),$$

As matrizes $\mathbf{J}_\alpha(\boldsymbol{\theta})$ e $\mathbf{K}_\alpha(\boldsymbol{\theta})$ estão bem definidas para todos os valores de μ_i e ϕ_i .M-ME. A matriz $V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma})$ para o M-ME é dada por

$$V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{\Psi}_\alpha^{-1}(\boldsymbol{\theta}) \boldsymbol{\Omega}_\alpha(\boldsymbol{\theta}) \boldsymbol{\Psi}_\alpha^{-1}(\boldsymbol{\theta}),$$

em que

$$\boldsymbol{\Psi}_\alpha(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}^\top \mathbf{A} \gamma_{11}^{(1+\alpha)} \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^\top \mathbf{A} \gamma_{12}^{(1+\alpha)} \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \gamma_{12}^{(1+\alpha)} \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \gamma_{22}^{(1+\alpha)} \boldsymbol{\mathcal{Z}} \end{bmatrix}$$

e

$$\boldsymbol{\Omega}_\alpha(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}^\top \mathbf{A} \left[\gamma_{11}^{(1+2\alpha)} - \gamma_1^{(1+\alpha)2} \right] \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^\top \mathbf{A} \left[\gamma_{12}^{(1+2\alpha)} - \gamma_1^{(1+\alpha)} \gamma_2^{(1+\alpha)} \right] \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \left[\gamma_{12}^{(1+2\alpha)} - \gamma_1^{(1+\alpha)} \gamma_2^{(1+\alpha)} \right] \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \left[\gamma_{22}^{(1+2\alpha)} - \gamma_2^{(1+\alpha)2} \right] \boldsymbol{\mathcal{Z}} \end{bmatrix},$$

sendo $\gamma_j^{(\alpha)} = \text{diag}\{\gamma_{j,i}^{(\alpha)}; i = 1, \dots, n\}$, para $j = 1, 2$, $\gamma_{11}^{(\alpha)} = \text{diag}\{\gamma_{11,i}^{(\alpha)}; i = 1, \dots, n\}$, $\gamma_{12}^{(\alpha)} =$

$\text{diag}\{\gamma_{12,i}^{(\alpha)}; i = 1, \dots, n\}$, $\gamma_{22}^{(\alpha)} = \text{diag}\{\gamma_{22,i}^{(\alpha)}; i = 1, \dots, n\}$, em que

$$\begin{aligned}\gamma_{11,i}^{(\alpha)} &= \frac{\phi_i^2 \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{[g'_\mu(\mu_i)]^2} \left[v_{i,\alpha} + (\mu_{i,\alpha}^* - \mu_i^*)^2 \right], \\ \gamma_{12,i}^{(\alpha)} &= \frac{\phi_i \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{g'_\mu(\mu_i) g'_\phi(\phi_i)} \left\{ \mu_i \left[v_{i,\alpha} + (\mu_{i,\alpha}^* - \mu_i^*)^2 \right] - \psi'((1 - \mu_{i,\alpha})\phi_{i,\alpha}) + (\mu_{i,\alpha}^* - \mu_i^*)(\mu_{i,\alpha}^\dagger - \mu_i^\dagger) \right\}, \\ \gamma_{22,i}^{(\alpha)} &= \frac{\mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{[g'_\phi(\phi_i)]^2} \left\{ \mu_i^2 \psi'(\mu_{i,\alpha}\phi_{i,\alpha}) + (1 - \mu_i)^2 \psi'((1 - \mu_{i,\alpha})\phi_{i,\alpha}) - \psi'(\phi_{i,\alpha}) \right. \\ &\quad \left. + \left[\mu_i(\mu_{i,\alpha}^* - \mu_i^*) + (\mu_{i,\alpha}^\dagger - \mu_i^\dagger) \right]^2 \right\}.\end{aligned}$$

As matrizes $\boldsymbol{\Psi}_\alpha$ e $\boldsymbol{\Omega}_\alpha$ estão bem definidas para $\mu_i\phi_i > 2\alpha/(1+2\alpha)$ e $(1-\mu_i)\phi_i > 2\alpha/(1+2\alpha)$. Uma condição suficiente para que isso valha para todo $0 \leq \alpha < 1$ é que $\mu_i\phi_i \geq 1$ e $(1-\mu_i)\phi_i \geq 1$.

M-LME. A matriz $V_\alpha(\boldsymbol{\beta}, \boldsymbol{\gamma})$ para o M-LME é dada por

$$V_\alpha(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_\alpha^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}) \boldsymbol{\Lambda}_\alpha^{-1}(\boldsymbol{\theta}),$$

em que

$$\boldsymbol{\Lambda}_\alpha(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}^\top \mathbf{A} \boldsymbol{\Lambda}_{11}^{(1+\alpha)} \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^\top \mathbf{A} \boldsymbol{\Lambda}_{12}^{(1+\alpha)} \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \boldsymbol{\Lambda}_{12}^{(1+\alpha)} \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \boldsymbol{\Lambda}_{22}^{(1+\alpha)} \boldsymbol{\mathcal{Z}} \end{bmatrix}$$

e

$$\boldsymbol{\Omega}_\alpha(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}^\top \mathbf{A} \left[\boldsymbol{\Lambda}_{11}^{(1+2\alpha)} - \boldsymbol{\Lambda}_1^{(1+\alpha)^2} \right] \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^\top \mathbf{A} \left[\boldsymbol{\Lambda}_{12}^{(1+2\alpha)} - \boldsymbol{\Lambda}_1^{(1+\alpha)} \boldsymbol{\Lambda}_2^{(1+\alpha)} \right] \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \left[\boldsymbol{\Lambda}_{12}^{(1+2\alpha)} - \boldsymbol{\Lambda}_1^{(1+\alpha)} \boldsymbol{\Lambda}_2^{(1+\alpha)} \right] \boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{Z}}^\top \mathbf{A} \left[\boldsymbol{\Lambda}_{22}^{(1+2\alpha)} - \boldsymbol{\Lambda}_2^{(1+\alpha)^2} \right] \boldsymbol{\mathcal{Z}} \end{bmatrix},$$

sendo $\boldsymbol{\Lambda}_j^{(\alpha)} = \text{diag}\{\Lambda_{j,i}^{(\alpha)}; i = 1, \dots, n\}$, para $j = 1, 2$, $\Lambda_{11}^{(\alpha)} = \text{diag}\{\Lambda_{11,i}^{(\alpha)}; i = 1, \dots, n\}$, $\Lambda_{12}^{(\alpha)} = \text{diag}\{\Lambda_{12,i}^{(\alpha)}; i = 1, \dots, n\}$, $\Lambda_{22}^{(\alpha)} = \text{diag}\{\Lambda_{22,i}^{(\alpha)}; i = 1, \dots, n\}$, em que

$$\begin{aligned}\Lambda_{11,i}^{(\alpha)} &= \frac{\phi_i^2 \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{[g'_\mu(\mu_i)]^2} \left[v'_{i,\alpha} + (\mu'_{i,\alpha} - \mu_i^*)^2 \right], \\ \Lambda_{12,i}^{(\alpha)} &= \frac{\phi_i \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{g'_\mu(\mu_i) g'_\phi(\phi_i)} \left\{ \mu_i \left[v'_{i,\alpha} + (\mu'_{i,\alpha} - \mu_i^*)^2 \right] - \psi'((1 - \mu_i)\phi'_{i,\alpha}) + (\mu'_{i,\alpha} - \mu_i^*)(\mu_{i,\alpha}^\dagger - \mu_i^\dagger) \right\}, \\ \Lambda_{22,i}^{(\alpha)} &= \frac{\mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{[g'_\phi(\phi_i)]^2} \left\{ \mu_i^2 \psi'(\mu_i\phi'_{i,\alpha}) + (1 - \mu_i)^2 \psi'((1 - \mu_i)\phi'_{i,\alpha}) - \psi'(\phi'_{i,\alpha}) \right. \\ &\quad \left. + \left[\mu_i(\mu'_{i,\alpha} - \mu_i^*) + (\mu_{i,\alpha}^\dagger - \mu_i^\dagger) \right]^2 \right\}.\end{aligned}$$

Note que as matrizes $\boldsymbol{\Lambda}_\alpha(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta})$ estão bem definidas para todos os valores de μ_i e ϕ_i , para todo $0 \leq \alpha < 1$.

6.2.6 Testes de hipóteses robustos

Um aspecto importante no ajuste de um modelo de regressão é o de testar hipóteses a respeito dos parâmetros do modelo. Os testes de hipóteses clássicos (como o teste Wald, por exemplo) baseados nas estimativas de máxima verossimilhança herdam a sensibilidade do estimador a dados atípicos, de modo que essas observações podem afetar o tamanho ou o poder do teste de forma considerável. É desejável que tenhamos um procedimento de teste que não seja fortemente influen-

ciado por essas observações. Chamaremos esse tipo de teste de “testes robustos”. Bianco e Martínez (2009) e Ribeiro e Ferrari (2022) propõem testes de hipóteses robustos para os modelos de regressão logística e beta, respectivamente. A ideia, basicamente, é obter um teste tipo-Wald robusto através da substituição da estimativa de máxima verossimilhança e do erro padrão pela estimativa robusta e erro padrão baseado na matriz de variância e covariância assintótica do estimador robusto. Neste trabalho, também consideraremos essa ideia.

Considere o problema de testar a hipótese nula $\mathcal{H}_0 : \Upsilon_j = \Upsilon_j^{(0)}$ contra a alternativa $\mathcal{H}_1 : \Upsilon_j \neq \Upsilon_j^{(0)}$. A estatística do teste tipo-Wald é definida por

$$W_\alpha = \left(\frac{\hat{\Upsilon}_{\alpha j} - \Upsilon_j^{(0)}}{\text{se}(\hat{\Upsilon}_{\alpha j})} \right)^2,$$

em que $\text{se}(\hat{\Upsilon}_{\alpha j})$ é o erro padrão assintótico de $\hat{\Upsilon}_{\alpha j}$ e $\hat{\Upsilon}_{\alpha j}$ representa qualquer um dos estimadores robustos definidos na Seção 6.2.5. Sob a hipótese nula, W_α tem distribuição assintótica χ_1^2 .

6.2.7 Seleção da constante de afinação α

Um dos aspectos cruciais em inferência robusta é a escolha da constante de afinação que frequentemente está presente nesses procedimentos. Essa constante produz um balanço (“*trade-off*”) entre robustez e eficiência; no nosso caso, valores de α afastados do zero levam a maior robustez e menor eficiência. A maioria dos trabalhos, em geral, baseia-se em estudos de simulação para sugerir valores específicos para essa constante e/ou sugere métodos para escolha da constante que dependem de um estimador piloto, que também deve ser fixado a priori (veja Croux e Haesbroeck (2003), Warwick e Jones (2005), Ghosh e Basu (2016), Ghosh (2019)). Ribeiro e Ferrari (2022) propuseram um algoritmo orientado pelos dados para a escolha dessa constante que se mostrou bastante eficaz. Esse algoritmo será adaptado para os estimadores robustos propostos para o modelo de regressão beta inflacionado em zero ou um. A ideia do algoritmo é selecionar α mais próximo de zero (isto é, de máxima verossimilhança) tal que as estimativas dos parâmetros sejam razoavelmente estáveis, de modo a ter eficiência completa na ausência de contaminação.

O algoritmo proposto pelas autoras é descrito a seguir. Consideremos que o vetor de parâmetros de interesse é $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)^\top$. Seja $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \dots \leq 1$ uma grade de valores para α igualmente espaçada e defina z_{α_k} como sendo

$$z_{\alpha_k} = \left(\frac{\hat{\delta}_{\alpha_k}^1}{\sqrt{n} \text{se}(\hat{\delta}_{\alpha_k}^1)}, \dots, \frac{\hat{\delta}_{\alpha_k}^p}{\sqrt{n} \text{se}(\hat{\delta}_{\alpha_k}^p)} \right)^\top,$$

em que $\text{se}(\hat{\delta}_{\alpha_k}^j)$ é o erro padrão assintótico de $\hat{\delta}_{\alpha_k}^j$. Defina as variações quadráticas padronizadas (*standardized quadratic variations*; SQV) como sendo

$$\text{SQV}_{\alpha_k} = p^{-1} \|z_{\alpha_k} - z_{\alpha_{k+1}}\|.$$

Note que se SQV_{α_k} é pequeno, então as estimativas com $\alpha = \alpha_k$ e $\alpha = \alpha_{k+1}$ são próximas.

Os passos do algoritmo são descritos a seguir.

1. Defina uma grade ordenada e igualmente espaçada para α : $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \dots \leq \alpha_{m_1}$.
2. Seja $L > 0$ um valor limitante pré-definido. Se a condição de estabilidade $\text{SQV}_{\alpha_k} < L$ é satisfeita para todo $k = 0, 1, \dots, m_1 - 1$, tome o valor ótimo de α como sendo $\alpha^* = \alpha_0 = 0$; caso contrário, considere α_{start} como sendo o próximo ponto na grade depois do maior α_k para o qual a condição de estabilidade não é satisfeita;
3. defina uma nova grade ordenada e igualmente espaçada para α começando de α_{start} : $\alpha_0 = \alpha_{\text{start}} < \alpha_1 < \alpha_2 < \dots < \alpha_m$, em que $\alpha_m \leq \alpha_{\text{max}}$;

4. se a condição de estabilidade é satisfeita para todo $k = 0, 1, \dots, m - 1$, então tome o valor ótimo de α como sendo $\alpha^* = \alpha_0 = \alpha_{\text{start}}$; caso contrário, considere α_{start} como sendo o próximo ponto na grade depois do maior α_k para o qual a condição de estabilidade não vale;
5. repita os passos 3 e 4 até atingir estabilidade nas estimativas para a grade atual ou atingir o valor máximo α_{max} ;
6. se α_{max} foi atingido sem estabilidade, repita os passos de 3 a 5 considerando $\alpha_{\text{start}} = 0$;
7. se α_{max} foi atingido sem estabilidade na última grade, tome $\alpha^* = 0$.

Esse algoritmo começa verificando se a condição de estabilidade é satisfeita para todos os valores de α menores ou iguais a α_{m_1} em uma grade ordenada e igualmente espaçada começando de $\alpha = 0$. Se sim, o método escolhe $\alpha^* = 0$ como valor ótimo, ou seja, o estimador de máxima verossimilhança é escolhido. Caso contrário, o algoritmo segue para o próximo ponto na grade depois do maior valor de α para o qual a condição de estabilidade não é satisfeita, e procura por uma sequência de m valores de α consecutivos e igualmente espaçados para o qual a condição de estabilidade vale. Se essa sequência não é encontrada, a procura começa novamente a partir de $\alpha = 0$. Pode ocorrer que o α ótimo não seja encontrado; neste caso, o algoritmo toma $\alpha^* = 0$. Conforme mencionado em [Ribeiro e Ferrari \(2022\)](#), no caso do SMLE e do MDPDE para θ , isso pode acontecer devido a esses estimadores robustos não estarem bem definidos para densidades beta ilimitadas.

Através de estudos de simulação e aplicações em dados reais, [Ribeiro e Ferrari \(2022\)](#) sugerem a seguinte configuração para o SMLE e MDPDE para θ :

- $L = 0.02$, $\alpha_{m_1} = 0.2$, $\alpha_{\text{max}} = 0.5$, $m = 3$ e uma grade com espaçamento 0.02 (e, portanto, $m_1 = 10$).

[Maluf et al. \(2022\)](#) também sugerem o uso desse algoritmo para o LSMLE e LMDPDE para θ e consideram a mesma configuração de [Ribeiro e Ferrari \(2022\)](#).

Neste trabalho, adaptamos o algoritmo proposto por [Ribeiro e Ferrari \(2022\)](#) para escolher separadamente a constante de afinação da parte discreta e a da parte contínua dos estimadores robustos propostos.

Através de estudos de simulação e aplicações em dados simulados, sugerimos que a escolha de α para o MDPDE para κ siga o algoritmo acima com a seguinte configuração:

- $L = 0.02$, $\alpha_{m_1} = 0.5$, $\alpha_{\text{max}} = 1$, $m = 3$ e uma grade com espaçamento 0.05 (e, portanto, $m_1 = 11$).

A escolha de α para os estimadores robustos de θ através do algoritmo acima apresenta um inconveniente: os erros padrão assintóticos dos estimadores de θ , utilizados para o cálculo do SQV, dependem de κ . Para que a escolha da constante de afinação dos estimadores associados à parte contínua independa das estimativas dos parâmetros da parte discreta, propomos utilizar erros padrão obtidos através de uma regressão beta que considera apenas as observações no intervalo unitário.

A maximização ou minimização das funções objetivo utilizadas para a obtenção dos estimadores propostos é feita através de métodos de otimização não linear, e estes necessitam da especificação de pontos iniciais para os parâmetros do modelo. Para a obtenção do SMLE e MDPDE para θ , utilizamos a sugestão de [Ribeiro e Ferrari \(2022\)](#). As autoras consideram pontos iniciais baseados na estimativa de máxima verossimilhança ou em estimativas robustas a partir de uma regressão linear na variável transformada (para detalhes, ver [Ribeiro e Ferrari \(2022\)](#)). Para o LSMLE e LMDPDE no modelo com precisão constante, utilizamos as estimativas de máxima verossimilhança; no modelo com precisão variável, consideramos as estimativas de máxima verossimilhança sob o modelo de precisão constante e os demais parâmetros do submodelo da precisão são iniciados em zero. Para o MDPDE para κ , utilizamos como ponto inicial a estimativa obtida do estimador tipo “Mallows” (veja [Carroll e Pederson \(1993\)](#) para detalhes), disponível no pacote `robust`, através da função `glmRob`.

Para o ajuste do SMLE e do MDPDE para θ , utilizamos as funções disponibilizadas por [Ribeiro e Ferrari \(2022\)](https://github.com/terezinharibeiro/RobustBetaRegression) no repositório do GitHub, que pode ser acessado em <https://github.com/terezinharibeiro/RobustBetaRegression>. As implementações do LSMLE e LMDPDE para θ e do MDPDE para κ foram feitas através de adaptações das funções disponibilizadas por [Ribeiro e Ferrari \(2022\)](https://github.com/ffqueiroz/RobustInflatedBetaRegression) e estão disponíveis no repositório do GitHub através do link: <https://github.com/ffqueiroz/RobustInflatedBetaRegression>.

6.3 Aplicações em dados simulados

Nas aplicações em dados simulados que serão apresentadas nesta seção, exemplificamos a aplicabilidade dos estimadores propostos em modelos de regressão beta inflacionados na presença de contaminação (observações atípicas). Em todas as aplicações, trabalhamos com o modelo de regressão beta inflacionado em zero com a seguinte estrutura de regressão para ϑ_i , μ_i e ϕ_i , respectivamente:

$$\begin{aligned}\log\left(\frac{\vartheta_i}{1-\vartheta_i}\right) &= \kappa_1 + \kappa_2\mathcal{S}_{i1} + \kappa_3\mathcal{S}_{i2}, \\ \log\left(\frac{\mu_i}{1-\mu_i}\right) &= \beta_1 + \beta_2\mathcal{X}_{i1}, \\ \log(\phi_i) &= \gamma_1,\end{aligned}$$

em que $\kappa_1 = 0$, $\kappa_2 = 2$, $\kappa_3 = 2$, $\beta_1 = -1.8$, $\beta_2 = -2$, e $\gamma_1 = 4.5$ e consideramos $n = 100$. As covariáveis \mathcal{S}_1 e \mathcal{S}_2 são obtidas como realizações independentes de uma variável aleatória $Z \sim \mathcal{N}(0, 1)$ e \mathcal{X}_1 é obtida de uma variável aleatória $U \sim U(0, 1)$. Em seguida, uma porcentagem dos dados é contaminada e nosso objetivo é avaliar o efeito dessa contaminação nos métodos de estimação estudados nesta tese. O algoritmo para a seleção das constantes de afinação apresentado na Seção 6.2.7 é utilizado.

Cenário 1

Neste cenário, contaminamos 5% das observações da parte contínua; isto é, observações tais que $y \in (0, 1)$. Por exemplo, em uma amostra de tamanho $n = 100$, suponha que $n^\dagger = 40$ observações são diferentes de zero; neste caso, serão contaminadas $40 \times 0.05 = 2$ observações na parte contínua. A contaminação segue as ideias de [Ribeiro e Ferrari \(2022\)](https://doi.org/10.1007/978-1-4939-9888-8) e é feita substituindo as observações geradas com as 5% menores médias da resposta por observações geradas a partir de um modelo de regressão beta com média $\mu'_i = (1 + \mu_i)/2$ (e precisão ϕ); veja a Figura 6.1.

Na Tabela 6.1 são apresentadas as estimativas dos parâmetros do modelo, com respectivos erros padrão, para os diferentes estimadores propostos, considerando os dados sem contaminação e contaminados de acordo com o cenário 1. Para os dados sem contaminação, observamos que as estimativas coincidem com a estimativa de máxima verossimilhança. De fato, na ausência de contaminação, o valor ótimo para a constante de afinação selecionado para todos os métodos foi zero. Adicionalmente, notamos que as estimativas dos parâmetros da parte discreta para os dados sem contaminação são iguais aquelas para os dados contaminados; isto é, a contaminação, feita apenas na parte contínua, não interfere nas estimativas referentes à parte discreta, o que é esperado.

Por outro lado, a contaminação afeta consideravelmente as estimativas de máxima verossimilhança dos parâmetros associados à parte contínua. Observe na Tabela 6.1 que a estimativa de máxima verossimilhança de β_2 passa de -1.95 para -1.00 e a de γ_1 passa de 4.67 para 2.65 . A diminuição da estimativa de γ_1 tem um significado intuitivo: o método de estimação entende a presença de contaminação como sendo um aumento na dispersão dos dados (diminuição da precisão). Os estimadores robustos, no entanto, apresentam estimativas bem próximas (se não iguais) entre si e àquelas para os dados não contaminados. Os valores ótimos para a constante de afinação para os estimadores dos parâmetros referentes à parte contínua do M-SE, M-ME, M-LSE e M-LME são, respectivamente, 0.08, 0.1, 0.1 e 0.12.

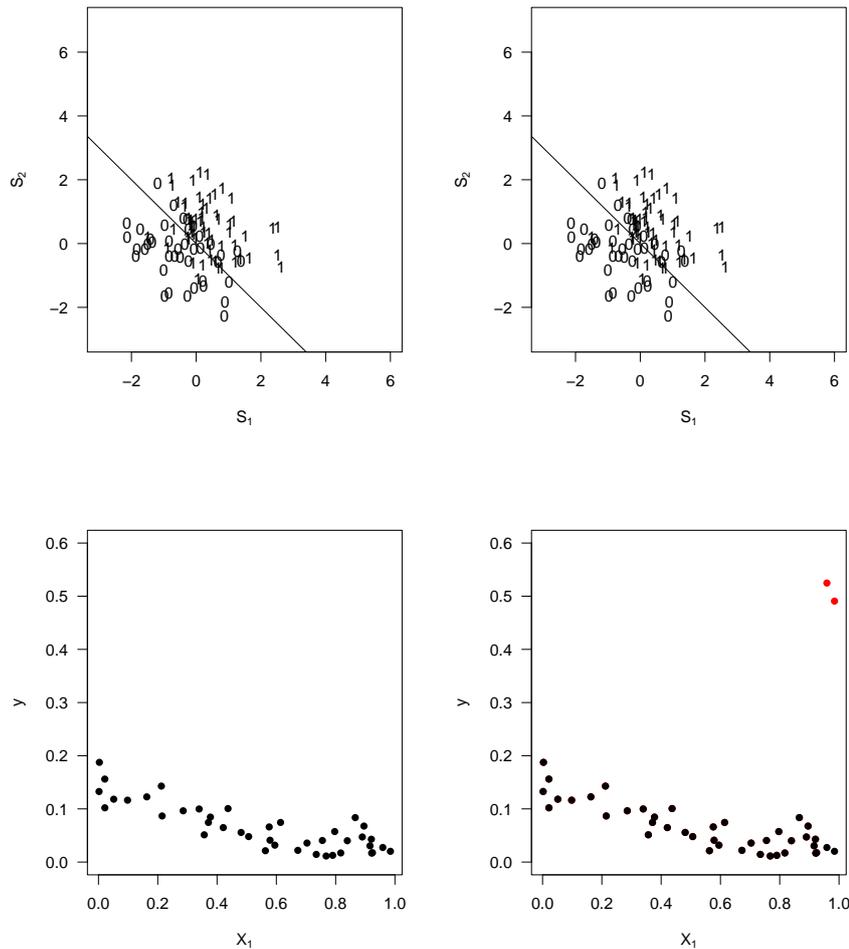


Figura 6.1: Diagramas de dispersão da amostra (parte contínua) sem contaminação (esquerda) e da amostra contaminada segundo o primeiro cenário (direita).

Na Figura 6.2 apresentamos os valores dos SQVs referentes aos estimadores LSMLE e LMDPDE (parte contínua) para os dados sem contaminação e contaminados. Conforme observamos, quando não há contaminação, a grande maioria dos SQVs está abaixo do limiar estabelecido pelo algoritmo ($L = 0.02$), significando que os valores das estimativas não mudam consideravelmente à medida que aumentamos o valor de α . Por outro lado, na presença de contaminação, as estimativas são instáveis (SQVs $> L$) e se estabilizam a partir do valor ótimo de α selecionado pelo algoritmo.

Cenário 2

Neste cenário, contaminamos 5% das observações seguindo as ideias de [Croux e Haesbroeck \(2003\)](#). A contaminação é feita substituindo as observações geradas com os 5% maiores valores de ϑ por observações em $(0, 1)$, ou seja, substituir $\mathbb{I}_{\{c\}}(y)$ por zero para essas observações. Note que substituir $\mathbb{I}_{\{c\}}(y_i)$ por zero, no modelo de regressão beta inflacionado, significa que o valor de y , que se esperava ser zero, passa a ser um valor em $(0, 1)$. Geramos y_i a partir de um modelo de regressão beta com média $\mu_i = g_\mu^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$ e precisão $\phi_i = g_\phi^{-1}(\mathbf{z}_i^\top \boldsymbol{\gamma})$. Em seguida, aumentamos os valores das covariáveis \mathcal{S}_1 e \mathcal{S}_2 de modo que as observações mal classificadas fiquem em um hiperplano que é paralelo ao hiperplano discriminante, cuja distância entre eles seja igual a $1.5\sqrt{p_0} = 1.5\sqrt{3}$; veja a primeira linha de gráficos da Figura 6.3. Nesse tipo de contaminação, produzimos pontos de alavanca ruins, que interferem significativamente na estimação via máxima verossimilhança. Além disso, alguns pontos também são adicionados à parte contínua do modelo, mas eles são gerados de

Tabela 6.1: Estimativas dos parâmetros, com respectivos erros padrão, considerando diferentes métodos de estimação para os dados sem contaminação e contaminados de acordo com o Cenário 1.

Par.	MLE				M-SE				M-ME			
	Sem cont.		Com cont.		Sem cont.		Com cont.		Sem cont.		Com cont.	
	Est.	e.p.										
κ_1	0.19	0.27	0.19	0.27	0.19	0.27	0.19	0.27	0.19	0.27	0.19	0.27
κ_2	1.71	0.41	1.71	0.41	1.71	0.41	1.71	0.41	1.71	0.41	1.71	0.41
κ_3	1.64	0.38	1.64	0.38	1.64	0.38	1.64	0.38	1.64	0.38	1.64	0.38
β_1	-1.75	0.09	-1.76	0.21	-1.75	0.09	-1.75	0.10	-1.75	0.09	-1.75	0.10
β_2	-1.95	0.19	-1.00	0.38	-1.95	0.19	-1.96	0.20	-1.95	0.19	-1.96	0.20
γ_1	4.67	0.19	2.65	0.19	4.67	0.19	4.55	0.19	4.67	0.19	4.57	0.19

Par.	M-LSE				M-LME			
	Sem cont.		Com cont.		Sem cont.		Com cont.	
	Est.	e.p.	Est.	e.p.	Est.	e.p.	Est.	e.p.
κ_1	0.19	0.27	0.19	0.27	0.19	0.27	0.19	0.27
κ_2	1.71	0.41	1.71	0.41	1.71	0.41	1.71	0.41
κ_3	1.64	0.38	1.64	0.38	1.64	0.38	1.64	0.38
β_1	-1.75	0.09	-1.75	0.10	-1.75	0.09	-1.75	0.10
β_2	-1.95	0.19	-1.97	0.20	-1.95	0.19	-1.97	0.20
γ_1	4.67	0.19	4.59	0.19	4.67	0.19	4.59	0.19

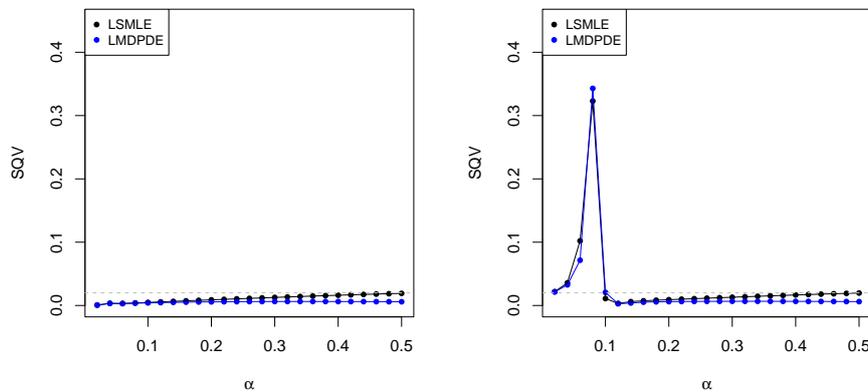


Figura 6.2: Gráfico dos SQVs referentes aos estimadores LSMLE e LMDPDE (parte contínua) para os dados sem contaminação (esquerda) e contaminados segundo o Cenário 1 (direita).

acordo com o modelo postulado (sem contaminação), então seguem a tendência apresentada pelas outras observações; veja a segunda linha de gráficos da Figura 6.3.

Os resultados dos ajustes são apresentados na Tabela 6.2. Observe que as estimativas dos parâmetros para todos os métodos coincidem quando consideramos os dados sem contaminação; o valor ótimo para α escolhido por todos eles foi zero. A presença de contaminação, no entanto, faz com que as estimativas de máxima verossimilhança dos parâmetros da parte discreta fiquem distorcidas. A estimativa de κ_2 , por exemplo, passa de 2.84 para 0.28. As estimativas dos parâmetros da parte contínua para os dados não contaminados são ligeiramente diferentes daquelas para os dados contaminados. Isso ocorre porque a contaminação da parte discreta, neste cenário, adiciona observações à parte contínua. No entanto, essa mudança pode ser considerada desprezível. Os estimadores robustos para os κ 's não são afetados pela presença de contaminação. Note que as estimativas dos κ 's via MDPDE para os dados contaminados são bem próximas das estimativas dos dados não contaminados. O valor ótimo para α selecionado para o MDPDE referente à parte discreta é 0.25. Para os estimadores associados à parte contínua, o valor ótimo de α selecionado é zero, como esperado.

Na Figura 6.4 apresentamos os valores dos SQVs referentes ao estimador MDPDE (parte discreta) para os dados sem contaminação e contaminados. Observe que, na ausência de contaminação, os SQVs estão todos abaixo do limiar. Na presença de contaminação, o SQV apresenta um pico para $\alpha = 0.2$ e a partir do próximo valor na grade, ou seja, 0.25, os valores de SQV se estabilizam, resultando neste valor para α ótimo.

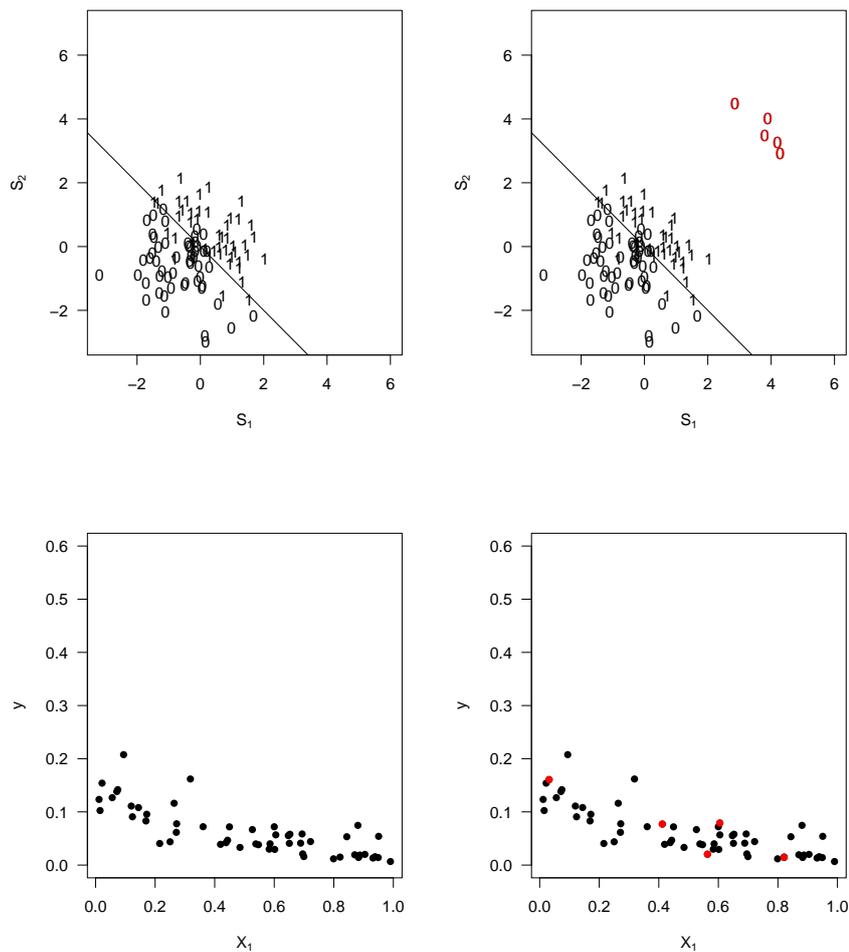


Figura 6.3: Diagramas de dispersão da amostra sem contaminação (esquerda) e da amostra contaminada segundo o Cenário 2 (direita) para a parte discreta (primeira linha) e contínua (segunda linha).

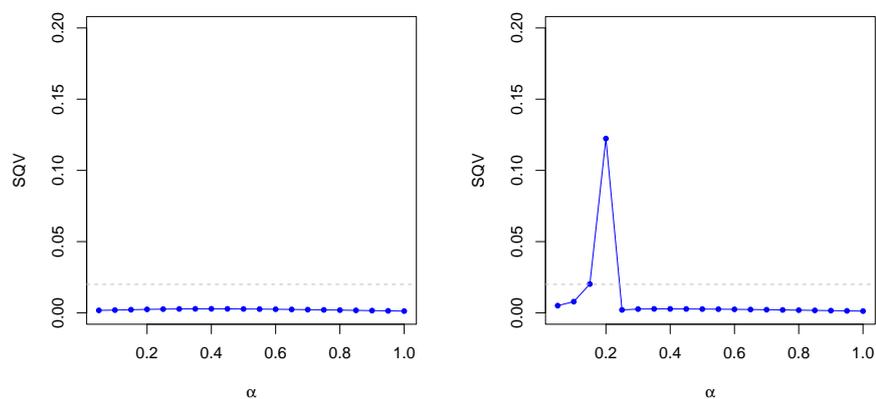


Figura 6.4: Gráfico dos SQVs referentes ao estimador MDPDE (parte discreta) para os dados sem contaminação (esquerda) e contaminados segundo o Cenário 2 (direita).

Cenário 3

Nesta aplicação, a contaminação é feita na parte contínua e discreta simultaneamente, de acordo com os cenários 1 e 2, respectivamente; veja Figura 6.5.

Tabela 6.2: Estimativas dos parâmetros, com respectivos erros padrão, considerando diferentes métodos de estimação para os dados sem contaminação e contaminados de acordo com o Cenário 2.

Par.	MLE				M-SE				M-ME			
	Sem cont.		Com cont.		Sem cont.		Com cont.		Sem cont.		Com cont.	
	Est.	e.p.										
κ_1	0.35	0.33	-0.31	0.21	0.35	0.33	0.25	0.32	0.35	0.33	0.25	0.32
κ_2	2.84	0.63	0.26	0.18	2.84	0.63	2.59	0.59	2.84	0.63	2.59	0.59
κ_3	2.63	0.59	0.28	0.18	2.63	0.59	2.41	0.56	2.63	0.59	2.41	0.56
β_1	-1.82	0.10	-1.80	0.10	-1.82	0.10	-1.80	0.10	-1.82	0.10	-1.80	0.10
β_2	-2.05	0.20	-2.10	0.22	-2.05	0.20	-2.10	0.22	-2.05	0.20	-2.10	0.22
γ_1	4.59	0.21	4.56	0.22	4.59	0.21	4.56	0.22	4.59	0.21	4.56	0.22

Par.	M-LSE				M-LME			
	Sem cont.		Com cont.		Sem cont.		Com cont.	
	Est.	e.p.	Est.	e.p.	Est.	e.p.	Est.	e.p.
κ_1	0.35	0.33	0.25	0.32	0.35	0.33	0.25	0.32
κ_2	2.84	0.63	2.59	0.59	2.84	0.63	2.59	0.59
κ_3	2.63	0.59	2.41	0.56	2.63	0.59	2.41	0.56
β_1	-1.82	0.10	-1.80	0.10	-1.82	0.10	-1.80	0.10
β_2	-2.05	0.20	-2.10	0.22	-2.05	0.20	-2.10	0.22
γ_1	4.59	0.21	4.56	0.22	4.59	0.21	4.56	0.22

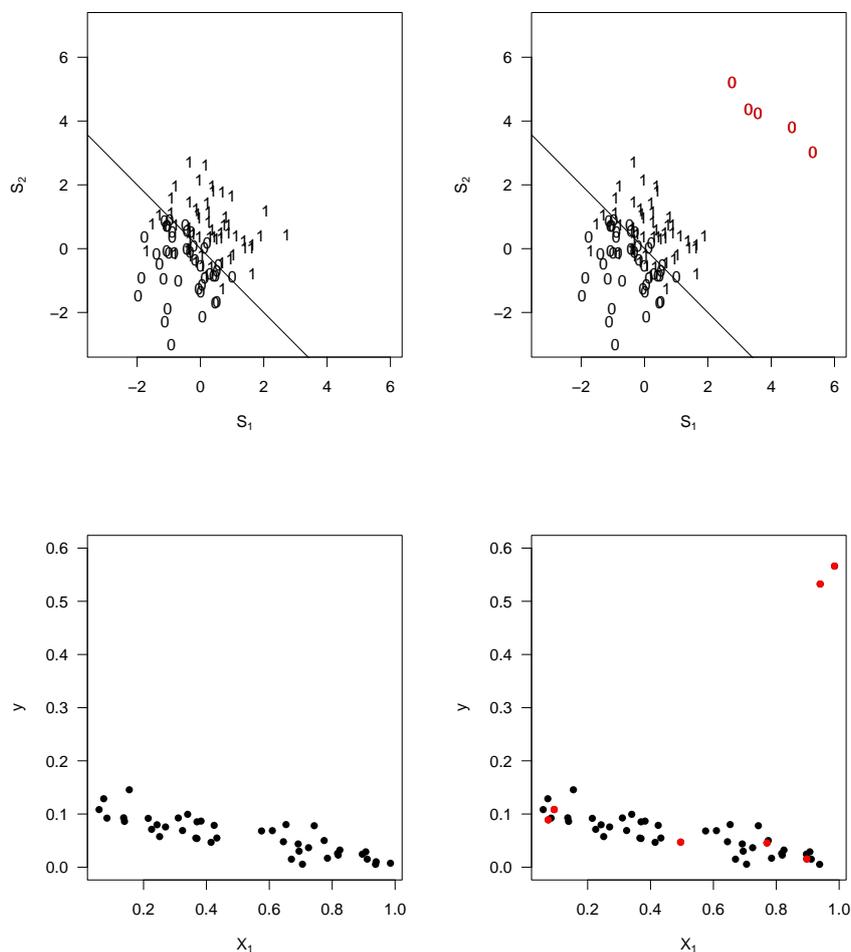


Figura 6.5: Diagramas de dispersão da amostra sem contaminação (esquerda) e da amostra contaminada segundo o Cenário 3 (direita) para a parte discreta (primeira linha) e contínua (segunda linha).

Os resultados dos ajustes considerando os diferentes estimadores são apresentados na Tabela 6.3. As conclusões são análogas às dos cenários anteriores. As estimativas para os parâmetros considerando os estimadores robustos são iguais às estimativas de máxima verossimilhança e o valor ótimo da constante α escolhido por todos eles é zero. No entanto, a contaminação afeta de forma considerável a estimação via máxima verossimilhança, ao contrário da estimação via métodos robustos, que não é influenciada pelas observações contaminadas (esse comportamento é observado nas partes discreta e contínua). Para os ajustes considerando os dados contaminados, os valores

ótimos obtidos para a constante de afinação para o MDPDE (parte discreta), SMLE, MDPDE, LSMLE e LMDPDE (parte contínua) são, respectivamente, 0.3, 0.06, 0.08, 0.1 e 0.1.

Tabela 6.3: Estimativas dos parâmetros, com respectivos erros padrão, considerando diferentes métodos de estimação para os dados sem contaminação e contaminados de acordo com o Cenário 3.

Par.	MLE				M-SE				M-ME			
	Sem cont.		Com cont.		Sem cont.		Com cont.		Sem cont.		Com cont.	
	Est.	e.p.										
κ_1	0.41	0.29	0.04	0.21	0.41	0.29	0.30	0.28	0.41	0.29	0.30	0.28
κ_2	1.74	0.46	-0.04	0.19	1.74	0.46	1.47	0.42	1.74	0.46	1.47	0.42
κ_3	2.08	0.47	0.35	0.18	2.08	0.47	1.75	0.42	2.08	0.47	1.75	0.42
β_1	-1.84	0.09	-1.89	0.22	-1.84	0.09	-1.90	0.09	-1.84	0.09	-1.90	0.09
β_2	-2.08	0.19	-0.94	0.38	-2.08	0.19	-1.91	0.18	-2.08	0.19	-1.91	0.18
γ_1	4.80	0.19	2.68	0.21	4.80	0.19	4.84	0.19	4.80	0.19	4.87	0.19

Par.	M-LSE				M-LME			
	Sem cont.		Com cont.		Sem cont.		Com cont.	
	Est.	e.p.	Est.	e.p.	Est.	e.p.	Est.	e.p.
κ_1	0.41	0.29	0.30	0.28	0.41	0.29	0.30	0.28
κ_2	1.74	0.46	1.47	0.42	1.74	0.46	1.47	0.42
κ_3	2.08	0.47	1.75	0.42	2.08	0.47	1.75	0.42
β_1	-1.84	0.09	-1.92	0.08	-1.84	0.09	-1.92	0.08
β_2	-2.08	0.19	-1.88	0.18	-2.08	0.19	-1.88	0.18
γ_1	4.80	0.19	4.93	0.19	4.80	0.19	4.92	0.19

Os valores dos SQVs referentes ao estimador MDPDE (parte discreta) e LSMLE e MDPDE (parte contínua) para os dados sem contaminação e contaminados são apresentados na Figura 6.6.

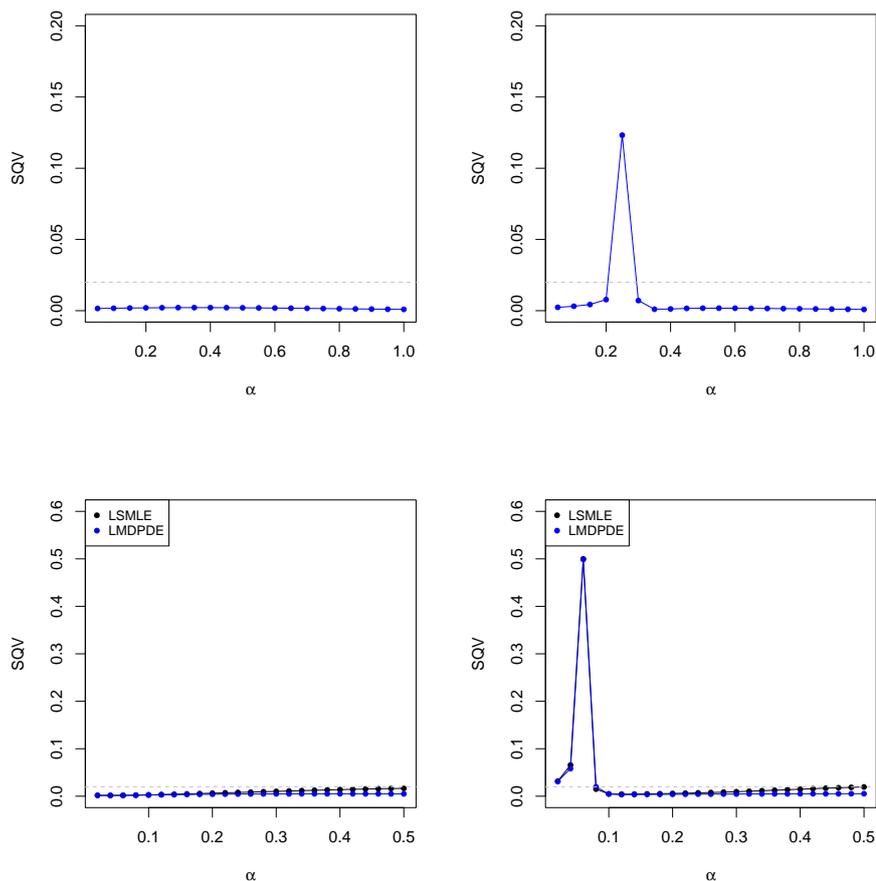


Figura 6.6: Gráfico dos SQVs referentes ao estimador MDPDE (parte discreta, primeira linha) e aos estimadores LSMLE e LMDPDE (parte contínua, segunda linha) para os dados sem contaminação (esquerda) e contaminados segundo o Cenário 3 (direita).

Para os ajustes referentes aos dados não contaminados, os SQVs estão abaixo do limiar. Na presença de contaminação, os SQVs referentes ao MDPDE (parte discreta) apresentam valores baixos, porém com um pico em 0.25. A partir do valor ótimo escolhido 0.3 as estimativas se

estabilizam. O comportamento dos SQVs para a parte contínua é similar ao da parte discreta, no entanto, já observamos instabilidade nas estimativas desde valores de α próximos de zero. Os SQVs e, conseqüentemente, as estimativas, se estabilizam a partir do α ótimo escolhido.

6.4 Resultados numéricos

No que segue, estamos interessados em avaliar o desempenho dos estimadores propostos em amostras de tamanho finito. Para isso, consideramos os três cenários apresentados na aplicação em dados simulados da Seção 6.3 com dois tamanhos de amostra, $n = 100$ e $n = 200$, e 3000 réplicas Monte Carlo. Assim como nas aplicações em dados simulados, os valores das constantes de afinação são escolhidos através do algoritmo descrito na Seção 6.2.7.

Os boxplots das estimativas dos parâmetros são apresentados nas Figuras 6.7-6.8, 6.9-6.10 e 6.11-6.12 para os cenários 1, 2 e 3, respectivamente, considerando os estimadores MLE e MDPDE para a parte discreta e MLE, MDPDE e SMLE para a parte contínua. Boxplots dos valores ótimos da constante de afinação são apresentados na Figura 6.13. Os correspondentes boxplots considerando os estimadores LMDPDE e LSMLE para a parte contínua estão disponíveis no Apêndice E.2 e são similares àqueles do MDPDE e do SMLE.

Assim como observado nas aplicações em dados simulados, quando há contaminação apenas na parte contínua, Cenário 1, somente as estimativas de máxima verossimilhança dos β 's e γ são afetadas; veja a Figura 6.8. Os MLEs de β e γ são fortemente afetados pela presença de contaminação, apresentando viés severo. As estimativas de máxima verossimilhança para β_2 na presença de contaminação, por exemplo, estão centradas em torno de -0.88 ($n = 100$) e o valor correto para β_2 é -2.0 . As estimativas de máxima verossimilhança de γ são as mais afetadas; na presença de contaminação estão centradas em torno de 2.6 , enquanto o valor verdadeiro é 4.5 . Os estimadores robustos, por outro lado, apresentam desempenho satisfatório na presença e na ausência de contaminação. Na ausência de contaminação, o desempenho do SMLE e do MDPDE é similar ao do MLE; isso se dá pelo algoritmo utilizado para a seleção da constante de afinação α que retorna $\alpha = 0$ (MLE) quando não identifica presença de observações atípicas nos dados. Na presença de contaminação, esses estimadores se comportam de forma parecida com o MLE na ausência de contaminação; ou seja, eles não são influenciados pelos *outliers*. O MDPDE e o SMLE apresentam comportamentos similares e estimativas centradas no verdadeiro valor do parâmetro (para os dados contaminados e sem contaminação). Conforme esperado, o MLE dos κ 's não é influenciado pela contaminação feita na parte contínua; veja a Figura 6.7. Além disso, os boxplots das estimativas via MLE e MDPDE para os κ 's são bem parecidos, senão idênticos, indicando que o algoritmo de seleção de α identificou corretamente que não há observações atípicas e escolheu $\alpha = 0$.

No Cenário 2, a contaminação é feita apenas na parte discreta e, conseqüentemente, apenas o MLE para os κ 's são afetados na presença de contaminação; veja as Figuras 6.9 e 6.10. Na presença de contaminação, as estimativas de máxima verossimilhança para os κ 's ficam bem distantes do valor verdadeiro. As estimativas de κ_3 , por exemplo, ficam centradas em torno de 0.15 ($n = 200$), na presença de contaminação, enquanto o valor verdadeiro é 2.0 . Observe que, quando há contaminação, a variabilidade das estimativas de máxima verossimilhança para κ_2 e κ_3 é bem pequena; os desvios padrão das estimativas de κ_2 e κ_3 , para $n = 100$, são, respectivamente, 0.10 e 0.12 . Esse comportamento é esperado. Conforme comentado na Seção 6.1.2, Croux *et al.* (2002) mostram que, na presença de contaminação, o MLE para os κ 's (exceto o intercepto) tende a zero. O estimador robusto (MDPDE), por outro lado, apresenta estimativas razoáveis tanto para os dados sem contaminação quanto para os dados contaminados. Na ausência de contaminação, o algoritmo escolhe $\alpha = 0$ na maioria dos casos. Na presença de contaminação, o MDPDE se comporta de forma parecida com o MLE na ausência de contaminação, indicando que as observações atípicas não influenciam o método de estimação.

No Cenário 3, a contaminação é feita na parte contínua e na parte discreta. Neste cenário, as estimativas de máxima verossimilhança para todos os parâmetros do modelo são fortemente

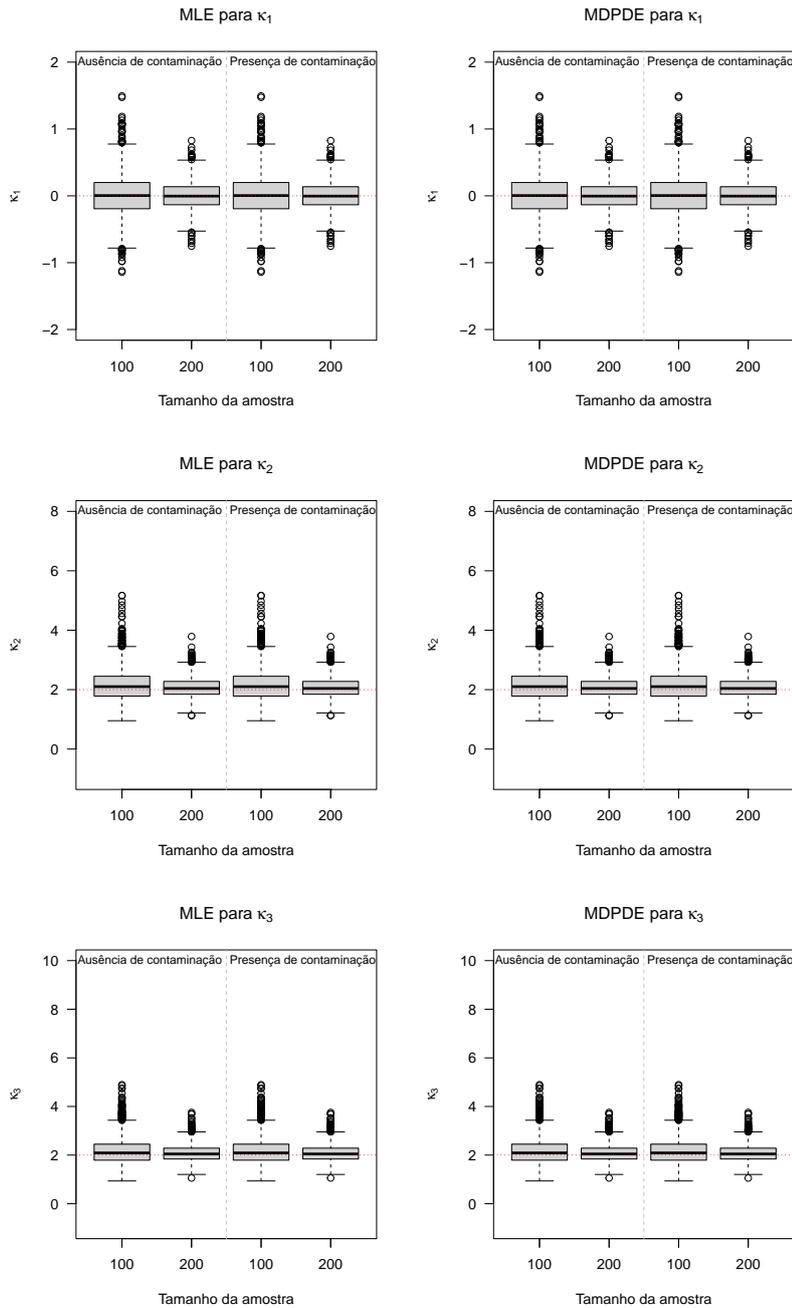


Figura 6.7: *Boxplots das estimativas dos parâmetros κ_1 , κ_2 e κ_3 sob o cenário 1: MLE (esquerda) e MDPDE (direita).*

influenciadas pelas observações atípicas. O desempenho dos estimadores é parecido com o observado nos Cenários 1 e 2: na ausência de contaminação, o algoritmo para a escolha da constante de afinação seleciona $\alpha = 0$ na maioria dos casos e as estimativas são próximas das de máxima verossimilhança; na presença de contaminação, os estimadores robustos se comportam como o MLE na ausência de contaminação e apresentam estimativas centradas em torno do valor verdadeiro.

Os resultados de simulação indicam que os estimadores propostos são mais indicados na presença de observações atípicas e que a estimação via máxima verossimilhança, neste caso, não é confiável. Em geral, na presença de contaminação, a robustez obtida através dos estimadores propostos tem como custo um aumento na variabilidade, principalmente em amostras de tamanho pequeno ($n = 100$).

O algoritmo para a escolha da constante de afinação dos estimadores robustos apresentou

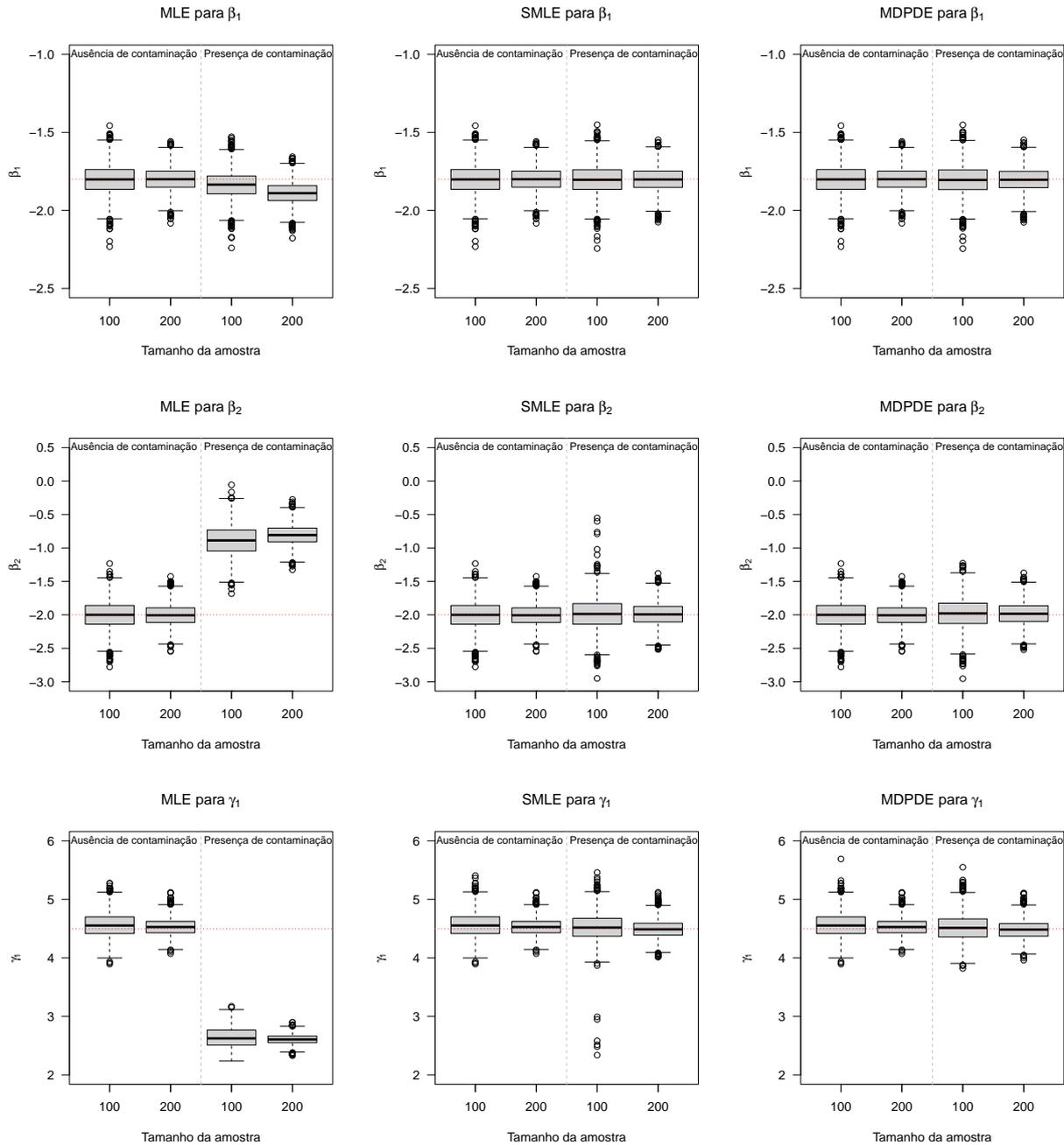


Figura 6.8: *Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 1: MLE (esquerda), SMLE (centro) e MDPDE (direita).*

desempenho excelente. Os boxplots dos valores ótimos escolhidos são apresentados na Figura 6.13. Note que, para o Cenário 1 (contaminação apenas na parte contínua), os valores ótimos de α escolhidos para o MDPDE para os κ 's são iguais a zero para todas as amostras (com e sem contaminação). Já para o SMLE e para o MDPDE, o algoritmo selecionou $\alpha = 0$ na maioria dos casos na ausência de contaminação e em torno de 0.1 para os dados contaminados. Analogamente, no Cenário 2 (contaminação apenas na parte discreta), o algoritmo seleciona $\alpha = 0$ para o MDPDE na ausência de contaminação e em torno de 0.3 na presença de contaminação. Já para o SMLE e MDPDE, $\alpha = 0$ é escolhido na grande maioria dos casos. Finalmente, no Cenário 3, a escolha de α é feita de forma análoga. Nos três cenários, o algoritmo consegue identificar a necessidade de utilizar um método robusto quando há contaminação nos dados e seleciona o MLE quando não há contaminação. Isso é importante em procedimentos robustos para se obter um balanço entre

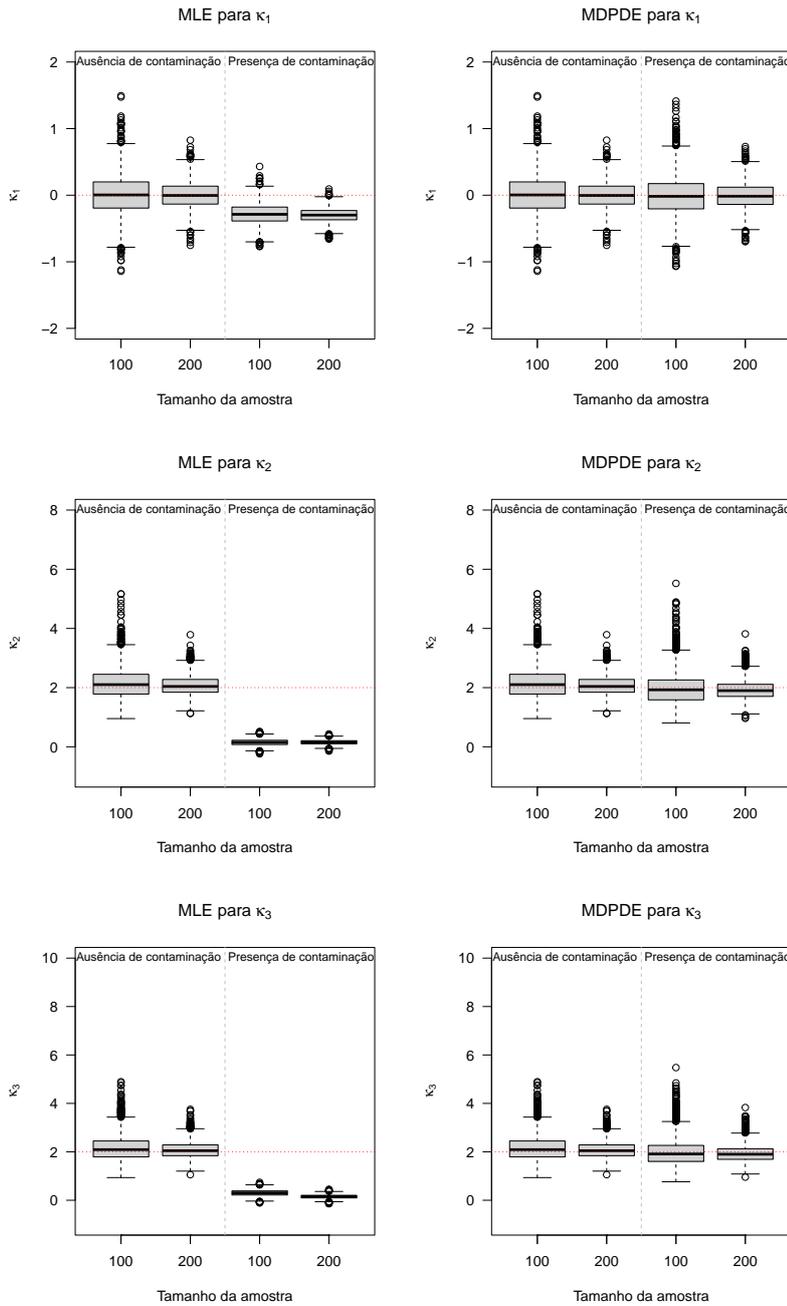


Figura 6.9: *Boxplots das estimativas dos parâmetros κ_1 , κ_2 e κ_3 sob o cenário 2: MLE (esquerda) e MDPDE (direita).*

eficiência e robustez e obter eficiência máxima (MLE) quando não há observações atípicas.

Resultados similares podem ser observados para o LMDPDE e o LSMLE nas Figuras E.1-E.4 (Apêndice E.2).

É importante mencionar que o desempenho desse algoritmo na escolha da constante de afinação para o modelo beta já foi estudado em Ribeiro e Ferrari (2022) (ver também Maluf *et al.* (2022)). No entanto, até o momento, não conhecemos nenhum trabalho que tenha implementado e estudado o desempenho de um algoritmo orientado pelos dados para a escolha da constante de afinação em regressão para dados binários, conforme feito aqui.

No que segue, fazemos um estudo a respeito da eficiência dos estimadores propostos na presença e na ausência de contaminação. Para tanto, utilizamos os erros quadráticos médios totais (EQMT) para comparar a eficiência dos estimadores. Na Tabela 6.4, apresentamos a razão entre os EQMTs

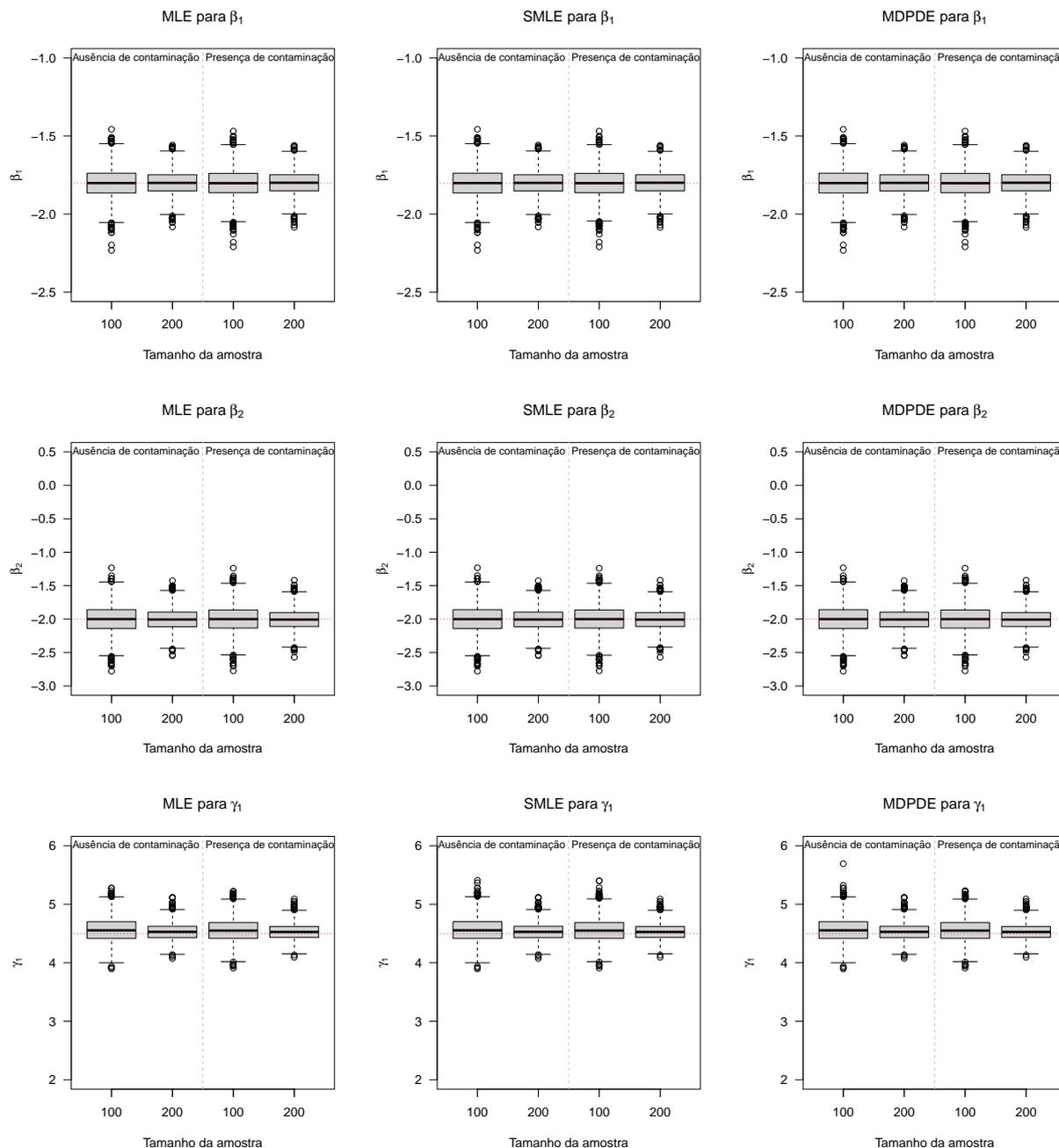


Figura 6.10: *Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 2: MLE (esquerda), SMLE (centro) e MDPDE (direita).*

do MLE, M-ME, M-LME, M-SE e M-LSE sob os cenários 1, 2 e 3. Na ausência de contaminação, notamos que as razões entre os EQMTs de todos os estimadores são iguais a um. Já para os dados contaminados, os EQMTs dos estimadores robustos são consideravelmente menores que o do MLE, principalmente em amostras de tamanho moderado/grande ($n = 200$). Note que para $n = 200$, Cenário 3, o EQMT do MLE é, aproximadamente, 34 vezes maior que o do SMLE. Os estimadores M-ME, M-LME, M-SE e M-LSE têm comportamentos similares com relação ao EQMT; as razões entre os EQMTs desses estimadores são igual ou próxima de um para todos os cenários.

Finalmente, estudamos o desempenho do teste tipo-Wald robusto, proposto na Seção 6.2.6, para testar $\mathcal{H}_0 : \Upsilon_j = \Upsilon_j^{(0)}$ contra a alternativa $\mathcal{H}_1 : \Upsilon_j \neq \Upsilon_j^{(0)}$ e rejeita \mathcal{H}_0 ao nível nominal ζ quando a estatística W_α excede o quantil de ordem $1 - \zeta$ da distribuição χ_1^2 . Fixamos $\zeta = 5\%$ e calculamos os níveis empíricos dos teste tipo-Wald robusto sob \mathcal{H}_0 baseados no MLE (teste Wald

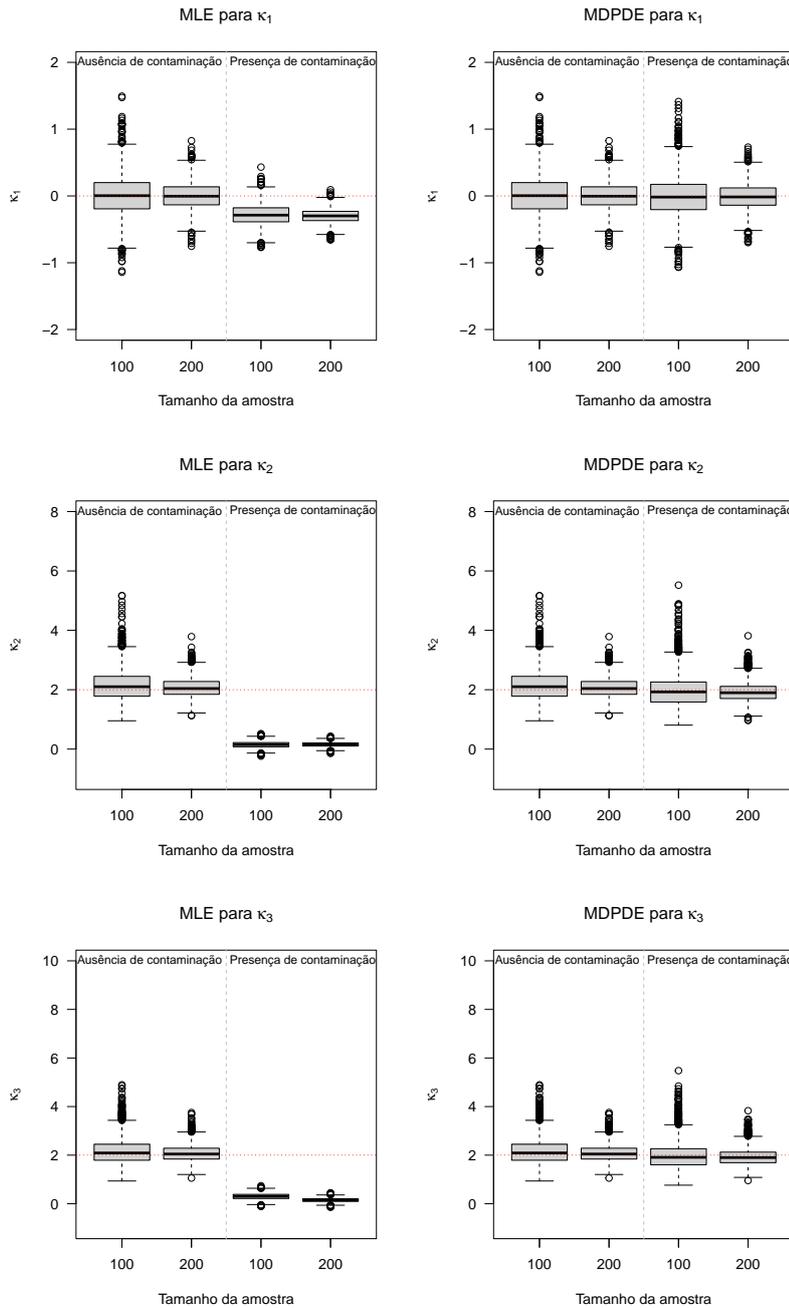


Figura 6.11: *Boxplots das estimativas dos parâmetros κ_1 , κ_2 e κ_3 sob o cenário 3: MLE (esquerda) e MDPDE (direita).*

usual) e nos estimadores robustos propostos. Esse procedimento foi feito para os Cenários 1, 2 e 3 e os valores escolhidos para $\Upsilon_j^{(0)}$ são aqueles fixados nos respectivos cenários. Os resultados estão apresentados na Tabela 6.5.

De maneira geral, podemos observar que os testes Wald usual e robustos apresentam comportamentos similares na ausência de contaminação. Os níveis empíricos tendem a ser mais próximos do nível nominal (5%) quando o tamanho da amostra é moderado/grande ($n = 200$). Por outro lado, o desempenho do teste Wald usual é bastante comprometido na presença de contaminação nos dados. Na maioria dos casos, observamos probabilidades do erro do tipo I iguais a um ou a zero; veja os testes Wald usual para κ_2 e para β_1 nos cenários 2 e 1, respectivamente, para $n = 100$ e dados contaminados. Isso indica que, na presença de observações atípicas, o teste Wald deve ser evitado. Conforme esperado, esse mau desempenho é observado apenas nos testes associados

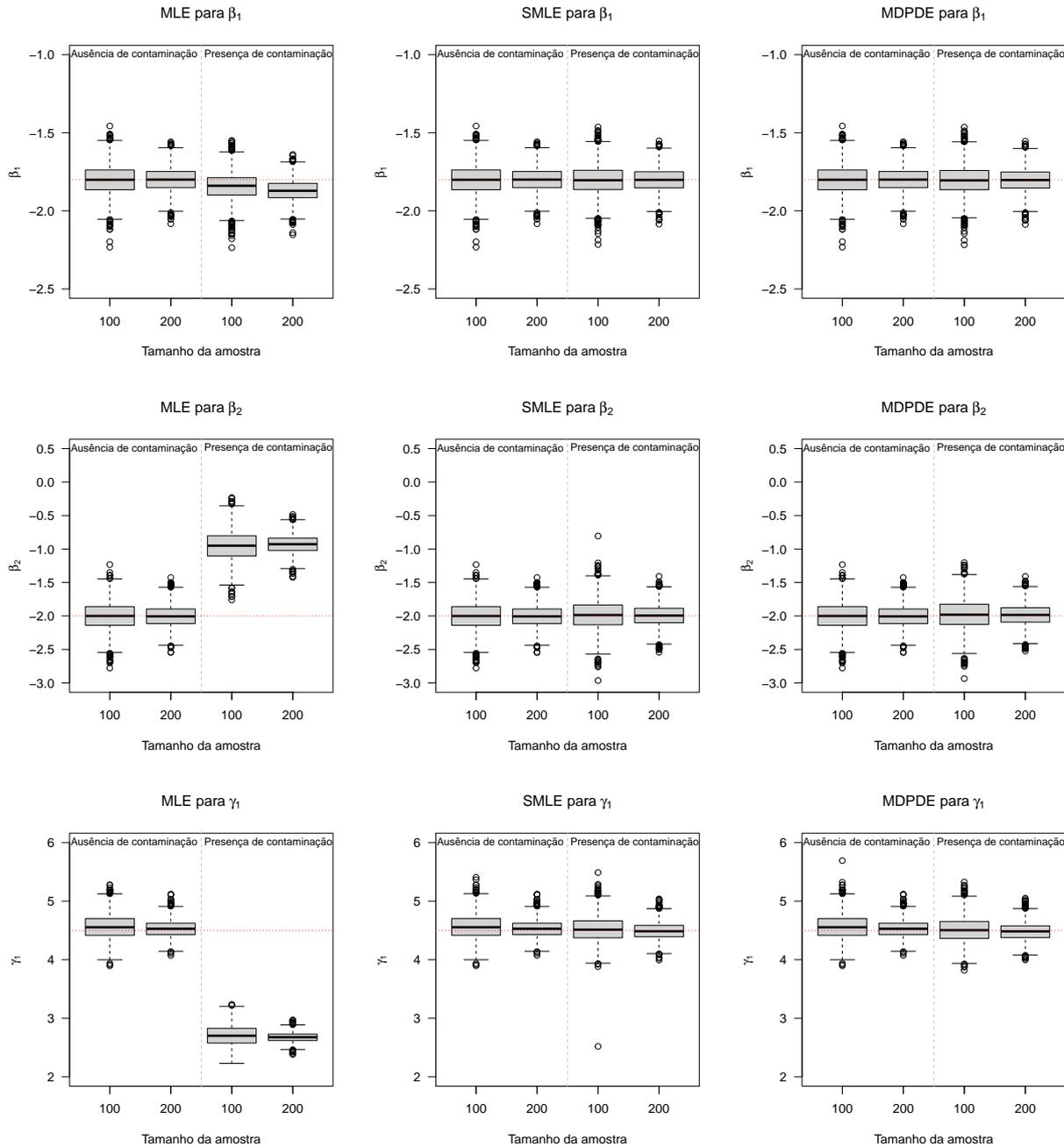


Figura 6.12: *Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 3: MLE (esquerda), SMLE (centro) e MDPDE (direita).*

aos parâmetros da parte que foi contaminada. Por exemplo, no Cenário 1, na presença de contaminação, o desempenho do teste Wald usual é satisfatório para os parâmetros associados à parte discreta (κ 's), o contrário do que acontece para os parâmetros associados à parte contínua (β 's e γ_1). Neste cenário, os níveis empíricos para κ_1 , κ_2 e κ_3 , para $n = 100$, são iguais a 4%, enquanto que os de β_1 , β_2 e γ_1 são 0%, 95% e 100%, respectivamente. De forma análoga, no Cenário 2, apenas os testes Wald usuais associados aos parâmetros da parte discreta são comprometidos e, no Cenário 3, os testes Wald usuais associados a todos os parâmetros são comprometidos, uma vez que a contaminação é feita na parte discreta e contínua.

Os testes tipo-Wald robustos, no entanto, apresentam desempenhos razoáveis tanto na ausência como na presença de contaminação. Na maioria dos casos, quando há contaminação, os níveis empíricos dos testes robustos apresentam-se próximos dos níveis nominais. Veja, por exemplo, no

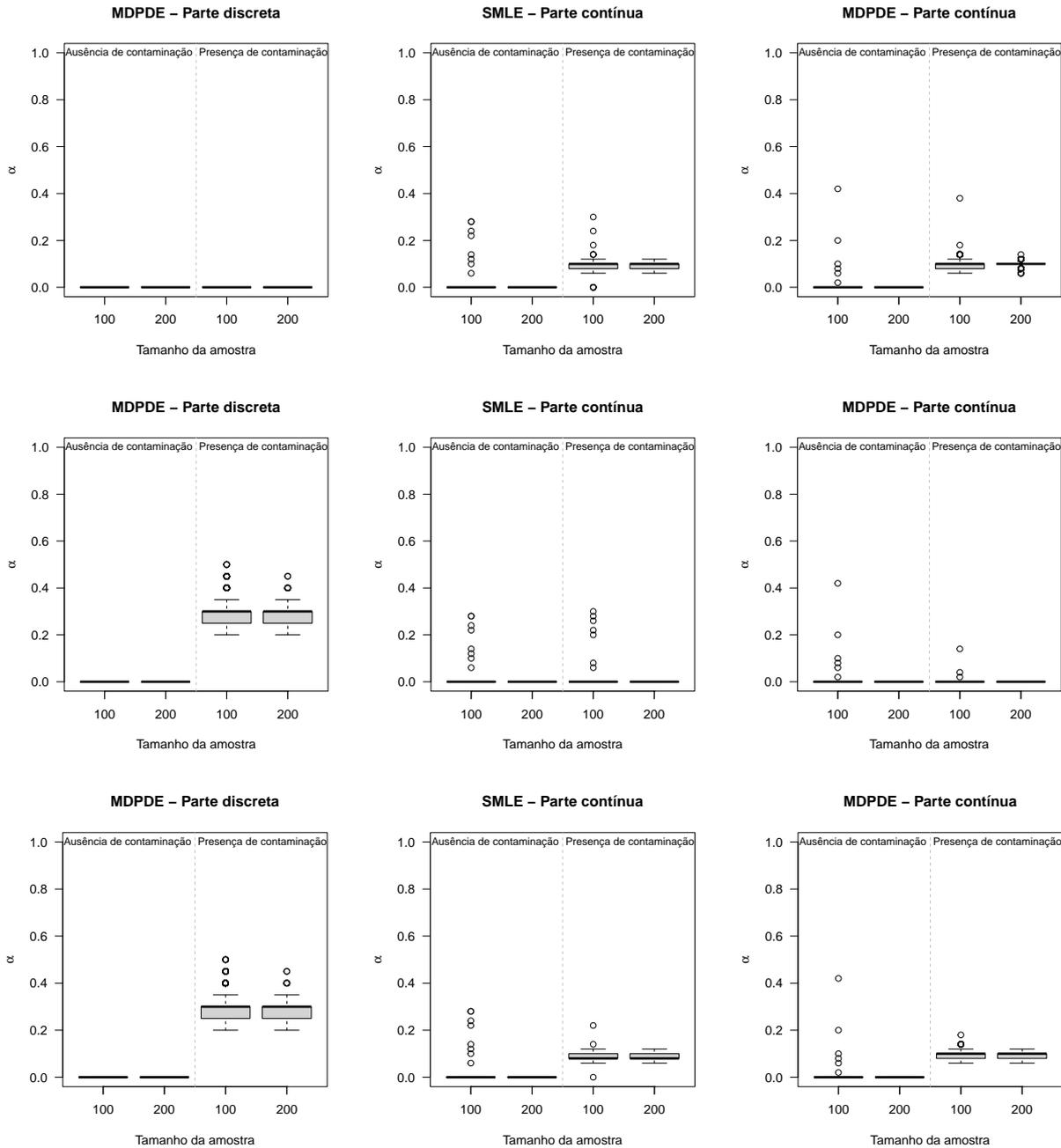


Figura 6.13: *Boxplots dos valores ótimos para a constante de afinação α para a parte discreta considerando o MDPDE e para a parte contínua considerando o LSMLE e o LMDPDE, sob os cenários 1 (primeira linha), 2 (segunda linha) e 3 (terceira linha).*

Cenário 3, para os dados contaminados, os testes para β_1 , β_2 e γ_1 , considerando o MDPDE e $n = 200$, apresentam níveis empíricos iguais a 5%, 5% e 7%, respectivamente. Em geral, quando comparamos os testes associados aos parâmetros da parte contínua para os diferentes estimadores propostos, não observamos uma superioridade de um com relação aos outros. Todos apresentam desempenhos parecidos. No entanto, em geral, a presença de contaminação nos dados faz com que as probabilidades do erro do tipo I dos testes tipo-Wald robustos fiquem maiores do que o esperado. Assim, ainda que esses testes sejam mais indicados do que o teste Wald usual, é preciso ter cuidado na tomada de decisão.

Nas Figuras 6.14 e 6.15 apresentamos as curvas das discrepâncias quantílicas relativas das es-

Tabela 6.4: Razão entre os erros quadráticos médios totais dos estimadores sob os Cenários 1, 2 e 3.

n	Cenário 1				Cenário 2				Cenário 3			
	Sem cont.		Com cont.		Sem cont.		Com cont.		Sem cont.		Com cont.	
	100	200	100	200	100	200	100	200	100	200	100	200
$\frac{MLE}{M-SE}$	1.00	1.00	6.69	15.89	1.00	1.00	7.93	20.93	1.00	1.00	12.96	33.73
$\frac{MLE}{M-ME}$	1.00	1.00	6.76	15.83	1.00	1.00	7.94	20.93	1.00	1.00	12.97	33.61
$\frac{MLE}{M-LSE}$	1.00	1.00	6.63	15.90	1.00	1.00	7.93	20.93	1.00	1.00	12.88	33.75
$\frac{MLE}{M-LME}$	1.00	1.00	6.77	15.89	1.00	1.00	7.94	20.93	1.00	1.00	12.98	33.72
$\frac{M-SE}{M-ME}$	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\frac{M-SE}{M-LSE}$	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
$\frac{M-SE}{M-LME}$	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\frac{M-ME}{M-LSE}$	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
$\frac{M-ME}{M-LME}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\frac{M-LSE}{M-LME}$	1.00	1.00	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00

tatísticas usuais e robustas ² (considerando o SMLE e o MDPDE) para todos os parâmetros do modelo considerando os dados contaminados e sem contaminação para o Cenário 3. Essa discrepância é definida como sendo a diferença entre o quantil exato da estatística de teste (obtido por simulação) e o quantil (assintótico) teórico, dividido pelo quantil teórico. Os quantis teóricos são obtidos através da distribuição qui-quadrado com 1 grau de liberdade. Quanto mais próximo de zero a discrepância quantílica relativa estiver, teremos maior indício que o limite em distribuição das estatísticas de teste sob a hipótese nula é qui-quadrado com 1 grau de liberdade. Quando não há contaminação, a discrepância quantílica relativa é próxima de zero para todas as estatísticas (usuais e robustas). Na presença de contaminação, as discrepâncias quantílicas relativas para as estatísticas usuais são distantes de zero para todos os testes (referentes aos parâmetros da parte discreta e contínua). Esse comportamento é observado para $n = 100$ e $n = 200$. Por exemplo, para $n = 200$, a discrepância quantílica relativa da estatística de teste usual para testar $\mathcal{H}_0 : \gamma_1 = 4.5$, na presença de contaminação, apresenta valores superiores a 1000. Esse comportamento indica que, na presença de contaminação, a aproximação das estatísticas usuais pela distribuição χ_1^2 pode não ser indicada. As discrepâncias quantílicas relativas das estatísticas robustas apresentam valores próximos de zero tanto na presença quanto na ausência de contaminação. Desta forma, é razoável assumir que, no contexto das simulações, a distribuição qui-quadrado com um grau de liberdade é uma boa aproximação para a distribuição destas estatísticas.

Em resumo, os estudos de simulação revelaram que a estimação via máxima verossimilhança no modelo de regressão beta inflacionado em zero ou um sofre forte influência de observações atípicas nos dados. Os desempenhos dos estimadores propostos neste trabalho são similares ao do estimador de máxima verossimilhança na ausência de contaminação; na presença de contaminação, apresentaram uma superioridade considerável. O método utilizado para a escolha da constante de afinação mostrou-se eficaz, identificando a necessidade de introduzir robustez no método de estimação e retornando o MLE na ausência de contaminação. Além disso, na presença de contaminação, os testes robustos apresentaram melhor desempenho que os testes usuais, retornando níveis empíricos próximos dos níveis nominais e indicando que a aproximação das estatísticas de testes pela χ_1^2 é razoável. Adicionalmente, o teste Wald usual pode apresentar resultados distorcidos e não é recomendado quando há observações atípicas nos dados.

²Chamamos de estatísticas usuais aquelas referentes aos testes Wald usuais; isto é, que são baseadas no MLE. De forma análoga, as estatísticas robustas são aquelas referentes aos testes tipo-Wald robustos; isto é, testes baseados nos estimadores robustos.

Tabela 6.5: Níveis empíricos sob \mathcal{H}_0 do teste tipo-Wald robusto sob os Cenários 1, 2 e 3 ao nível nominal de 5%.

Cenário 1												
$n = 100$		Sem cont.					Com cont.					
Estimador	κ_1	κ_2	κ_3	β_1	β_2	γ_1	κ_1	κ_2	κ_3	β_1	β_2	γ_1
MLE	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.04	0.04	0.00	0.95	1.00
M-SE	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.04	0.04	0.06	0.07	0.08
M-ME	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.04	0.04	0.06	0.07	0.09
M-LSE	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.04	0.04	0.06	0.08	0.08
M-LME	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.04	0.04	0.06	0.07	0.08
$n = 200$		Sem cont.					Com cont.					
Estimador	κ_1	κ_2	κ_3	β_1	β_2	γ_1	κ_1	κ_2	κ_3	β_1	β_2	γ_1
MLE	0.05	0.04	0.04	0.06	0.05	0.06	0.05	0.04	0.04	0.00	1.00	1.00
M-SE	0.05	0.04	0.04	0.06	0.05	0.06	0.05	0.04	0.04	0.06	0.07	0.07
M-ME	0.05	0.04	0.04	0.06	0.05	0.06	0.05	0.04	0.04	0.06	0.07	0.07
M-LSE	0.05	0.04	0.04	0.06	0.05	0.06	0.05	0.04	0.04	0.06	0.07	0.07
M-LME	0.05	0.04	0.04	0.06	0.05	0.06	0.05	0.04	0.04	0.06	0.07	0.06
Cenário 2												
$n = 100$		Sem cont.					Com cont.					
Estimador	κ_1	κ_2	κ_3	β_1	β_2	γ_1	κ_1	κ_2	κ_3	β_1	β_2	γ_1
MLE	0.04	0.04	0.04	0.06	0.06	0.07	0.20	1.00	1.00	0.06	0.05	0.07
M-SE	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.05	0.06
M-ME	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.05	0.06
M-LSE	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.05	0.06
M-LME	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.05	0.06
$n = 200$		Sem cont.					Com cont.					
Estimador	κ_1	κ_2	κ_3	β_1	β_2	γ_1	κ_1	κ_2	κ_3	β_1	β_2	γ_1
MLE	0.05	0.04	0.04	0.06	0.05	0.06	0.54	1.00	1.00	0.06	0.05	0.06
M-SE	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.05	0.04	0.05
M-ME	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.05	0.04	0.05
M-LSE	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.05	0.04	0.05
M-LME	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.05	0.04	0.05
Cenário 3												
$n = 100$		Sem cont.					Com cont.					
Estimador	κ_1	κ_2	κ_3	β_1	β_2	γ_1	κ_1	κ_2	κ_3	β_1	β_2	γ_1
MLE	0.04	0.04	0.04	0.06	0.06	0.07	0.20	1.00	1.00	0.00	0.93	1.00
M-SE	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.06	0.07
M-ME	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.06	0.07
M-LSE	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.07	0.07
M-LME	0.04	0.04	0.04	0.06	0.06	0.07	0.04	0.09	0.08	0.06	0.06	0.07
$n = 200$		Sem cont.					Com cont.					
Estimador	κ_1	κ_2	κ_3	β_1	β_2	γ_1	κ_1	κ_2	κ_3	β_1	β_2	γ_1
MLE	0.05	0.04	0.04	0.06	0.05	0.06	0.54	1.00	1.00	0.00	1.00	1.00
M-SE	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.05	0.05	0.06
M-ME	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.05	0.05	0.07
M-LSE	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.06	0.06	0.06
M-LME	0.05	0.04	0.04	0.06	0.05	0.06	0.04	0.09	0.10	0.05	0.06	0.06

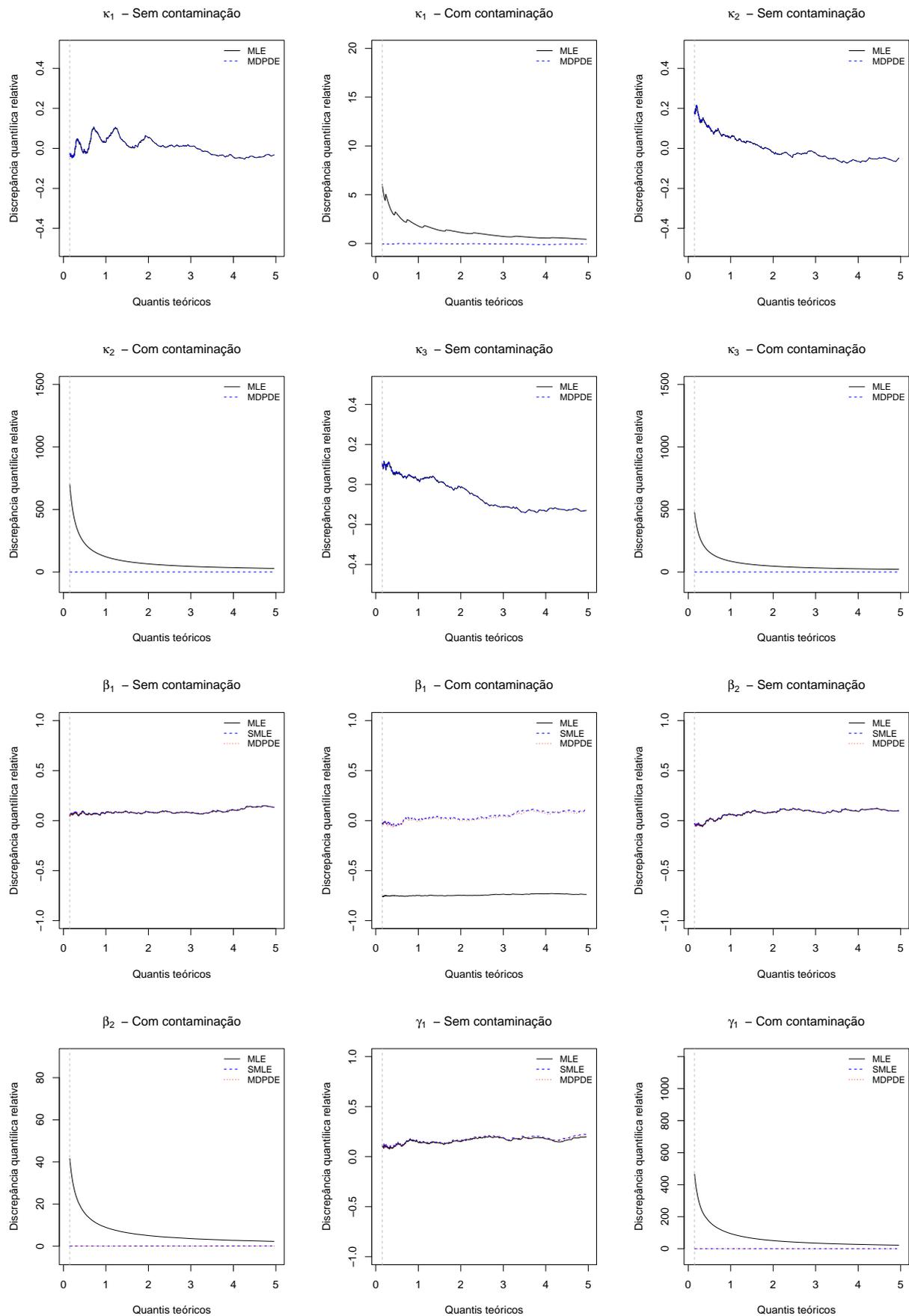


Figura 6.14: Discrepância quantílica relativa para o teste tipo-Wald usual e robusto; Cenário 3 e $n = 100$.

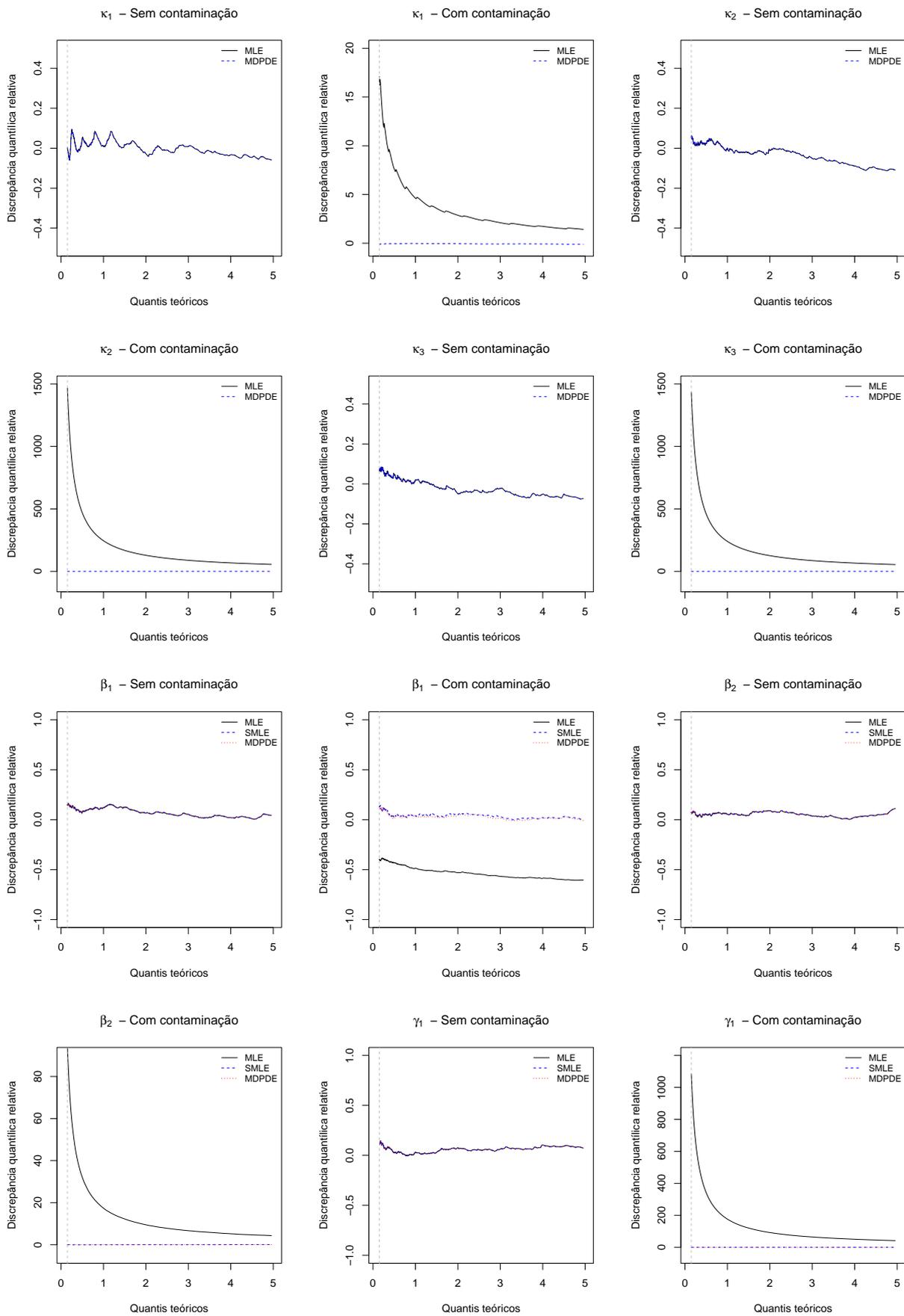


Figura 6.15: Discrepância quantílica relativa para o teste tipo-Wald usual e robusto; Cenário 3 e $n = 200$.

Capítulo 7

Discussões e pesquisas futuras

Nesta tese, apresentamos uma nova classe de modelos para dados com suporte em um intervalo limitado. Essa nova classe, denominada classe *power logit*, é definida a partir de uma transformação aplicada a uma variável aleatória com distribuição simétrica padrão. As distribuições *power logit* generalizam as distribuições GJS ao custo de introduzir um parâmetro adicional. As aplicações na Seção 2.5 revelam que as distribuições GJS não são suficientemente flexíveis para produzir um ajuste adequado aos dados; o uso das distribuições *power logit* correspondentes melhora consideravelmente os ajustes. Uma característica interessante das distribuições *power logit* é que seus três parâmetros são interpretáveis em termos da mediana, dispersão e assimetria da variável resposta. Além disso, essas distribuições tendem a ser naturalmente mais flexíveis do que as distribuições de dois parâmetros, como as distribuições beta, simplex e Kumaraswamy. As distribuições *power logit* podem depender de um parâmetro extra que indexa a distribuição simétrica subjacente. O parâmetro extra, se houver, pode ser visto como uma constante de afinação que é escolhida para promover ainda mais flexibilidade para ajustar adequadamente diferentes tipos de dados, particularmente com observações próximas dos extremos do intervalo.

Modelos de regressão, em que se supõe que a variável resposta segue uma distribuição na classe de *power logit*, são apresentados. As aplicações em dados reais nas Seções 3.6 e 4.4 sugerem que os modelos propostos são úteis para modelar dados contínuos limitados. Neste trabalho, oferecemos um amplo conjunto de ferramentas para inferência via máxima verossimilhança, diagnóstico e influência nesses modelos. Três tipos de resíduos são definidos. Chamamos a atenção para o resíduo padronizado, que pode destacar mais claramente as observações de alavanca do que os resíduos deviance e quantílico. As aplicações em dados simulados na Seção 3.3.1 mostram que os resíduos propostos podem identificar diferentes desvios do modelo postulado e detectar a presença de observações atípicas. Além disso, medidas de ajuste, como AIC e Υ , juntamente com gráficos de resíduos com envelopes simulados podem ser empregadas para selecionar a função geradora de densidades $r(\cdot)$. Adicionalmente, o modelo de regressão log-log, que é o modelo limite do modelo de regressão *power logit* quando o parâmetro de assimetria λ tende para zero, é desenvolvido e pode ser considerado como uma alternativa parcimoniosa aos modelos de regressão *power logit* quando a estimativa de λ é próxima de zero.

O pacote `PLreg`, implementado no *software* R, permite que os usuários empreguem os modelos de regressão *power logit* em suas próprias análises de dados. O pacote é desenvolvido nesta tese e fornece uma ampla gama de recursos para ajustar os modelos de regressão *power logit*, log-log e GJS, bem como para a realização de análises de diagnóstico e influência. Além disso, `PLreg` fornece procedimentos para escolha do parâmetro extra quando necessário. Diversos métodos, como o `summary`, `plot` e `AIC`, estão disponíveis para o pacote. `PLreg` está disponível no CRAN, o repositório oficial do *software* R. No Capítulo 4 esse pacote é apresentado e diversos exemplos são conduzidos a fim de exemplificar as suas principais funções.

O modelo de regressão *power logit* inflacionado em zero ou um é desenvolvido no Capítulo 5. Esse modelo é uma extensão natural do modelo de regressão *power logit* e é útil quando a variável de interesse não assume apenas valores no intervalo $(0, 1)$ mas, também, os valores zero ou

um. Este modelo generaliza o modelo de regressão GJS inflacionado em zero ou um, proposto por [Queiroz e Lemonte \(2021\)](#) e é um competidor para o modelo de regressão beta inflacionado em zero ou um ([Ospina e Ferrari, 2012](#)). Aspectos inferenciais, medidas de diagnóstico, resíduo e influência são desenvolvidos. De forma análoga, o modelo de regressão log-log inflacionado é apresentado como um caso limite do modelo de regressão *power logit* inflacionado. Adicionalmente, o pacote `PLreg` pode ser usado, de forma intermediária, para ajustar os modelos *power logit* e log-log inflacionados. A aplicação em dados reais apresentada na Seção 5.5 evidencia a utilidade do modelo de regressão proposto.

A segunda parte desta tese propôs-se a desenvolver métodos inferenciais robustos em modelos de regressão beta inflacionados em zero ou um. Esses modelos são uma mistura de dois outros: um modelo de regressão para dados binários (parte discreta) e um modelo de regressão beta (parte contínua). A inferência em modelos de regressão para dados binários e em modelos de regressão beta é fortemente influenciada por observações atípicas; isso é estudado nas Seções 6.1.2 e 6.1.3. Os estimadores propostos são baseados na junção de métodos inferenciais robustos aplicados às duas partes do modelo de forma separada. Para a parte discreta, consideramos um estimador baseado na divergência potência entre densidades e, para a parte contínua, consideramos estimadores baseados em L_q -verossimilhanças reparametrizadas e, também, em divergências potências entre densidades. Propriedades de robustez dos estimadores propostos são estudadas. Simulações de Monte Carlo e aplicações em dados simulados ilustram as vantagens da metodologia robusta sobre a inferência baseada em máxima verossimilhança usual quando há presença de observações atípicas nos dados.

Os estimadores robustos propostos no Capítulo 6 dependem de duas constantes de afinação, uma para os parâmetros referentes à parte discreta do modelo e outra para aqueles da parte contínua. A escolha dessas constantes é crucial para esses estimadores. Um algoritmo orientado pelos dados para a escolha dessas constantes é proposto na Seção 6.2.7 e tem como base o trabalho de [Ribeiro e Ferrari \(2022\)](#). Como subproduto do algoritmo proposto, obtemos uma metodologia inédita para a escolha de constantes de afinação orientada pelos dados para métodos robustos em modelos de regressão para dados binários.

Concluimos esta tese com algumas direções interessantes para futuras pesquisas, algumas das quais já em desenvolvimento.

1. Desenvolver testes de hipóteses não encaixadas no modelo de regressão *power logit* para testar $\mathcal{H}_0 : \lambda = 0$ contra $\mathcal{H}_0 : \lambda > 0$; ver [Vuong \(1989\)](#). Esse procedimento é uma ferramenta adicional para auxiliar na escolha entre o modelo de regressão *power logit* e log-log.
2. Propor métodos de seleção de modelos (veja, por exemplo, [Schneider et al. \(2020\)](#)); aqui a escolha da distribuição *power logit* para analisar os dados é feita com base em métodos de diagnóstico.
3. Considerar diferentes estruturas de regressão nos modelos de regressão *power logit*, como modelos não lineares, espaciais, séries temporais, componentes mistos e florestas aleatórias.
4. Aplicar os métodos inferenciais robustos no modelo de regressão beta inflacionado em conjuntos de dados reais.
5. Avaliar o desempenho de outros estimadores robustos para a parte discreta no modelo de regressão beta inflacionado.
6. Propor métodos inferenciais robustos no modelo de regressão beta inflacionado em zero e um.

Apêndice A

A transformação *power logit*

A seguir, apresentamos os resultados de uma breve simulação a fim de ilustrar a eficácia da transformação *power logit* em transformar dados limitados assimétricos em simétricos. Para tanto, geramos 1000 amostras de tamanho $n = 200$ de uma variável aleatória W com função densidade de probabilidade $f(w; \boldsymbol{\theta})$ e suporte em $(0, 1)$. Em seguida, para cada amostra, calculamos o coeficiente de assimetria amostral, dado por $\hat{\mu}_3(\mathbf{w}) = \sum_{i=1}^{200} (w_i - \bar{w})^3 / [(200 - 1)s^3]$, em que \bar{w} e s são a média e o desvio padrão amostral, respectivamente. Consideramos as distribuições a seguir.

1. Distribuição beta: $W \sim \text{beta}(\mu, \sigma)$; μ é a média de W e σ é um parâmetro de precisão.

$$f(w; \boldsymbol{\theta}) = \frac{\Gamma(\sigma)}{\Gamma(\mu\sigma)\Gamma((1-\mu)\sigma)} w^{\mu\sigma-1} (1-w)^{(1-\mu)\sigma-1}, \quad 0 < w < 1, \quad \boldsymbol{\theta} = (\mu, \sigma)^\top.$$

2. Distribuição kumaraswamy: $W \sim \text{Kumaraswamy}(\mu, \sigma)$; μ é a mediana de W e σ é um parâmetro de precisão.

$$f(w; \boldsymbol{\theta}) = \frac{\sigma \log(0.5)}{\log(1-\mu^\sigma)} w^{\sigma-1} (1-w^\sigma)^{\frac{\log(0.5)}{\log(1-\mu^\sigma)}-1}, \quad 0 < w < 1, \quad \boldsymbol{\theta} = (\mu, \sigma)^\top.$$

3. Distribuição gama unitária: $W \sim \text{UnitGamma}(\mu, \sigma)$, μ é a mediana de W e σ é um parâmetro de precisão.

$$f(w; \boldsymbol{\theta}) = \frac{\left[\frac{\mu^{1/\sigma}}{1-\mu^{1/\sigma}} \right]^\sigma}{\Gamma(\sigma)} w^{\frac{\mu^{1/\sigma}}{1-\mu^{1/\sigma}}-1} \left[\log\left(\frac{1}{w}\right) \right]^{\sigma-1}, \quad 0 < w < 1, \quad \boldsymbol{\theta} = (\mu, \sigma)^\top.$$

4. Distribuição beta generalizada: $W \sim \text{Gbeta}(\mu, \sigma, \nu)$; μ , σ e ν são parâmetros de forma.

$$f(w; \boldsymbol{\theta}) = \frac{\Gamma(\sigma)}{\Gamma(\mu\sigma)\Gamma((1-\mu)\sigma)} \frac{1}{\nu w} w^{\mu\sigma/\nu} (1-w^{1/\nu})^{(1-\mu)\sigma-1}, \quad 0 < w < 1, \quad \boldsymbol{\theta} = (\mu, \sigma, \nu)^\top.$$

Se $X \sim \text{beta}(\mu, \sigma)$, então $X^\nu \sim \text{Gbeta}(\mu, \sigma, \nu)$.

Calculamos a média dos coeficientes de assimetria das 1000 amostras para y (dados originais), $\text{logit}(y)$ (dados transformados pela função logito), e $\text{logit}(y^\lambda)$ (dados transformados pela função *power logit*). O valor de λ foi escolhido para cada amostra de modo a minimizar o coeficiente de assimetria amostral. A Figura A.1 apresenta os resultados. Note que cada ponto nos gráficos representa a média de 1000 coeficientes de assimetria obtidos de amostras de tamanho 200 das distribuições especificadas. Por exemplo, a Figura A.1(a) apresenta as médias dos coeficientes de assimetria para amostras geradas de uma distribuição beta com diferentes valores de μ e $\phi = 8$. Conforme observamos na Figura, quando a distribuição de y é altamente assimétrica (coeficiente de

assimetria médio menor que -1 ou maior que 1), a distribuição de $\text{logit}(y)$ também é assimétrica. Se a distribuição de y é assimétrica à direita (esquerda), a distribuição de $\text{logit}(y)$ é assimétrica à esquerda (direita). A distribuição de $\text{logit}(y^\lambda)$ é praticamente simétrica para todas as situações, exceto em alguns casos extremos (veja a Figura A.1(g) com μ muito pequeno). Além disso, quando a distribuição de y é aproximadamente simétrica, a distribuição de $\text{logit}(y)$ também é simétrica e o valor escolhido de λ é aproximadamente 1. Em geral, os gráficos na Figura A.1 sugerem que a transformação logito frequentemente falha ao transformar dados assimétricos em simétricos. Em contraste, a transformação *power logit* mostra-se eficaz, desde que o parâmetro potência (λ) seja escolhido corretamente.

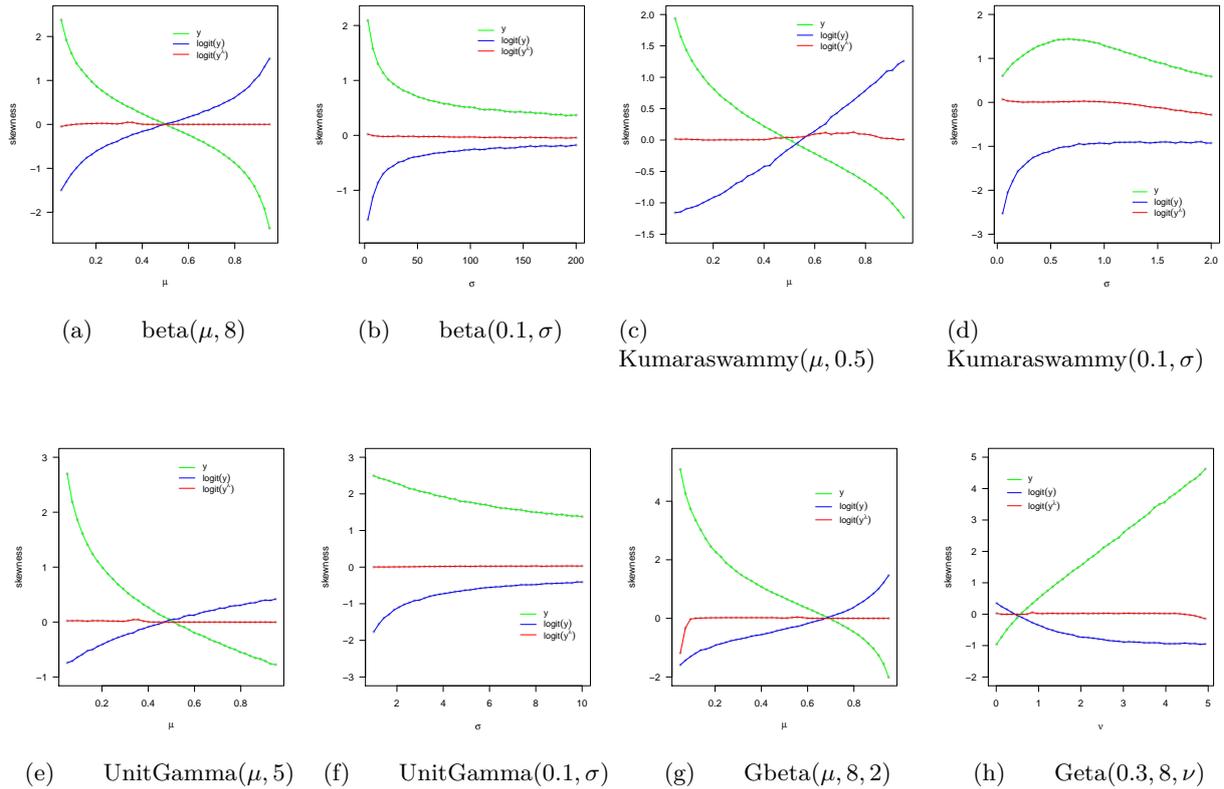


Figura A.1: Gráficos da média dos coeficientes de assimetria de y , $\text{logit}(y)$ e $\text{logit}(y^\lambda)$ para dados gerados de diferentes distribuições com suporte em $(0, 1)$.

Apêndice B

Demonstrações

B.1 Demonstrações das proposições do Capítulo 2

A seguir, são apresentadas as demonstrações das proposições apresentadas na Seção 2.2.

Demonstração da Proposição 1

Seja $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$. Esse resultado segue diretamente do fato de que

$$Z = \frac{1}{\sigma} \log \left(\frac{Y^{(\lambda)}}{\mu^{(\lambda)}} \right) \Rightarrow Y = \left[\frac{\mu^\lambda e^{\sigma Z}}{1 - \mu^\lambda (1 - e^{\sigma Z})} \right]^{1/\lambda},$$

com $Z \sim \text{S}(0, 1; r)$.

Demonstração da Proposição 2

Para verificar que σ é parâmetro de dispersão, seguiremos os argumentos de [Mitnik e Baek \(2013\)](#) e vamos considerar a ordem de dispersão quantílica (*quantile spread order*) introduzida recentemente por [Townsend e Colonius \(2005\)](#). Considere F^{-1} a função quantílica de uma variável aleatória Z . A dispersão quantílica de Z é dada pela função $QS_Z(u) = F^{-1}(1 - u) - F^{-1}(u)$, para $u \in (0, 1/2)$. Agora, considere $Z_1, F_{Z_1}^{-1}, Z_2, F_{Z_2}^{-1}$ como definido anteriormente. Dizemos que Z_1 é menor que Z_2 em ordem de dispersão quantílica, ou $Z_1 \leq_{QS} Z_2$, se $QS_{Z_1}(u) \leq QS_{Z_2}(u)$, para $u \in (0, 1/2)$. Note que, conceitualmente, a ordem \leq_{QS} corresponde à comparação de Z_1 e Z_2 com relação à dispersão, uma vez que estamos verificando se a distância entre dois quaisquer quantis simétricos de Z_2 é pelo menos tão grande quanto a distância para os correspondentes quantis de Z_1 ([Mitnik e Baek, 2013](#)).

Dessa forma, se $Y_1 \sim \text{PL}(\mu, \sigma_1, \lambda; r)$ e $Y_2 \sim \text{PL}(\mu, \sigma_2, \lambda; r)$, então precisamos mostrar que Y_1 é menor que Y_2 em ordem de dispersão quantílica se, e somente se, $\sigma_1 \leq \sigma_2$. Temos que

$$QS_{Y_1}(u) = \left[\frac{\mu^\lambda e^{\sigma_1 z_{1-u}}}{1 - \mu^\lambda (1 - e^{\sigma_1 z_{1-u}})} \right]^{1/\lambda} - \left[\frac{\mu^\lambda e^{\sigma_1 z_u}}{1 - \mu^\lambda (1 - e^{\sigma_1 z_u})} \right]^{1/\lambda},$$

em que $u \in (0, 1/2)$. Verifiquemos se $QS_{Y_1}(u) - QS_{Y_2}(u) \leq 0$ se, e só se, $\sigma_1 \leq \sigma_2$. Note que

$$\begin{aligned} QS_{Y_1}(u) - QS_{Y_2}(u) &= \left[\frac{\mu^\lambda e^{\sigma_1 z_{1-u}}}{1 - \mu^\lambda (1 - e^{\sigma_1 z_{1-u}})} \right]^{1/\lambda} - \left[\frac{\mu^\lambda e^{\sigma_1 z_u}}{1 - \mu^\lambda (1 - e^{\sigma_1 z_u})} \right]^{1/\lambda} \\ &\quad - \left[\frac{\mu^\lambda e^{\sigma_2 z_{1-u}}}{1 - \mu^\lambda (1 - e^{\sigma_2 z_{1-u}})} \right]^{1/\lambda} + \left[\frac{\mu^\lambda e^{\sigma_2 z_u}}{1 - \mu^\lambda (1 - e^{\sigma_2 z_u})} \right]^{1/\lambda}. \end{aligned}$$

Como $Z \sim S(0, 1; r)$ e $u \in (0, 1/2)$, então $z_u < 0$ e $z_{1-u} > 0$, e $z_u = -z_{1-u}$. Assim,

$$\begin{aligned} QS_{Y_1}(u) - QS_{Y_2}(u) &= \left[\frac{\mu^\lambda e^{\sigma_1 z_{1-u}}}{1 - \mu^\lambda (1 - e^{\sigma_1 z_{1-u}})} \right]^{1/\lambda} - \left[\frac{\mu^\lambda e^{\sigma_2 z_{1-u}}}{1 - \mu^\lambda (1 - e^{\sigma_2 z_{1-u}})} \right]^{1/\lambda} \\ &+ \left[\frac{\mu^\lambda e^{-\sigma_2 z_{1-u}}}{1 - \mu^\lambda (1 - e^{-\sigma_2 z_{1-u}})} \right]^{1/\lambda} - \left[\frac{\mu^\lambda e^{-\sigma_1 z_{1-u}}}{1 - \mu^\lambda (1 - e^{-\sigma_1 z_{1-u}})} \right]^{1/\lambda}. \end{aligned}$$

Agora, note que

$$\left[\frac{\mu^\lambda e^{\sigma_1 z_{1-u}}}{1 - \mu^\lambda (1 - e^{\sigma_1 z_{1-u}})} \right]^{1/\lambda} \leq \left[\frac{\mu^\lambda e^{\sigma_2 z_{1-u}}}{1 - \mu^\lambda (1 - e^{\sigma_2 z_{1-u}})} \right]^{1/\lambda}$$

e

$$\left[\frac{\mu^\lambda e^{-\sigma_2 z_{1-u}}}{1 - \mu^\lambda (1 - e^{-\sigma_2 z_{1-u}})} \right]^{1/\lambda} \leq \left[\frac{\mu^\lambda e^{-\sigma_1 z_{1-u}}}{1 - \mu^\lambda (1 - e^{-\sigma_1 z_{1-u}})} \right]^{1/\lambda}$$

se, e só se, $\sigma_1 \leq \sigma_2$. Logo, o σ é parâmetro de dispersão.

Demonstração da Proposição 3

Seja $Y \sim \text{PL}(\mu, \sigma, 1; r)$ e defina $W = 1 - Y$. Para todo $w \in (0, 1)$, temos

$$F_W(w; \mu, \sigma) = \mathbb{P}(W \leq w) = 1 - \mathbb{P}(Y \leq 1 - w)$$

assim,

$$\begin{aligned} f_W(w; \mu, \sigma) &= f_Y(1 - w; \mu, \sigma) = \frac{1}{\sigma y(1 - y)} r \left(\left\{ \frac{1}{\sigma} \log \left(\frac{(1 - w)/w}{\mu/(1 - \mu)} \right) \right\}^2 \right) \\ &= \frac{1}{\sigma w(1 - w)} r \left(\left\{ \frac{1}{\sigma} \left[-\log \left(\frac{w}{1 - w} \right) + \log \left(\frac{1 - \mu}{\mu} \right) \right] \right\}^2 \right) \\ &= \frac{1}{\sigma w(1 - w)} r \left(\left\{ \frac{1}{\sigma} \log \left(\frac{w/(1 - w)}{(1 - \mu)/\mu} \right) \right\}^2 \right). \end{aligned}$$

Assim, $W = 1 - Y \sim \text{PL}(1 - \mu, \sigma, 1; r)$.

Demonstração da Proposição 4

1. Seja $Y \sim \text{PL}(\mu, \sigma, 1; r)$, então

$$\begin{aligned} f_Y(y; \mu, \sigma) &= \frac{1}{\sigma y(1 - y)} r \left(\left\{ \frac{1}{\sigma} \log \left(\frac{y/(1 - y)}{\mu/(1 - \mu)} \right) \right\}^2 \right) \\ &= \frac{r(\{\sigma^{-1}[t(y) - t(\mu)]\}^2)}{\sigma y(1 - y)}, \end{aligned}$$

em que $t(y) = \log(y/(1 - y))$. Logo, $Y \sim \text{GJS}(\mu, \sigma; r)$.

2. Assuma que $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$ e seja $W = Y^\lambda$, então, para $w \in (0, 1)$,

$$F_W(w; \mu, \sigma) = \mathbb{P}(W \leq w) = \mathbb{P}(Y^\lambda \leq w) = F_Y(w^{1/\lambda}; \mu, \sigma).$$

Logo,

$$\begin{aligned} f_W(w; \mu, \sigma, \lambda) &= \frac{w^{1/\lambda}}{\lambda w} f_Y(w^{1/\lambda}; \mu, \sigma, \lambda) = \frac{w^{1/\lambda}}{\lambda w} \frac{\lambda}{\sigma w^{1/\lambda}(1-w)} r \left(\left\{ \frac{1}{\sigma} \log \left(\frac{w/(1-w)}{\mu^\lambda/(1-\mu^\lambda)} \right) \right\}^2 \right) \\ &= \frac{1}{\sigma w(1-w)} r \left(\left\{ \frac{1}{\sigma} \left[\log \left(\frac{w}{1-w} \right) - \log \left(\frac{\mu^\lambda}{1-\mu^\lambda} \right) \right] \right\}^2 \right), \end{aligned}$$

de modo que $W = Y^\lambda \sim \text{GJS}(\mu^\lambda, \sigma; r)$. Analogamente, pode-se mostrar que se $Y^\lambda \sim \text{GJS}(\mu^\lambda, \sigma; r)$, então $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$.

3. Consultar [Lemonte e Bazán \(2016\)](#).

Demonstração da Proposição 5

Usando a regra de l'Hôpital, pode-se mostrar que

$$\frac{y^{(\lambda)}}{\mu^{(\lambda)}} \rightarrow \frac{\log \mu}{\log y}, \quad \text{quando } \lambda \rightarrow 0^+.$$

Assim,

$$\lim_{\lambda \rightarrow 0^+} z = -\frac{1}{\sigma} \log \left(\frac{\log y}{\log \mu} \right) = -\frac{1}{\sigma} \log \left(\frac{-\log y}{-\log \mu} \right).$$

Dessa forma, para $y \in (0, 1)$, temos

$$\begin{aligned} \lim_{\lambda \rightarrow 0^+} F_Y(y; \mu, \sigma, \lambda) &= \mathbb{P} \left(Z \leq -\frac{1}{\sigma} \log \left(\frac{-\log y}{-\log \mu} \right) \right) \\ &= \mathbb{P} (-\log(-\log \mu) + \sigma Z \leq -\log(-\log y)) \\ &= \mathbb{P} (W \leq -\log(-\log y)) \\ &= \mathbb{P} (\exp\{-\exp\{-W\}\} \leq y), \end{aligned}$$

em que $W = -\log(-\log \mu) + \sigma Z \sim \text{S}(-\log(-\log \mu), \sigma^2; r)$. Logo,

$$Y \xrightarrow{\mathcal{D}} \exp\{-\exp\{-W\}\}, \quad \lambda \rightarrow 0^+.$$

Ou, equivalentemente,

$$-\log(-\log Y) \xrightarrow{\mathcal{D}} W, \quad \lambda \rightarrow 0^+.$$

Demonstração da Proposição 6

Seja $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$ e defina $W = Y^c$, para $c > 0$, então

$$\begin{aligned} F_W(w; \mu, \sigma, \lambda) &= \mathbb{P}(Y^c \leq w) \\ &= F_Y(w^{1/c}; \mu, \sigma, \lambda). \end{aligned}$$

Assim,

$$\begin{aligned} f_W(w; \mu, \sigma, \lambda) &= \frac{w^{1/c}}{cw} f_Y(w^{1/c}; \mu, \sigma, \lambda) \\ &= \frac{w^{1/c}}{cw} \frac{\lambda}{\sigma w^{1/c}(1-w^{1/c})} r \left(\left\{ \frac{1}{\sigma} \log \left(\frac{w^{1/c}/(1-w^{1/c})}{\mu^{c\lambda/c}/(1-\mu^{c\lambda/c})} \right) \right\}^2 \right) \\ &= \frac{1}{\sigma w(1-w^{1/c})} r \left(\left\{ \frac{1}{\sigma} \left[\log \left(\frac{w^{1/c}}{1-w^{1/c}} \right) - \log \left(\frac{\mu^{c\lambda/c}}{1-\mu^{c\lambda/c}} \right) \right] \right\}^2 \right). \end{aligned}$$

Logo, $Y^c \sim \text{PL}(\mu^c, \sigma, \lambda/c; r)$.

B.2 Derivadas da função de log-verossimilhança no modelo de regressão *power logit*

Derivadas de segunda ordem da função de log-verossimilhança

Conforme apresentado na Seção 3.2, a função de log-verossimilhança para θ é dada por

$$\ell(\theta) = \sum_{i=1}^n \ell_i(\mu_i, \sigma_i, \lambda), \quad (\text{B.1})$$

em que

$$\ell_i(\mu_i, \sigma_i, \lambda) = \log \lambda - \log \sigma_i - \log\{1 - y_i^\lambda\} + \log\{r(z_i^2)\},$$

sendo $z_i = h(y_i; \mu_i, \sigma_i, \lambda)$. As segundas derivadas de $\ell(\theta)$ com respeito aos β 's, τ 's e λ são dadas por

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \beta_R \partial \beta_m} &= \sum_{y_i=1}^n \frac{\partial}{\partial \mu_i} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} \frac{d\mu_i}{d\eta_{1i}} \right) \frac{d\mu_i}{d\eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_m} x_{iR} \\ &= \sum_{y_i=1}^n \left(\frac{\partial \ell_i^2(\mu_i, \sigma_i, \lambda)}{\partial \mu_i^2} \frac{d\mu_i}{d\eta_{1i}} + \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} \frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_{1i}} \right) \frac{d\mu_i}{d\eta_{1i}} x_{iR} x_{im}, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \tau_1 \partial \tau_N} &= \sum_{y_i=1}^n \frac{\partial}{\partial \sigma_i} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} \right) \frac{d\sigma_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_N} s_{i1} \\ &= \sum_{i=1}^n \left(\frac{\partial \ell_i^2(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i^2} \frac{d\sigma_i}{d\eta_{2i}} + \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i} \frac{\partial}{\partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} \right) \frac{d\sigma_i}{d\eta_{2i}} s_{i1} s_{iN}, \end{aligned}$$

$$\frac{\partial^2 \ell(\theta)}{\partial \lambda^2} = \sum_{y_i=1}^n \frac{\partial^2 \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \lambda^2}.$$

Como

$$\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} = \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i), \quad \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i} = \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1]$$

e

$$\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \lambda} = \frac{1}{\lambda} + \frac{y_i^\lambda \log y_i}{1 - y_i^\lambda} - \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{\log y_i}{(1 - y_i^\lambda)} - \frac{\log \mu_i}{(1 - \mu_i^\lambda)} \right],$$

após algumas simplificações, segue que

$$\begin{aligned} \frac{\partial \ell_i^2(\mu_i, \sigma_i, \lambda)}{\partial \mu_i^2} &= \frac{-\lambda}{[\sigma_i \mu_i (1 - \mu_i^\lambda)]^2} \left\{ \sigma_i [1 - \mu_i^\lambda (1 + \lambda)] z_i v(z_i) + \lambda [v(z_i) + z_i v'(z_i)] \right\}, \\ \frac{\partial \ell_i^2(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i^2} &= \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^2} z_i^2 [3v(z_i) + z_i v'(z_i)], \\ \frac{\partial \ell_i^2(\mu_i, \sigma_i, \lambda)}{\partial \lambda^2} &= -\frac{1}{\lambda^2} + \frac{y_i^\lambda \log^2 y_i}{(1 - y_i^\lambda)^2} - \frac{1}{\sigma_i^2} [v(z_i) + z_i v'(z_i)] \left(\frac{\log y_i}{1 - y_i^\lambda} - \frac{\log \mu_i}{1 - \mu_i^\lambda} \right)^2 \\ &\quad - \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{y_i^\lambda \log^2 y_i}{(1 - y_i^\lambda)^2} - \frac{\mu_i^\lambda \log^2 \mu_i}{(1 - \mu_i^\lambda)^2} \right], \end{aligned}$$

e

$$\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_{1i}} = -\frac{\ddot{d}_1(\mu_i)}{\dot{d}_1(\mu_i)^2} \quad \text{e} \quad \frac{\partial}{\partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} = -\frac{\ddot{d}_2(\sigma_i)}{\dot{d}_2(\sigma_i)^2}.$$

Adicionalmente, para $R = 1, \dots, p$ e $l = 1, \dots, q$, temos

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_R \partial \tau_l} = \sum_{i=1}^n \frac{\partial}{\partial \sigma_i} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} \right) \frac{d\sigma_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_l} \frac{d\mu_i}{d\eta_{1i}} x_{iR},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_R \partial \lambda} = \sum_{i=1}^n \frac{\partial}{\partial \lambda} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} \right) \frac{d\mu_i}{d\eta_{1i}} x_{iR},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \tau_l \partial \lambda} = \sum_{i=1}^n \frac{\partial}{\partial \lambda} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i} \right) \frac{d\sigma_i}{d\eta_{2i}} s_{il},$$

em que

$$\begin{aligned} \frac{\partial}{\partial \sigma_i} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} \right) &= \frac{\partial}{\partial \sigma_i} \left(\frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i) \right) \\ &= \frac{-\lambda}{\sigma_i^2 \mu_i (1 - \mu_i^\lambda)} z_i [2v(z_i) + z_i v'(z_i)], \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \lambda} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \mu_i} \right) &= \frac{\partial}{\partial \lambda} \left(\frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i) \right) \\ &= \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} \left\{ \frac{1 - \mu_i^\lambda (1 - \lambda \log \mu_i)}{\lambda (1 - \mu_i^\lambda)} z_i v(z_i) \right. \\ &\quad \left. + \frac{1}{\sigma_i} \left(\frac{\log y_i}{1 - y_i^\lambda} - \frac{\log \mu_i}{1 - \mu_i^\lambda} \right) [v(z_i) + z_i v'(z_i)] \right\}, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \lambda} \left(\frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \sigma_i} \right) &= \frac{\partial}{\partial \lambda} \left(\frac{1}{\sigma_i} [z_i^2 v(z_i) - 1] \right) \\ &= \frac{1}{\sigma_i^2} \left(\frac{\log y_i}{1 - y_i^\lambda} - \frac{\log \mu_i}{1 - \mu_i^\lambda} \right) z_i [2v(z_i) + z_i v'(z_i)]. \end{aligned}$$

Dessa forma, é possível escrever a informação observada da forma (3.7).

Apêndice C

Cálculo da matriz de perturbação Δ

Aqui são apresentados os cálculos para a obtenção da matriz de perturbação Δ , sob diferentes esquemas de perturbação, para os modelos de regressão *power logit* e *power logit* inflacionado em zero ou um.

C.1 Modelo de regressão *power logit*

C.1.1 Ponderação de casos

No esquema de perturbação de ponderação de casos, a função de log-verossimilhança perturbada para o modelo de regressão *power logit* é dada por

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \ell_i(\mu_i, \sigma_i, \lambda),$$

em que $\ell_i(\mu_i, \sigma_i, \lambda)$ é definida em 3.2. Dessa forma, para $i = 1, \dots, n$ e $R = 1, \dots, p$, temos que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} = \frac{\partial}{\partial \beta_R} \left(\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i} \right) = \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \beta_R}.$$

De (3.3), temos que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} = \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i) \frac{1}{d_1(\mu_i)} x_{iR}.$$

Analogamente, para $l = 1, \dots, q$ e $i = 1, \dots, n$,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} = \frac{\partial}{\partial \tau_l} \left(\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i} \right) = \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \tau_l}.$$

Assim, segue de (3.4) que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} = \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1] \frac{1}{d_2(\sigma_i)} s_{il}.$$

Em notação matricial, e avaliando as derivadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, temos que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{X}^\top \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 \widehat{\mathbf{D}}_\beta, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{S}^\top \widehat{\mathbf{T}}_2 \widehat{\mathbf{D}}_\tau, \end{aligned} \tag{C.1}$$

em que $\mathbf{D}_\beta = \text{diag}\{\mu_1^*, \dots, \mu_n^*\}$ e $\mathbf{D}_\tau = \text{diag}\{\sigma_1^*, \dots, \sigma_n^*\}$.

C.1.2 Perturbação individual nas covariáveis contínuas

Sejam \mathbf{X} e \mathbf{S} as matrizes de covariáveis do modelo de regressão *power logit*. Vamos assumir que $\mathbf{X} \neq \mathbf{S}$ e que a j -ésima coluna de \mathbf{X} , digamos \mathbf{x}_j , com $j = 1, \dots, p$, é perturbada aditivamente da forma definida em (3.14). A função de log-verossimilhança do modelo perturbado é dada por (3.15). Desse modo, para $R = 1, \dots, p$ e $\forall i = 1, \dots, n$, segue que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R} \right) \\ &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \frac{\partial \eta_{1i\omega}}{\partial \beta_R} \right) \\ &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \right) \frac{\partial \eta_{1i\omega}}{\partial \beta_R} + \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{1i\omega}}{\partial \beta_R} \right) \\ &= \frac{\partial}{\partial \mu_{i\omega}} \left(\frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \right) \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \frac{\partial \eta_{1i\omega}}{\partial \omega_i} \frac{\partial \eta_{1i\omega}}{\partial \beta_R} + \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{1i\omega}}{\partial \beta_R} \right) \\ &= \left(\frac{\partial^2 \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}^2} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} + \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{\partial}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \right) \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \frac{\partial \eta_{1i\omega}}{\partial \omega_i} \frac{\partial \eta_{1i\omega}}{\partial \beta_R} \\ &\quad + \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{1i\omega}}{\partial \beta_R} \right), \end{aligned}$$

em que

$$\frac{d\mu_{i\omega}}{d\eta_{1i\omega}} = \frac{1}{\dot{d}_1(\mu_{i\omega})}, \quad \frac{\partial}{\partial \mu_{i\omega}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} = -\frac{\ddot{d}_1(\mu_{i\omega})}{\dot{d}_1(\mu_{i\omega})^2}, \quad \frac{\partial \eta_{1i\omega}}{\partial \omega_i} = \beta_j \sigma_x.$$

Note que, se $R = j$, temos que $\partial \eta_{1i\omega} / \partial \beta_R = x_{iR} + \sigma_x \omega_i$ e se $R \neq j$, $\partial \eta_{1i\omega} / \partial \beta_R = x_{iR}$. Logo,

$$\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{1i\omega}}{\partial \beta_R} = \begin{cases} \sigma_x, & R = j, \\ 0, & R \neq j, \end{cases}$$

De modo que, se $r \neq j$, temos

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} = \left(\frac{\partial^2 \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}^2} \frac{1}{\dot{d}_1(\mu_{i\omega})} - \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{\ddot{d}_1(\mu_{i\omega})}{\dot{d}_1(\mu_{i\omega})^2} \right) \frac{\beta_j \sigma_x}{\dot{d}_1(\mu_{i\omega})} x_{iR},$$

e se $r = j$, temos que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_r \partial \omega_i} &= \left(\frac{\partial^2 \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}^2} \frac{1}{\dot{d}_1(\mu_{i\omega})} - \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{\ddot{d}_1(\mu_{i\omega})}{\dot{d}_1(\mu_{i\omega})^2} \right) \frac{\beta_j \sigma_x}{\dot{d}_1(\mu_{i\omega})} (x_{iR} + \sigma_x \omega_i) \\ &\quad + \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{\sigma_x}{\dot{d}_1(\mu_{i\omega})}. \end{aligned}$$

Defina

$$-w_{i\omega}^{(1)} = \frac{\partial^2 \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}^2} \frac{1}{\dot{d}_1(\mu_{i\omega})} - \frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega}} \frac{\ddot{d}_1(\mu_{i\omega})}{\dot{d}_1(\mu_{i\omega})^2},$$

e

$$\mu_{i\omega}^* = \frac{\lambda}{\sigma_i \mu_{i\omega} (1 - \mu_{i\omega}^\lambda)}.$$

Dessa forma,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} = \begin{cases} -w_{i\omega}^{(1)} \frac{\beta_j \sigma_x}{\dot{d}_1(\mu_{i\omega})} x_{iR}, & R \neq j, \\ -w_{i\omega}^{(1)} \frac{\beta_j \sigma_x}{\dot{d}_1(\mu_{i\omega})} (x_{iR} + \sigma_x \omega_i) + z_{i\omega} v(z_{i\omega}) \mu_{i\omega}^* \frac{\sigma_x}{\dot{d}_1(\mu_{i\omega})}, & R = j, \end{cases} \quad (\text{C.2})$$

em que $z_{i\omega} = h(y; \mu_{i\omega}, \sigma_i, \lambda)$. Avaliando (C.2) em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, temos

$$\left. \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\sigma_x \hat{\boldsymbol{\beta}}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1 \widehat{\mathbf{T}}_1 + \sigma_x \mathbf{c}_\beta^j \hat{\boldsymbol{\mu}}^{*\top} \widehat{\mathbf{W}}_1 \widehat{\mathbf{T}}_1, \quad (\text{C.3})$$

sendo $\mathbf{c}_\beta^j = (c_1 \dots, c_p)^\top$ um vetor de dimensão $p \times 1$ tal que $c_i = 0 \forall i \neq j$ e $c_i = 1$ para $i = j$, e $\hat{\boldsymbol{\beta}}_j$ denota o j -ésimo elemento de $\hat{\boldsymbol{\beta}} = (\beta_1, \dots, \beta_p)^\top$, $j = 1, \dots, p$.

Da mesma forma, para $l = 1, \dots, q$ e $\forall i = 1, \dots, n$, segue que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} &= \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} \\ &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_l} \right) \\ &= \frac{\partial}{\partial \mu_{i\omega}} \left(\frac{\partial \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} \right) \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \frac{\partial \eta_{1i\omega}}{\partial \omega_i} \frac{\partial \eta_{2i}}{\partial \tau_l} \\ &= \frac{\partial^2 \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega} \partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} \frac{d\mu_{i\omega}}{d\eta_{1i\omega}} \frac{\partial \eta_{1i\omega}}{\partial \omega_i} \frac{\partial \eta_{2i}}{\partial \tau_l}. \end{aligned}$$

Definindo

$$-w_{i\omega}^{(4)} = \frac{\partial^2 \ell_i(\mu_{i\omega}, \sigma_i, \lambda)}{\partial \mu_{i\omega} \partial \sigma_i},$$

segue que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} = -w_{i\omega}^{(4)} \frac{1}{d_1(\mu_{i\omega})} \frac{1}{d_2(\sigma_i)} \beta_j \sigma_x s_{il}. \quad (\text{C.4})$$

Avaliando (C.4) em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, temos que

$$\left. \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\sigma_x \hat{\boldsymbol{\beta}}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2. \quad (\text{C.5})$$

Agora, vamos assumir que a k -ésima coluna de \mathbf{S} , digamos \mathbf{s}_k com $k = 1, \dots, q$, é perturbada aditivamente da forma definida em (3.16). A função de log-verossimilhança do modelo perturbado é dada por (3.17).

Assim, para $l = 1, \dots, q$ e $\forall i = 1, \dots, n$, segue que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l} \right) \\ &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \frac{\partial \eta_{2i\omega}}{\partial \tau_l} \right) \\ &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \right) \frac{\partial \eta_{2i\omega}}{\partial \tau_l} + \frac{\partial \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{2i\omega}}{\partial \tau_l} \right) \\ &= \frac{\partial}{\partial \sigma_{i\omega}} \left(\frac{\partial \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \right) \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \frac{\partial \eta_{2i\omega}}{\partial \omega_i} \frac{\partial \eta_{2i\omega}}{\partial \tau_l} + \frac{\partial \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{2i\omega}}{\partial \tau_l} \right) \\ &= \left(\frac{\partial^2 \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}^2} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} + \frac{\partial \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}} \frac{\partial}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \right) \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \frac{\partial \eta_{2i\omega}}{\partial \omega_i} \frac{\partial \eta_{2i\omega}}{\partial \tau_l} \\ &\quad + \frac{\partial \ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{2i\omega}}{\partial \tau_l} \right), \end{aligned}$$

em que

$$\frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} = \frac{1}{d_2(\sigma_{i\omega})}, \quad \frac{\partial}{\partial \sigma_{i\omega}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} = -\frac{\ddot{d}_1(\sigma_{i\omega})}{d_1(\sigma_{i\omega})^2}, \quad \frac{\partial \eta_{2i\omega}}{\partial \omega_i} = \tau_k \sigma_s.$$

Se $l = k$, temos que $\partial\eta_{2i\omega}/\partial\tau_l = s_{il} + \sigma_s\omega_i$ e se $l \neq k$, temos que $\partial\eta_{2i\omega}/\partial\tau_l = s_{il}$. De modo que,

$$\frac{\partial}{\partial\omega_i} \frac{\partial\eta_{2i\omega}}{\partial\tau_l} = \begin{cases} \sigma_s, & l = k, \\ 0, & l \neq k, \end{cases}$$

Portanto, se $l \neq k$,

$$\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\tau_l\partial\omega_i} = \left(\frac{\partial^2\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}^2} \frac{1}{\dot{d}_2(\sigma_{i\omega})} - \frac{\partial\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}} \frac{\ddot{d}_2(\sigma_{i\omega})}{\dot{d}_2(\sigma_{i\omega})^2} \right) \frac{\tau_k\sigma_s}{\dot{d}_2(\sigma_{i\omega})} s_{il},$$

e, se $l = k$,

$$\begin{aligned} \frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\tau_l\partial\omega_i} &= \left(\frac{\partial^2\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}^2} \frac{1}{\dot{d}_2(\sigma_{i\omega})} - \frac{\partial\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}} \frac{\ddot{d}_2(\sigma_{i\omega})}{\dot{d}_2(\sigma_{i\omega})^2} \right) \frac{\tau_k\sigma_s}{\dot{d}_2(\sigma_{i\omega})} (s_{il} + \sigma_s\omega_i) \\ &+ \frac{\partial\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}} \frac{\sigma_s}{\dot{d}_2(\sigma_{i\omega})}. \end{aligned}$$

Defina, para $i = 1, \dots, n$,

$$-w_{i\omega}^{(2)} = \frac{\partial^2\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}^2} \frac{1}{\dot{d}_2(\sigma_{i\omega})} - \frac{\partial\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}} \frac{\ddot{d}_2(\sigma_{i\omega})}{\dot{d}_2(\sigma_{i\omega})^2}$$

e

$$\sigma_{i\omega}^* = \frac{1}{\sigma_{i\omega}} [z_{i\omega}^2 v(z_{i\omega}^2) - 1],$$

em que $z_{i\omega} = h(y; \mu, \sigma_{i\omega}, \lambda)$. Logo,

$$\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\tau_l\partial\omega_i} = \begin{cases} -w_{i\omega}^{(2)} \frac{\tau_k\sigma_s}{\dot{d}_2(\sigma_{i\omega})} s_{il}, & r \neq j, \\ -w_{i\omega}^{(2)} \frac{\tau_k\sigma_s}{\dot{d}_2(\sigma_{i\omega})} (s_{il} + \sigma_s\omega_i) + \frac{\sigma_s\sigma_{i\omega}^*}{\dot{d}_2(\sigma_{i\omega})}, & r = j, \end{cases} \quad (\text{C.6})$$

Avaliando (C.6) em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, temos

$$\left. \frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\sigma_s \hat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2 \hat{\mathbf{T}}_2 + \sigma_s \mathbf{c}_\tau^k \hat{\boldsymbol{\sigma}}^{*\top} \hat{\mathbf{T}}_2, \quad (\text{C.7})$$

sendo $\mathbf{c}_\tau^k = (c_1, \dots, c_q)^\top$ um vetor de dimensão $q \times 1$ tal que $c_i = 0 \forall i \neq k$ e $c_i = 1$ para $i = k$, e $\hat{\tau}_k$ denota o k -ésimo elemento de $\hat{\boldsymbol{\tau}} = (\tau_1, \dots, \tau_q)^\top$, $k = 1, \dots, q$. Agora, para $R = 1, \dots, p$ e $\forall i = 1, \dots, n$, segue que

$$\begin{aligned} \frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\beta_R\partial\omega_i} &= \frac{\partial}{\partial\omega_i} \left(\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\beta_R} \right) \\ &= \frac{\partial}{\partial\omega_i} \left(\frac{\partial\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\mu_i} \frac{d\mu_i}{d\eta_{1i}} \frac{\partial\eta_{1i}}{\partial\beta_R} \right) \\ &= \frac{\partial}{\partial\sigma_{i\omega}} \left(\frac{\partial\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\mu_i} \frac{d\mu_i}{d\eta_{1i}} \right) \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \frac{\partial\eta_{2i\omega}}{\partial\omega_i} \frac{\partial\eta_{1i}}{\partial\beta_R} \\ &= \frac{\partial^2\ell_i(\mu_i, \sigma_{i\omega}, \lambda)}{\partial\sigma_{i\omega}\partial\mu_i} \frac{d\mu_i}{d\eta_{1i}} \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}} \frac{\partial\eta_{2i\omega}}{\partial\omega_i} \frac{\partial\eta_{1i}}{\partial\beta_R}. \end{aligned}$$

Assim, segue que

$$\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\beta_R\partial\omega_i} = -w_{i\omega}^{(4)} \frac{1}{\dot{d}_1(\mu_i)} \frac{1}{\dot{d}_2(\sigma_{i\omega})} \tau_k \sigma_s x_{iR}. \quad (\text{C.8})$$

Avaliando (C.8) em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, temos

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\sigma_s \widehat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4 \widehat{\mathbf{T}}_1 \widehat{\mathbf{T}}_2. \quad (\text{C.9})$$

C.2 Modelo de regressão *power logit* inflacionado em zero ou um

Nesta Seção, omitiremos alguns detalhes técnicos, que podem ser encontrados na Seção C.1.

C.2.1 Ponderação de casos

Temos que

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \ell_1(\boldsymbol{\kappa}|\boldsymbol{\omega}) + \ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda|\boldsymbol{\omega}),$$

em que

$$\ell_1(\boldsymbol{\kappa}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \ell_i(\alpha_i), \quad \ell_2(\boldsymbol{\kappa}|\boldsymbol{\omega}) = \sum_{i \in \wp} \omega_i \ell_i(\mu_i, \sigma_i, \lambda),$$

em que $\ell_i(\alpha_i)$ e $\ell_i(\mu_i, \sigma_i, \lambda)$ são definidas em 5.4. Assim, para $i = 1, \dots, n$ e $t = 1, \dots, m$, temos que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \kappa_t \partial \omega_i} = \frac{\partial^2 \ell_1(\boldsymbol{\kappa}|\boldsymbol{\omega})}{\partial \kappa_t \partial \omega_i} = \frac{\partial}{\partial \kappa_t} \left(\frac{\partial \ell_1(\boldsymbol{\kappa}|\boldsymbol{\omega})}{\partial \omega_i} \right) = \frac{\partial \ell_i(\alpha_i)}{\partial \kappa_t} = \frac{[\mathbb{I}_{\{c\}}(y_i) - \alpha_i]}{\alpha_i(1 - \alpha_i)} \frac{z_{iR}}{d_0(\alpha_i)}.$$

Analogamente, para $R = 1, \dots, p$, $l = 1, \dots, q$ e $i \in \wp$,

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} &= \frac{\partial^2 \ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} = \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \beta_R} = \frac{\lambda}{\sigma_i \mu_i (1 - \mu_i^\lambda)} z_i v(z_i) \frac{1}{d_1(\mu_i)} x_{iR}, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} &= \frac{\partial^2 \ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} = \frac{\partial \ell_i(\mu_i, \sigma_i, \lambda)}{\partial \tau_l} = \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1] \frac{1}{d_2(\sigma_i)} s_{il}. \end{aligned}$$

Note que, para todo $i \notin \wp$,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_R \partial \omega_i} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} = 0.$$

Portanto, colocando em notação matricial e avaliando a derivada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, temos que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{Z}^\top \widehat{\mathbf{A}} \widehat{\mathbf{T}}_0 \widehat{\mathbf{D}}_\kappa, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{X}^\top \widehat{\mathbf{W}}^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{D}}_\beta^\dagger, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{S}^\top \widehat{\mathbf{T}}_2^\dagger \widehat{\mathbf{D}}_\tau^\dagger, \end{aligned}$$

em que $\mathbf{D}_\kappa = \text{diag}\{\alpha_1^*, \dots, \alpha_n^*\}$.

C.2.2 Perturbação individual nas covariáveis contínuas

Perturbação individual das covariáveis do componente discreto

Vamos assumir que $\mathbf{Z} \neq \mathbf{X} \neq \mathbf{S}$ e que a u -ésima coluna de \mathbf{Z} , digamos z_u com $u = 1, \dots, m$, é perturbada aditivamente, de modo que

$$z_{iu\omega} = z_{iu} + \sigma_z \omega_i.$$

A função de log-verossimilhança do modelo perturbado é dada por

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \ell_1(\boldsymbol{\kappa}|\boldsymbol{\omega}) + \ell_2(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \sum_{i=1}^n \ell_i(\alpha_{i\omega}) + \sum_{i \in \wp} \ell_i(\mu_i, \sigma_i, \lambda).$$

Assim, para $t = 1, \dots, m$, segue que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \kappa_t \partial \omega_i} &= \frac{\partial^2 \ell_1(\boldsymbol{\kappa}|\boldsymbol{\omega})}{\partial \kappa_t \partial \omega_i} = \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_1(\boldsymbol{\kappa}|\boldsymbol{\omega})}{\partial \kappa_t} \right) \\ &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} \right) \\ &= \frac{\partial}{\partial \omega_i} \left(\frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \right) \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} + \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} \right) \\ &= \frac{\partial}{\partial \alpha_{i\omega}} \left(\frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \right) \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \frac{\partial \eta_{0i\omega}}{\partial \omega_i} \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} + \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} \right) \\ &= \left(\frac{\partial^2 \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}^2} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} + \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{\partial}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \right) \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \frac{\partial \eta_{0i\omega}}{\partial \omega_i} \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} \\ &\quad + \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} \left(\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} \right), \end{aligned}$$

com

$$\frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} = \frac{1}{\dot{d}_0(\alpha_{i\omega})}, \quad \frac{\partial}{\partial \alpha_{i\omega}} \frac{d\alpha_{i\omega}}{d\eta_{0i\omega}} = -\frac{\ddot{d}_0(\alpha_{i\omega})}{\dot{d}_0(\alpha_{i\omega})^2}, \quad \frac{\partial \eta_{0i\omega}}{\partial \omega_i} = \kappa_u \sigma_z.$$

Além disso, se $t = u$, temos que $\partial \eta_{0i\omega} / \partial \omega_i = z_{it} + \sigma_z \omega_i$ e se $t \neq u$, temos que $\partial \eta_{0i\omega} / \partial \omega_i = z_{it}$. Assim,

$$\frac{\partial}{\partial \omega_i} \frac{\partial \eta_{0i\omega}}{\partial \kappa_t} = \begin{cases} \sigma_z, & t = u, \\ 0, & t \neq u, \end{cases}$$

De modo que, se $t \neq u$, temos

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \kappa_t \partial \omega_i} = \left(\frac{\partial^2 \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}^2} \frac{1}{\dot{d}_0(\alpha_{i\omega})} - \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{\ddot{d}_0(\alpha_{i\omega})}{\dot{d}_0(\alpha_{i\omega})^2} \right) \frac{\kappa_u \sigma_z}{\dot{d}_0(\alpha_{i\omega})} z_{it},$$

e se $t = u$, temos que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \kappa_t \partial \omega_i} &= \left(\frac{\partial^2 \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}^2} \frac{1}{\dot{d}_0(\alpha_{i\omega})} - \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{\ddot{d}_0(\alpha_{i\omega})}{\dot{d}_0(\alpha_{i\omega})^2} \right) \frac{\kappa_u \sigma_z}{\dot{d}_0(\alpha_{i\omega})} (z_{it} + \sigma_z \omega_i) \\ &\quad + \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{\sigma_z}{\dot{d}_0(\alpha_{i\omega})}. \end{aligned}$$

Seja

$$-w_{i\omega}^{(0)} = \frac{\partial^2 \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}^2} \frac{1}{\dot{d}_0(\alpha_{i\omega})} - \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{\ddot{d}_0(\alpha_{i\omega})}{\dot{d}_0(\alpha_{i\omega})^2},$$

para $i = 1, \dots, n$, segue que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \kappa_R \partial \omega_i} = \begin{cases} -w_{i\omega}^{(0)} \frac{\kappa_u \sigma_z}{\dot{d}_0(\alpha_{i\omega})} z_{it}, & t \neq u, \\ -w_{i\omega}^{(0)} \frac{\kappa_u \sigma_z}{\dot{d}_0(\alpha_{i\omega})} (z_{it} + \sigma_z \omega_i) + \frac{\partial \ell_i(\alpha_{i\omega})}{\partial \alpha_{i\omega}} \frac{\sigma_z}{\dot{d}_0(\alpha_{i\omega})}, & t = u, \end{cases} \quad (\text{C.10})$$

Dessa forma, avaliando (C.10) em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, segue que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\sigma_z \hat{\kappa}_u \mathbf{Z}^\top \widehat{\mathbf{W}}_0 \widehat{\mathbf{T}}_0 + \sigma_z \mathbf{c}_{\boldsymbol{\kappa}}^u \hat{\boldsymbol{\alpha}}^{*\top} \widehat{\mathbf{A}} \widehat{\mathbf{T}}_0,$$

em que $\mathbf{c}_{\boldsymbol{\kappa}}^u = (c_1, \dots, c_m)^\top$ é um vetor de dimensão $m \times 1$ tal que $c_i = 0$ com $i \neq u$, e $c_i = 1$ para $i = u$, e $\hat{\kappa}_u$ denota o u -ésimo elemento de $\hat{\boldsymbol{\kappa}} = (\kappa_1, \dots, \kappa_m)^\top$, $u = 1, \dots, m$.

Note que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_r \partial \omega_i} &= 0, \quad R = 1, \dots, p, \quad i = 1, \dots, n, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_l \partial \omega_i} &= 0, \quad l = 1, \dots, q, \quad i = 1, \dots, n, \end{aligned}$$

e, portanto,

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{0}_{p,n}, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{0}_{q,n}. \end{aligned}$$

Perturbação individual das covariáveis do componente contínuo

Se perturbamos individualmente as covariáveis da mediana ou da dispersão, o cálculo das derivadas é equivalente ao feito na Seção C.1.2, com a diferença de que a derivada é igual a zero no conjunto $\{i; i \notin \wp\}$ e que

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \kappa_t \partial \omega_i} = 0, \quad \forall t = 1, \dots, m \quad \text{e} \quad \forall i = 1, \dots, n.$$

Dessa forma, perturbando individualmente as covariáveis da mediana, temos

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{0}_{m,n^\dagger}, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\sigma_x \hat{\beta}_j \mathbf{X}^\top \widehat{\mathbf{W}}_1^\dagger \widehat{\mathbf{T}}_1^\dagger + \sigma_x \mathbf{c}_{\boldsymbol{\beta}}^j \hat{\boldsymbol{\mu}}^{*\dagger\top} \widehat{\mathbf{W}}^\dagger \widehat{\mathbf{T}}_1^\dagger, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\sigma_x \hat{\beta}_j \mathbf{S}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger. \end{aligned}$$

Analogamente, se perturbarmos individualmente as covariáveis da dispersão, obtemos

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \mathbf{0}_{m,n^\dagger}, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\sigma_s \hat{\tau}_k \mathbf{X}^\top \widehat{\mathbf{W}}_4^\dagger \widehat{\mathbf{T}}_1^\dagger \widehat{\mathbf{T}}_2^\dagger, \\ \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\sigma_s \hat{\tau}_k \mathbf{S}^\top \widehat{\mathbf{W}}_2^\dagger \widehat{\mathbf{T}}_2^\dagger + \sigma_s \mathbf{c}_{\boldsymbol{\tau}}^k \hat{\boldsymbol{\sigma}}^{*\dagger\top} \widehat{\mathbf{T}}_2^\dagger. \end{aligned}$$

Por brevidade, não apresentaremos o desenvolvimento do cálculo das derivadas para as perturbações simultâneas. Estes podem ser obtidos de forma análoga ao feito para as perturbações individuais.

Apêndice D

Gráficos

D.1 Gráficos de assimetria para as distribuições *power logit*

As Figuras de D.1 a D.7 apresentam gráficos da medida de assimetria $\gamma_{0.125}$ para algumas distribuições na classe *power logit*. Evidentemente, outras formas para $\gamma_{0.125}$ podem ser obtidas através da mudança dos parâmetros das distribuições.

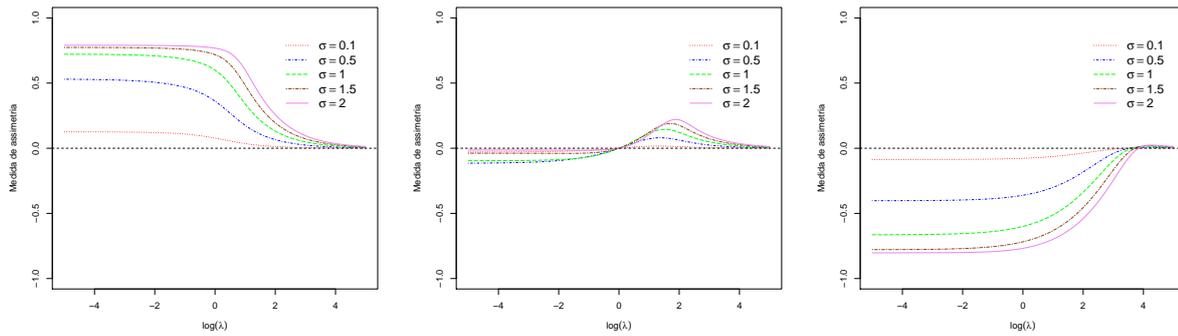


Figura D.1: Medida de assimetria para a distribuição *PL-LO* considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .

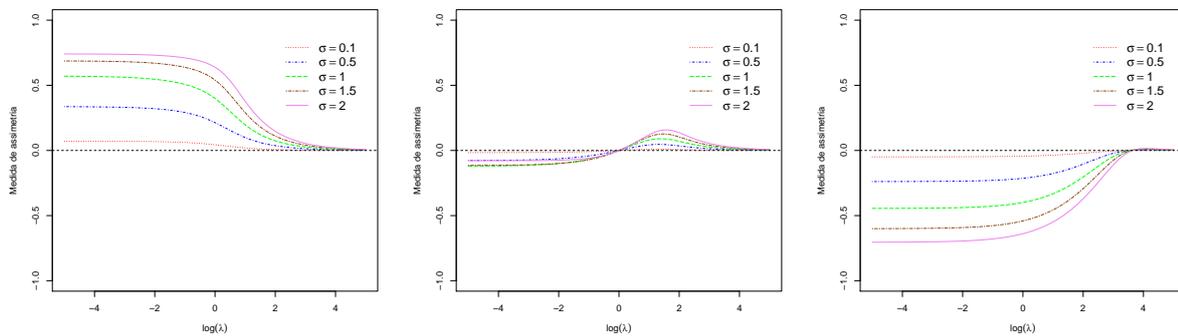


Figura D.2: Medida de assimetria para a distribuição *PL-PE_(1.5)* considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .

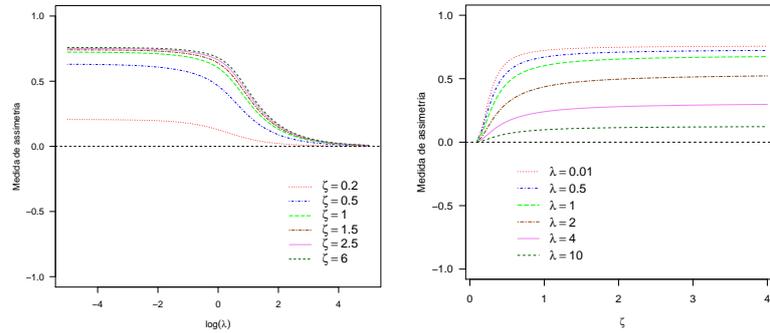


Figura D.3: Medida de assimetria para a distribuição $PL-PE_{(\zeta)}$ considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ .

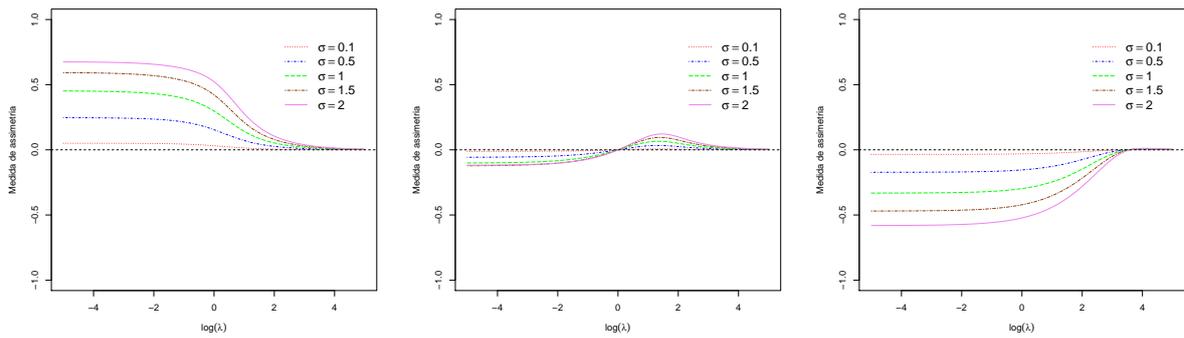


Figura D.4: Medida de assimetria para a distribuição $PL-SN_{(1.5)}$ considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .

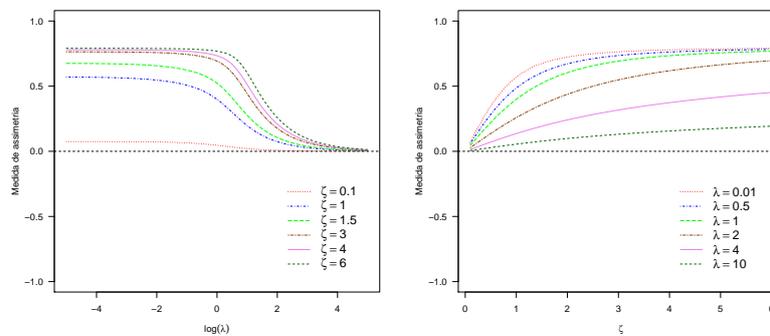


Figura D.5: Medida de assimetria para a distribuição $PL-SN_{(\zeta)}$ considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ .

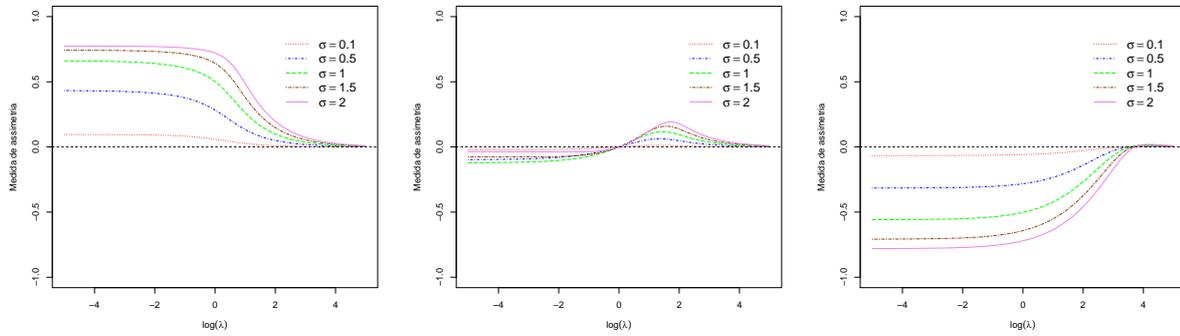


Figura D.6: Medida de assimetria para a distribuição $PL-Hyp_{(1.2)}$ considerando $\mu = 0.1$, $\mu = 0.5$ e $\mu = 0.9$, respectivamente, e variando σ e λ .

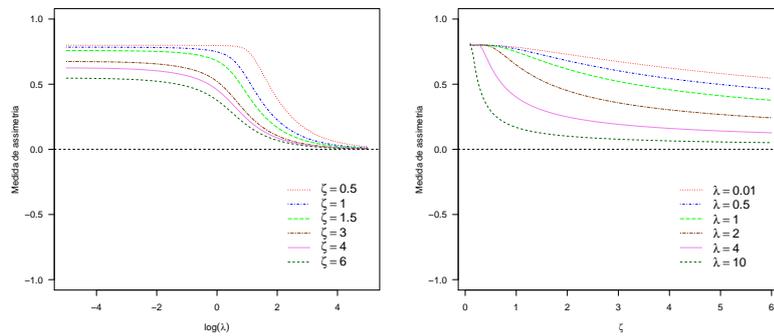


Figura D.7: Medida de assimetria para a distribuição $PL-Hyp_{(\zeta)}$ considerando $\mu = 0.1$, $\sigma = 2$ e variando λ e ζ .

Apêndice E

Resultados de simulações

E.1 Simulações caso iid - Capítulo 2

Performance dos estimadores de máxima verossimilhança penalizada na classe *power logit*

A Tabela E.1 apresenta a média, estimativa do viés e desvio padrão das 3000 estimativas de máxima verossimilhança (usual e penalizada) para alguns modelos da classe *power logit*.

Tabela E.1: Média, estimativa do viés e erro-padrão das 3000 estimativas de máxima verossimilhança (usual e penalizada) para os modelos PL-N, PL-LA e PL-SN_(1.5).

		$n = 30$			$n = 60$			$n = 120$			$n = 240$		
		Média	Viés	EP	Média	Viés	EP	Média	Viés	EP	Média	Viés	EP
PL-N	$\hat{\mu}$	0.70	0.00	0.02	0.70	0.00	0.01	0.70	-0.00	0.01	0.70	0.00	0.01
	$\tilde{\mu}$	0.70	0.00	0.02	0.70	0.00	0.01	0.70	0.00	0.01	0.70	0.00	0.01
	$\hat{\sigma}$	0.60	0.10	0.30	0.56	0.06	0.18	0.52	0.02	0.11	0.51	0.01	0.07
	$\tilde{\sigma}$	0.54	0.04	0.18	0.53	0.03	0.14	0.51	0.01	0.10	0.51	0.01	0.07
	$\hat{\lambda}$	2.17	1.17	2.99	1.62	0.62	1.87	1.26	0.26	1.13	1.10	0.10	0.79
	$\tilde{\lambda}$	1.49	0.49	1.82	1.36	0.36	1.48	1.17	0.17	1.05	1.06	0.06	0.77
PL-LA	$\hat{\mu}$	0.70	0.00	0.02	0.70	0.00	0.01	0.70	0.00	0.01	0.70	0.00	0.01
	$\tilde{\mu}$	0.70	0.00	0.02	0.70	0.00	0.01	0.70	0.00	0.01	0.70	0.00	0.01
	$\hat{\sigma}$	0.64	0.14	0.34	0.59	0.09	0.23	0.55	0.05	0.15	0.52	0.02	0.10
	$\tilde{\sigma}$	0.54	0.04	0.20	0.54	0.04	0.17	0.53	0.03	0.13	0.51	0.01	0.09
	$\hat{\lambda}$	2.60	1.60	3.35	1.93	0.93	2.27	1.50	0.50	1.49	1.24	0.24	0.99
	$\tilde{\lambda}$	1.52	0.52	1.94	1.45	0.45	1.66	1.30	0.30	1.29	1.16	0.16	0.93
PL-SN _(1.5)	$\hat{\mu}$	0.70	0.00	0.01	0.70	0.00	0.01	0.70	0.00	0.01	0.70	0.00	0.00
	$\tilde{\mu}$	0.70	0.00	0.01	0.70	0.00	0.01	0.70	0.00	0.01	0.70	0.00	0.00
	$\hat{\sigma}$	0.53	0.03	0.18	0.53	0.03	0.14	0.52	0.02	0.12	0.51	0.01	0.09
	$\tilde{\sigma}$	0.48	-0.02	0.11	0.50	0.00	0.11	0.51	0.01	0.11	0.51	0.01	0.09
	$\hat{\lambda}$	1.42	0.42	1.77	1.34	0.34	1.49	1.24	0.24	1.29	1.15	0.15	1.02
	$\tilde{\lambda}$	0.83	-0.17	1.10	1.02	0.02	1.17	1.09	0.09	1.15	1.07	0.07	0.97

Estimação do parâmetro extra (ζ)

A Tabela E.2 apresenta a mediana (M) e a distância interquartilica (IR) das $M = 3000$ réplicas referentes à estimação de μ , σ e λ através do método da máxima verossimilhança e ζ minimizando a medida Υ_{ζ} .

Tabela E.2: Mediana (M) e distância interquartilica (IR) das M réplicas para a escolha de ζ .

$\theta = (\mu = 0.3, \sigma = 1, \lambda = 0.5)^\top$								
ζ		$n = 40$		$n = 80$		$n = 150$		
		M	IQ	M	IQ	M	IQ	
PL-Hyp(ζ)	1	μ	0.30	0.09	0.30	0.06	0.30	0.04
		σ	0.98	0.29	0.99	0.22	1.00	0.17
		λ	0.49	0.38	0.51	0.28	0.52	0.21
		ζ	0.94	0.04	0.97	0.03	0.98	0.02
	1.5	μ	0.30	0.07	0.30	0.04	0.30	0.03
		σ	0.98	0.33	0.98	0.24	0.99	0.19
		λ	0.48	0.47	0.48	0.36	0.49	0.27
		ζ	1.41	0.07	1.45	0.05	1.47	0.04
	2	μ	0.30	0.05	0.30	0.04	0.30	0.03
		σ	0.97	0.35	0.99	0.27	0.99	0.20
		λ	0.47	0.54	0.48	0.42	0.49	0.31
		ζ	1.87	0.10	1.93	0.07	1.96	0.05
PL-PE(ζ)	1.5	μ	0.30	0.06	0.30	0.04	0.30	0.03
		σ	0.98	0.34	0.99	0.25	0.99	0.19
		λ	0.49	0.51	0.49	0.38	0.49	0.29
		ζ	1.59	0.62	1.54	0.44	1.53	0.31
	2	μ	0.30	0.06	0.30	0.04	0.30	0.03
		σ	0.98	0.33	0.99	0.23	1.00	0.18
		λ	0.49	0.50	0.50	0.36	0.49	0.26
		ζ	2.08	0.98	2.05	0.73	2.04	0.52
	2.5	μ	0.30	0.06	0.30	0.05	0.30	0.03
		σ	0.98	0.32	0.99	0.23	1.00	0.17
		λ	0.49	0.49	0.50	0.35	0.49	0.25
		ζ	2.57	1.34	2.53	1.03	2.53	0.74
PL-SN(ζ)	1.5	μ	0.30	0.04	0.30	0.03	0.30	0.02
		σ	0.96	0.38	0.98	0.29	0.99	0.22
		λ	0.46	0.64	0.49	0.49	0.48	0.37
		ζ	1.59	0.09	1.55	0.07	1.53	0.06
	2	μ	0.30	0.05	0.30	0.04	0.30	0.03
		σ	0.97	0.35	0.99	0.25	0.99	0.19
		λ	0.49	0.57	0.49	0.42	0.49	0.31
		ζ	2.12	0.13	2.06	0.10	2.03	0.08
	3	μ	0.30	0.07	0.30	0.05	0.30	0.04
		σ	0.97	0.30	0.99	0.21	0.99	0.16
		λ	0.49	0.49	0.50	0.35	0.49	0.26
		ζ	3.17	0.20	3.10	0.16	3.05	0.13
PL-t(ζ)	2	μ	0.30	0.08	0.30	0.05	0.30	0.04
		σ	1.00	0.36	1.03	0.27	1.05	0.20
		λ	0.49	0.42	0.56	0.33	0.58	0.25
		ζ	1.93	0.35	2.00	0.25	2.03	0.19
	4	μ	0.30	0.07	0.30	0.05	0.30	0.04
		σ	0.97	0.35	0.99	0.27	1.00	0.19
		λ	0.45	0.49	0.49	0.37	0.50	0.28
		ζ	3.41	0.71	3.67	0.54	3.83	0.41
	6	μ	0.30	0.07	0.30	0.05	0.30	0.04
		σ	0.97	0.34	0.99	0.25	1.00	0.19
		λ	0.45	0.51	0.48	0.37	0.50	0.28
		ζ	4.66	1.16	5.20	0.89	5.54	0.65

E.2 Simulações Capítulo 6

Nesta seção, apresentamos os resultados de simulação considerando os estimadores LSMLE e LMDPDE.

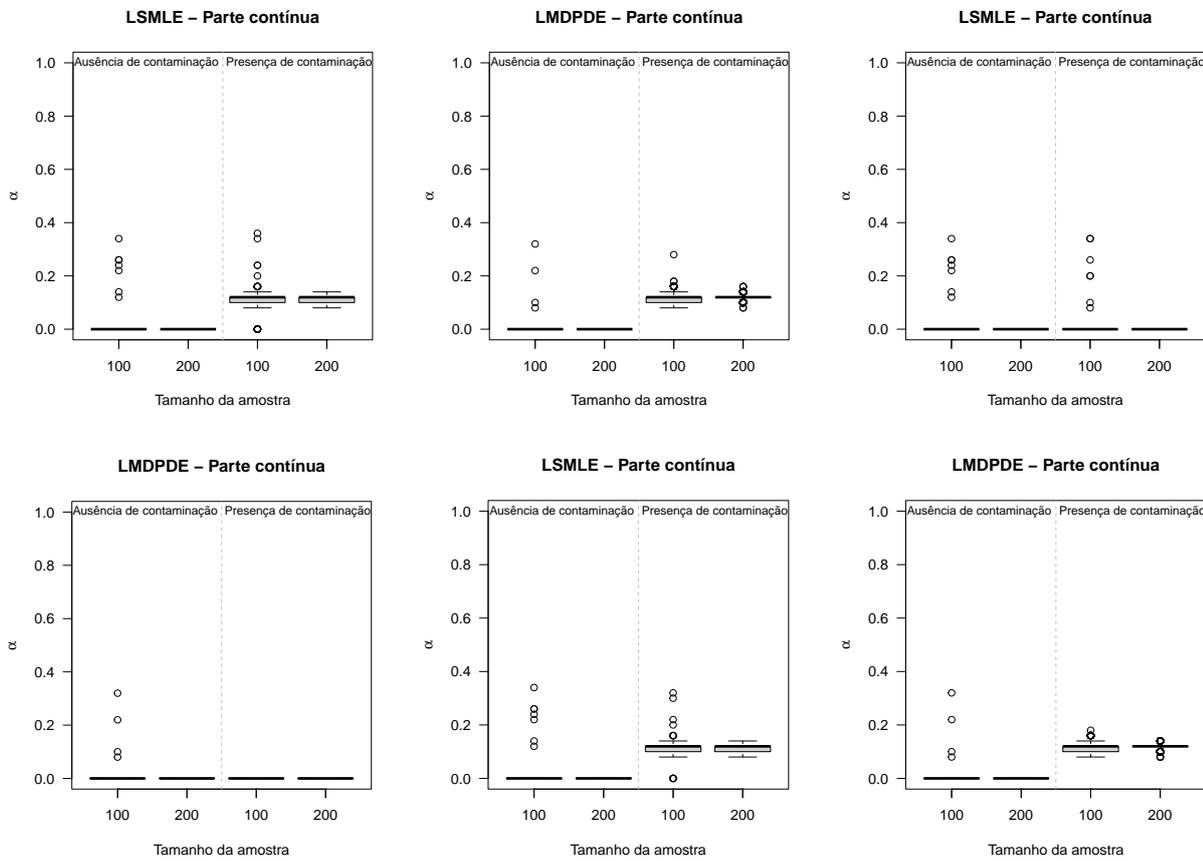


Figura E.1: Boxplots dos valores ótimos para a constante de afinação α para a parte contínua considerando o LSMLE e o LMDPDE, sob o cenário 1 (primeira linha), cenário 2 (segunda linha) e cenário 3 (terceira linha).

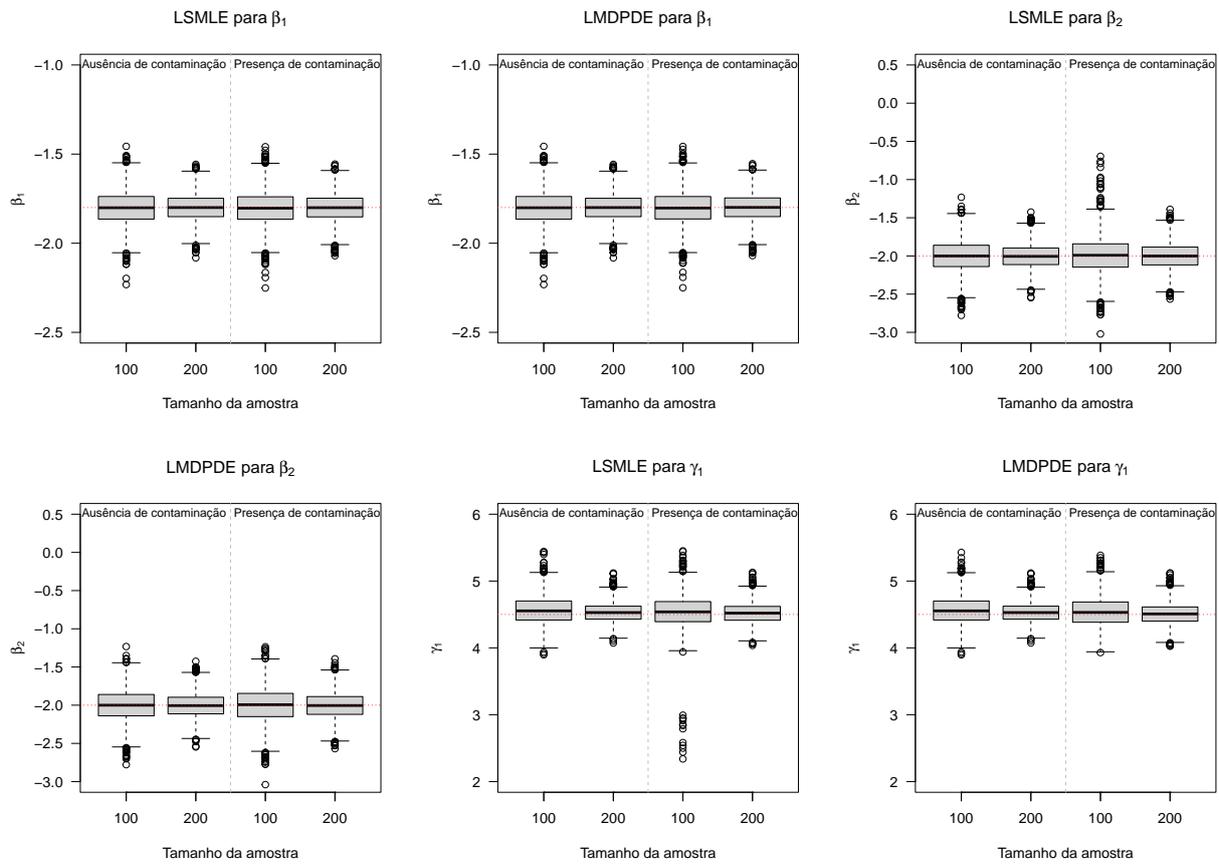


Figura E.2: *Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 1: MLE (esquerda), LSMLE (centro) e LMDPDE (direita).*

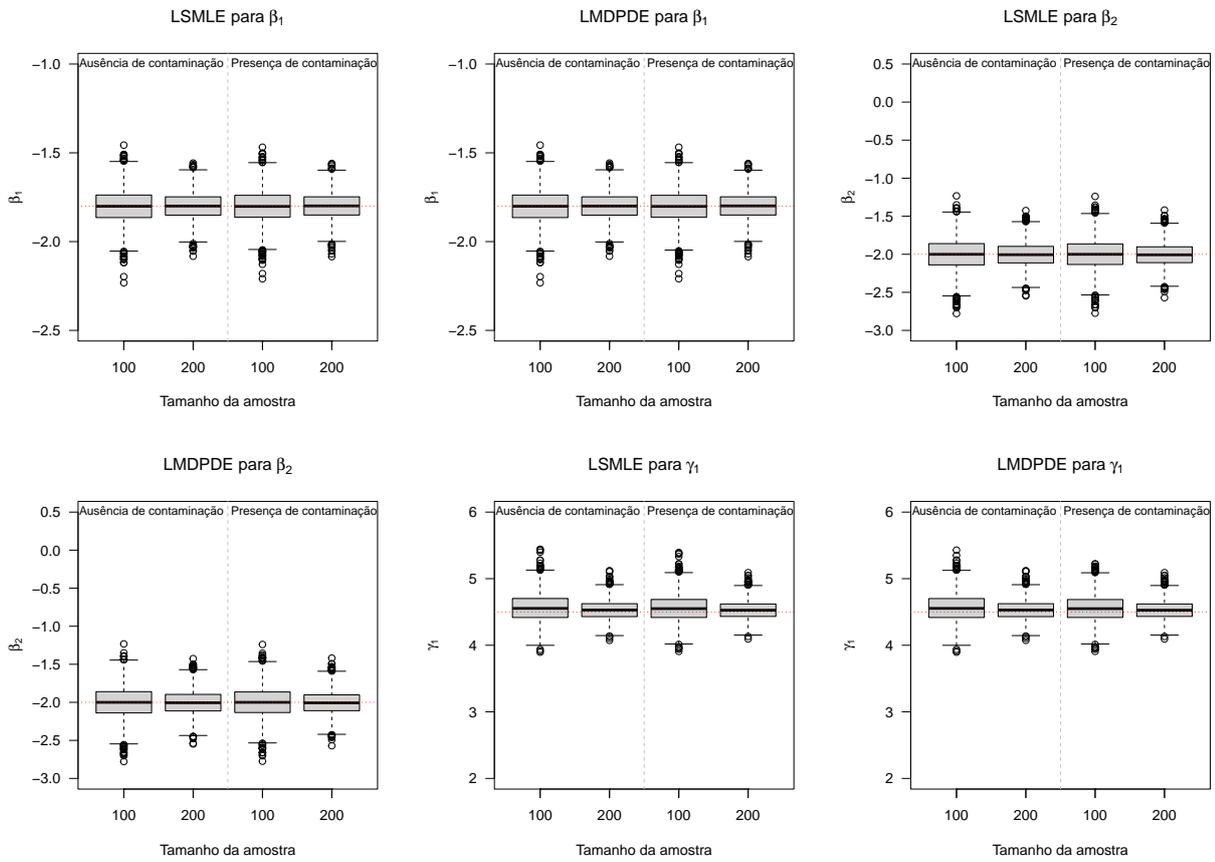


Figura E.3: Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 2: MLE (esquerda), LSMLE (centro) e LMDPDE (direita).

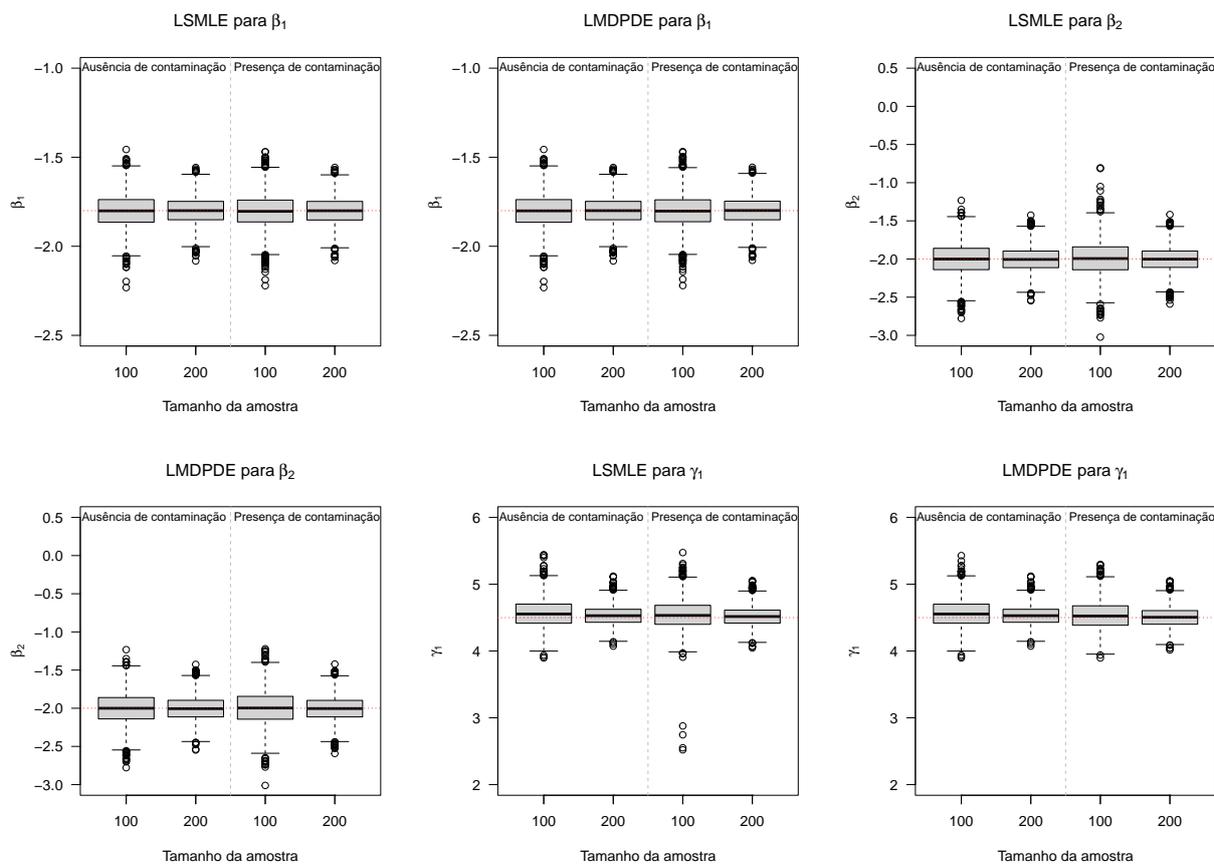


Figura E.4: Boxplots das estimativas dos parâmetros β_1 , β_2 e γ_1 sob o cenário 3: MLE (esquerda), LSMLE (centro) e LMDPDE (direita).

Apêndice F

O modelo de regressão log-log

Antes de apresentarmos o desenvolvimento do modelo de regressão log-log, vamos verificar que os parâmetros desta classe de distribuições, apresentada na Definição 2.3, são interpretados em termos da mediana e da dispersão da variável resposta. Para isso, considere a seguinte proposição.

Proposição F.1 *Seja $Y \sim \text{log-log}(\mu, \sigma; r)$. O quantil de ordem u da distribuição de Y é dado por*

$$y_u = \mu^{\exp\{-\sigma z_u\}}$$

e z_u é o quantil de ordem u de $S \sim S(0, 1; r)$.

Consequentemente, $y_{0.5} = \mu$, isto é, μ é a mediana da distribuição de Y . Para verificarmos que σ é parâmetro de dispersão, vamos utilizar o mesmo argumento feito para a classe *power logit*. Dessa forma, se $Y_1 \sim \text{log-log}(\mu, \sigma_1; r)$ e $Y_2 \sim \text{log-log}(\mu, \sigma_2; r)$, então precisamos mostrar que Y_1 é menor que Y_2 em ordem de dispersão quantílica se, e somente se, $\sigma_1 \leq \sigma_2$. Considere $y_u = \mu^{\exp\{-\sigma z_u\}} = \delta^{\exp\{\sigma z_u\}}$, com $\delta = \mu^{-1} > 1$.

Temos que

$$QS_Y(u) = \delta^{\exp\{\sigma z_{1-u}\}} - \delta^{\exp\{\sigma z_u\}},$$

em que $u \in (0, 1/2)$. Verifiquemos se $QS_{Y_1}(u) - QS_{Y_2}(u) \leq 0$ se, e só se, $\sigma_1 \leq \sigma_2$. Note que

$$\begin{aligned} QS_{Y_1}(u) - QS_{Y_2}(u) &= \delta^{\exp\{\sigma_1 z_{1-u}\}} - \delta^{\exp\{\sigma_1 z_u\}} \\ &\quad - \left[\delta^{\exp\{\sigma_2 z_{1-u}\}} - \delta^{\exp\{\sigma_2 z_u\}} \right]. \end{aligned}$$

Como $Z \sim S(0, 1; r)$ e $u \in (0, 1/2)$, então $z_u < 0$ e $z_{1-u} > 0$, e $z_{1-u} = -z_u$. Assim,

$$\begin{aligned} QS_{Y_1}(u) - QS_{Y_2}(u) &= \delta^{\exp\{\sigma_1 z_{1-u}\}} - \sigma^{\exp\{\sigma_2 z_{1-u}\}} \\ &\quad - \left[\delta^{\exp\{-\sigma_1 z_{1-u}\}} - \delta^{\exp\{-\sigma_2 z_{1-u}\}} \right]. \end{aligned}$$

Agora, note que

$$\delta^{\exp\{\sigma_1 z_{1-u}\}} - \sigma^{\exp\{\sigma_2 z_{1-u}\}} \leq 0$$

e

$$\delta^{\exp\{-\sigma_1 z_{1-u}\}} - \delta^{\exp\{-\sigma_2 z_{1-u}\}} \geq 0$$

se, e só se, $\sigma_1 \leq \sigma_2$. Logo, o σ é parâmetro de dispersão.

F.1 Estimação

Conforme apresentado na Seção 3.5, função de log-verossimilhança para $\boldsymbol{\theta}$ é dada por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\mu_i, \sigma_i, \lambda),$$

sendo $\ell_i(\mu_i, \sigma_i) = -\log[-\sigma_i y_i \log(y_i)] + \log\{r(z_i^2)\}$. Dessa forma, cálculo dos componentes do vetor escore é feito como segue.

$$U_R = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \beta_R} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \sigma_i)}{\partial \mu_i} \frac{d\mu_i}{d\eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_R}, \quad R = 1, \dots, p,$$

$$U_l = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \tau_l} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \sigma_i)}{\partial \sigma_i} \frac{d\sigma_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_l}, \quad l = 1, \dots, q,$$

em que

$$\frac{\partial \ell_i(\mu_i, \sigma_i)}{\partial \mu_i} = \frac{1}{\sigma_i \mu_i (-\log \mu_i)} z_i v(z_i), \quad \frac{\partial \ell_i(\mu_i, \sigma_i)}{\partial \sigma_i} = \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1].$$

Além disso, temos que

$$\frac{d\mu_i}{d\eta_{1i}} = \frac{1}{\dot{d}_1(\mu_i)} = e \quad \text{e} \quad \frac{d\sigma_i}{d\eta_{2i}} = \frac{1}{\dot{d}_2(\sigma_i)}.$$

Assim, uma vez que $\partial \eta_{1i} / \partial \beta_R = x_{iR}$ e $\partial \eta_{2i} / \partial \tau_l = s_{il}$, segue que

$$U_R = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \beta_R} = \sum_{i=1}^n \frac{1}{\sigma_i \mu_i (-\log \mu_i)} z_i v(z_i) \frac{1}{\dot{d}_1(\mu_i)} x_{iR}, \quad R = 1, \dots, p, \quad (\text{F.1})$$

$$U_l = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \tau_l} = \sum_{i=1}^n \frac{1}{\sigma_i} [z_i^2 v(z_i) - 1] \frac{1}{\dot{d}_2(\sigma_i)} s_{il}, \quad l = 1, \dots, q. \quad (\text{F.2})$$

Note que

$$\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \beta_R} = \lim_{\lambda \rightarrow 0^+} \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \beta_R} \quad \text{e} \quad \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \tau_l} = \lim_{\lambda \rightarrow 0^+} \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)}{\partial \tau_l}.$$

Dessa forma, definindo as quantidades necessárias, obtemos a expressão do vetor escore de $\boldsymbol{\theta}$ em forma matricial. Para o cálculo da matriz da informação observada, utilizamos as mesmas expressões apresentadas na Seção B.2, alterando apenas as derivadas da função de log-verossimilhança com relação a μ_i e σ_i . As derivadas de segunda ordem são apresentadas a seguir.

$$\begin{aligned} \frac{\partial \ell_i^2(\mu_i, \sigma_i)}{\partial \mu_i^2} &= \frac{1}{[\sigma_i \mu_i (-\log \mu_i)]^2} \{ \sigma_i (1 + \log \mu_i) z_i v(z_i) - v(z_i) - z_i v'(z_i) \}, \\ \frac{\partial \ell_i^2(\mu_i, \sigma_i)}{\partial \sigma_i^2} &= \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^2} z_i^2 [3v(z_i) + z_i v'(z_i)], \\ \frac{\partial \ell_i^2(\mu_i, \sigma_i)}{\partial \mu_i \partial \sigma_i} &= \frac{-1}{\sigma_i^2 \mu_i (-\log \mu_i)} z_i [2v(z_i) + z_i v'(z_i)]. \end{aligned}$$

Adicionalmente, a matriz da informação de Fisher de $\boldsymbol{\theta}$ para o modelo de regressão log-log pode ser escrita como

$$\mathbf{K}_n(\boldsymbol{\beta}, \boldsymbol{\tau}) = \begin{bmatrix} \mathbf{X}^\top \mathbf{W}(\boldsymbol{\mu}) \mathbf{X} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{S}^\top \mathbf{W}(\boldsymbol{\sigma}) \mathbf{S} \end{bmatrix},$$

em que $\mathbf{W}(\boldsymbol{\mu}) = \text{diag}\{d_r(\zeta)[\mu_1^*/\dot{d}_1(\mu_1)]^2, \dots, d_r(\zeta)[\mu_n^*/\dot{d}_1(\mu_n)]^2\}$ e $\mathbf{W}(\boldsymbol{\sigma}) = \text{diag}\{(f_r(\zeta) - 1)[1/\sigma_1 \dot{d}_2(\sigma_1)]^2, \dots, (f_r(\zeta) - 1)[1/\sigma_n \dot{d}_2(\sigma_n)]^2\}$. As definições de $d_r(\zeta)$ e $f_r(\zeta)$ foram apresentadas

na Seção 2.3.

Finalmente, os modelos de regressão log-log também podem depender de parâmetro(s) extra(s), denotado(s) aqui por ζ . Esse parâmetro (ou vetor de parâmetros) será estimado, também, através da minimização da medida Υ_ζ , que foi definida na Seção 2.3.2.

F.2 Diagnóstico

Com relação as medidas de diagnóstico, os métodos desenvolvidos para o modelo de regressão *power logit* podem ser adaptados de forma natural para o modelo de regressão log-log. Assim, podemos definir os resíduos quantílico, deviance e padronizado como segue.

Resíduo quantílico: O resíduo quantílico para o modelo de regressão log-log é dado por

$$r_i^q = \Phi^{-1}[R(\hat{z}_i)], \quad i = 1, \dots, n,$$

em que $R(\cdot)$ é a função de distribuição acumulada de Z . Note que, aqui, $z_i = \sigma_i^{-1}\{-\log(-\log y_i) - [-\log(-\log \mu_i)]\}$.

Resíduo deviance: Após algumas simplificações, o resíduo deviance r_i^d pode ser escrito como

$$r_i^d = \text{sign}(\hat{z}_i) \left\{ 2 \log \left[\frac{r(0)}{r(\hat{z}_i^2)} \right] \right\}^{1/2}, \quad i = 1, \dots, n.$$

Resíduo padronizado: Assim como no caso do modelo de regressão *power logit*, aqui nós também identificamos a equivalência entre o modelo de regressão log-log e o modelo de regressão não-linear simétrico heterocedástico. Basta notar que podemos escrever o modelo da forma

$$y_i^\dagger = \mu_i^\dagger(\boldsymbol{\beta}; \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

sendo $\varepsilon_i \sim S(0, \sigma_i^2; r)$, $y_i^\dagger = -\log(-\log y_i)$ e $\mu_i^\dagger(\boldsymbol{\beta}; \mathbf{x}_i) = \mu_i^\dagger = -\log(-\log \mu_i)$ uma função não linear de $\boldsymbol{\beta}$ contínua e duas vezes diferenciável, com matriz de derivadas $\mathbf{D} = \partial \boldsymbol{\mu}^\dagger / \partial \boldsymbol{\beta} = \sigma_i \boldsymbol{\mu}^* \mathbf{T}_1 \mathbf{X}$ (com posto $p < n$) para todo $\boldsymbol{\beta}$. Dessa forma, o resíduo padronizado fica dado por

$$r_i^p = \frac{\hat{y}_i^\dagger - \hat{\mu}_i^\dagger}{\hat{\sigma}_i \sqrt{\hat{\xi}_r \{1 - (\hat{d}_r \hat{\xi}_r)^{-1} \hat{h}_{ii}\}}}, \quad i = 1, \dots, n,$$

sendo \hat{h}_{ii} o i -ésimo elemento da diagonal de $\mathbf{H} = \boldsymbol{\Sigma}^{-1/2} \mathbf{D} (\mathbf{D}^\top \boldsymbol{\Sigma}^{-1} \mathbf{D})^{-1} \mathbf{D}^\top \boldsymbol{\Sigma}^{-1/2}$ avaliado na estimativa de máxima verossimilhança.

De forma análoga, a matriz de alavanca generalizada é dada pro

$$\text{GL}(\boldsymbol{\theta}) = \dot{\mathbf{L}}_\beta \mathbf{J}_n^{-1} \ddot{\mathbf{L}}_{\boldsymbol{\theta}y},$$

em que $\dot{\mathbf{L}}_\beta = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}$ e $\ddot{\mathbf{L}}_{\boldsymbol{\theta}y} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \mathbf{y}^\top$. Após algumas contas, similares aquelas feitas para o modelo de regressão *power logit*, temos que $\dot{\mathbf{L}}_\beta = [\mathbf{T}_1 \mathbf{X}, \mathbf{0}_{n,q}]$ e

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_R \partial y_j} &= \frac{1}{\sigma_j \mu_j (-\log \mu_j)} \frac{1}{\dot{d}_1(\mu_j)} x_{jR} \frac{1}{\sigma_j y_j (-\log y_j)} [v(z_j) + z_j v'(z_j)], \quad R = 1, \dots, p, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \tau_l \partial y_j} &= \frac{1}{\dot{d}_2(\sigma_j)} s_{jl} \frac{1}{\sigma_j y_j (-\log y_j)} \frac{1}{\sigma_j} z_j [2v(z_j) + z_j v'(z_j)], \quad l = 1, \dots, q. \end{aligned}$$

Assim,

$$\ddot{\mathbf{L}}_{\boldsymbol{\theta}y} = \begin{bmatrix} \mathbf{X}^\top \mathbf{T}_1 \mathbf{D}_\beta \mathbf{D}_y \mathbf{W}_\beta \\ \mathbf{S}^\top \mathbf{T}_2 \mathbf{D}_y \mathbf{W}_\tau \end{bmatrix}.$$

Apêndice G

Cálculo das matrizes de variâncias e covariâncias dos estimadores robustos

G.1 Cálculo da matriz de variâncias e covariâncias para o MDPDE no modelo de regressão para dados binários

Conforme apresentado na Seção 6.2.2, a matriz de variâncias e covariâncias do MDPDE para $\boldsymbol{\kappa}$ é dada por

$$\mathbf{V}_\alpha(\boldsymbol{\kappa}) = \mathbf{A}_\alpha^{-1}(\boldsymbol{\kappa})\mathbf{B}_\alpha(\boldsymbol{\kappa})\mathbf{A}_\alpha^{-1}(\boldsymbol{\kappa}),$$

com

$$\mathbf{A}_\alpha(\boldsymbol{\kappa}) = \sum_{i=1}^n \mathbb{E} \left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \right] \quad \text{e} \quad \mathbf{B}_\alpha(\boldsymbol{\kappa}) = \sum_{i=1}^n \mathbb{E} \left[\left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right)^\top \right].$$

Inicialmente, vamos obter a matriz $\mathbf{A}_\alpha(\boldsymbol{\kappa})$. Para isso, observe que

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\kappa}^\top} \left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) &= \frac{\partial}{\partial \vartheta_i} \left(\frac{\partial}{\partial \vartheta_i} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \frac{1}{g'_\vartheta(\vartheta_i)} \right) \frac{1}{g'_\vartheta(\vartheta_i)} \mathbf{S}_i^\top \mathbf{S}_i \\ &= \left[\left(\frac{\partial^2}{\partial \vartheta_i^2} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \frac{1}{g'_\vartheta(\vartheta_i)} + \left(\frac{\partial}{\partial \vartheta_i} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \frac{\partial}{\partial \vartheta_i} \frac{1}{g'_\vartheta(\vartheta_i)} \right] \frac{1}{g'_\vartheta(\vartheta_i)} \mathbf{S}_i^\top \mathbf{S}_i. \end{aligned}$$

Dessa forma,

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \left(\frac{\partial}{\partial \boldsymbol{\kappa}} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \right] &= \mathbb{E} \left[\left(\frac{\partial^2}{\partial \vartheta_i^2} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \frac{1}{g'_\vartheta(\vartheta_i)} + \left(\frac{\partial}{\partial \vartheta_i} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right) \frac{\partial}{\partial \vartheta_i} \frac{1}{g'_\vartheta(\vartheta_i)} \right] \frac{1}{g'_\vartheta(\vartheta_i)} \mathbf{S}_i^\top \mathbf{S}_i \\ &= \mathbb{E} \left[\frac{\partial^2}{\partial \vartheta_i^2} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) \right] \frac{1}{[g'_\vartheta(\vartheta_i)]^2} \mathbf{S}_i^\top \mathbf{S}_i. \end{aligned}$$

Após alguma álgebra, obtemos

$$\begin{aligned} \frac{\partial^2}{\partial \vartheta_i^2} \mathcal{V}_i(Y_i^c; \boldsymbol{\kappa}) &= \frac{(1 + \alpha)}{\vartheta_i(1 - \vartheta_i)} \left\{ \frac{(2\vartheta_i - 1)}{\vartheta_i(1 - \vartheta_i)} [(1 - \vartheta_i)\vartheta_i^{\alpha+1} - \vartheta_i(1 - \vartheta_i)^{\alpha+1} - (y_i^c - \vartheta_i)f_\kappa(y_i^c; \vartheta_i)^\alpha] \right. \\ &\quad + (2\vartheta_i - 1)[(1 - \vartheta_i)^\alpha - \vartheta_i^\alpha] + \alpha[(1 - \vartheta_i)\vartheta_i^\alpha + \vartheta_i(1 - \vartheta_i)^\alpha] + f_\kappa(y_i^c; \vartheta_i)^\alpha \\ &\quad \left. - \frac{\alpha}{\vartheta_i(1 - \vartheta_i)} (y_i^c - \vartheta_i)^2 f_\kappa(y_i^c; \vartheta_i)^\alpha \right\}. \end{aligned}$$

Usando o fato de que

$$\begin{aligned}\mathbb{E}[f_{\kappa}(Y_i^c; \vartheta_i)^\alpha] &= \vartheta_i^{1+\alpha} + (1 - \vartheta_i)^{1+\alpha}, \\ \mathbb{E}[(Y_i^c - \vartheta_i)f_{\kappa}(Y_i^c; \vartheta_i)^\alpha] &= (1 - \vartheta_i)\vartheta_i^{1+\alpha} - \vartheta_i(1 - \vartheta_i)^{1+\alpha}, \\ \mathbb{E}[(Y_i^c - \vartheta_i)^2 f_{\kappa}(Y_i^c; \vartheta_i)^\alpha] &= (1 - \vartheta_i)^2 \vartheta_i^{1+\alpha} + \vartheta_i^2 (1 - \vartheta_i)^{1+\alpha},\end{aligned}\tag{G.1}$$

pode-se verificar que

$$\mathbb{E}\left[\frac{\partial^2}{\partial \vartheta_i^2} \mathcal{V}_i(Y_i^c; \kappa)\right] = \frac{(1 + \alpha)}{\vartheta_i(1 - \vartheta_i)} [(1 - \vartheta_i)\vartheta_i^\alpha + \vartheta_i(1 - \vartheta_i)^\alpha].$$

Dessa forma, fazendo as devidas substituições, chegamos na expressão em (6.20).

Agora, para obtermos a matriz $\mathbf{B}_\alpha(\kappa)$, notemos que

$$\mathbb{E}\left[\left(\frac{\partial}{\partial \kappa} \mathcal{V}_i(Y_i^c; \kappa)\right) \left(\frac{\partial}{\partial \kappa} \mathcal{V}_i(Y_i^c; \kappa)\right)^\top\right] = \mathbb{E}\left[\left(\frac{\partial}{\partial \vartheta_i} \mathcal{V}_i(Y_i^c; \kappa)\right)^2\right] \frac{1}{[g'_\vartheta(\vartheta_i)]^2} \mathbf{S}_i^\top \mathbf{S}_i.$$

Como

$$\begin{aligned}\left(\frac{\partial}{\partial \vartheta_i} \mathcal{V}_i(Y_i^c; \kappa)\right)^2 &= \frac{(1 + \alpha)^2}{[\vartheta_i(1 - \vartheta_i)]^2} \left\{ [(1 - \vartheta_i)\vartheta_i^{1+\alpha} - \vartheta_i(1 - \vartheta_i)^{1+\alpha}]^2 \right. \\ &\quad - 2 [(1 - \vartheta_i)\vartheta_i^{1+\alpha} - \vartheta_i(1 - \vartheta_i)^{1+\alpha}] (y_i^c - \vartheta_i) f_{\kappa}(y_i^c; \vartheta_i)^\alpha \\ &\quad \left. + (y_i^c - \vartheta_i)^2 f_{\kappa}(y_i^c; \vartheta_i)^{2\alpha} \right\},\end{aligned}$$

usando (G.1) e o fato de que $\mathbb{E}[(Y_i^c - \vartheta_i)^2 f_{\kappa}(Y_i^c; \vartheta_i)^{2\alpha}] = (1 - \vartheta_i)^2 \vartheta_i^{1+2\alpha} + \vartheta_i^2 (1 - \vartheta_i)^{1+2\alpha}$, chegamos a

$$\mathbb{E}\left[\left(\frac{\partial}{\partial \vartheta_i} \mathcal{V}_i(Y_i^c; \kappa)\right)^2\right] = \frac{(1 + \alpha)^2}{\vartheta_i(1 - \vartheta_i)} [(1 - \vartheta_i)\vartheta_i^\alpha + \vartheta_i(1 - \vartheta_i)^\alpha]^2.$$

G.2 Cálculo da matriz de variâncias e covariâncias para o LSMLE

Nesta seção, calculamos as matrizes $\mathbf{J}_\alpha(\theta)$ e $\mathbf{K}_\alpha(\theta)$, apresentadas em (6.21) e (6.22), respectivamente. Temos

$$\mathbf{J}_\alpha(\theta) = \mathbb{E}\left\{\sum_{i \in \varphi} \frac{\partial}{\partial \theta^\top} [\mathbf{U}^*(\mathcal{Y}_i^*; \theta) h_{\theta}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha]\right\}$$

e

$$\mathbf{K}_\alpha(\theta) = \mathbb{E}\left\{\sum_{i \in \varphi} \mathbf{U}^*(\mathcal{Y}_i^*; \theta) \mathbf{U}^*(\mathcal{Y}_i^*; \theta)^\top h_{\theta}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha}\right\}.$$

Note que podemos particionar o vetor escore modificado como sendo $\mathbf{U}^*(y_i^*; \theta) = (\mathbf{U}_\beta^*(y_i^*; \theta)^\top, \mathbf{U}_\gamma^*(y_i^*; \theta)^\top)^\top$. De forma análoga, podemos particionar as matrizes $\mathbf{J}_\alpha(\theta)$ e $\mathbf{K}_\alpha(\theta)$ da seguinte forma

$$\mathbf{J}_\alpha(\theta) = \begin{bmatrix} \mathbf{J}_\alpha^{\beta\beta} & \mathbf{J}_\alpha^{\beta\gamma} \\ \mathbf{J}_\alpha^{\gamma\beta} & \mathbf{J}_\alpha^{\gamma\gamma} \end{bmatrix} \quad \text{e} \quad \mathbf{K}_\alpha(\theta) = \begin{bmatrix} \mathbf{K}_\alpha^{\beta\beta} & \mathbf{K}_\alpha^{\beta\gamma} \\ \mathbf{K}_\alpha^{\gamma\beta} & \mathbf{K}_\alpha^{\gamma\gamma} \end{bmatrix}.$$

No que segue, tomamos $i \in \wp$. Assim, temos que

$$\begin{aligned} \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta}) &= \frac{\partial}{\partial \mu_i} \log(h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)) \frac{1}{g'_\mu(\mu_i)} \boldsymbol{x}_i \\ &= \phi_i (y_i^* - \mu_i^*) \frac{1}{g'_\mu(\mu_i)} \boldsymbol{x}_i, \end{aligned}$$

e

$$\begin{aligned} \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta}) &= \frac{\partial}{\partial \phi_i} \log(h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)) \frac{\partial \phi_i}{\partial \phi'_{i,1+\alpha}} \frac{1}{g'_\phi(\phi'_{i,1-\alpha})} \boldsymbol{z}_i \\ &= [\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)] (1 - \alpha)^{-1} \frac{1}{g'_\phi(\phi'_{i,1-\alpha})} \boldsymbol{z}_i. \end{aligned}$$

Inicialmente, obteremos a matriz $\mathbf{J}_\alpha(\boldsymbol{\theta})$. Para tanto, note que

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}^\top} [\mathbf{U}^*(y_i^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(Y_i^*; \mu_i, \phi_i)^\alpha] &= \left[\frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \right] h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)^\alpha \\ &\quad + \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}^\top} \exp\{\alpha \log(h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i))\} \\ &= \left[\frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \right] h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)^\alpha \\ &\quad + \alpha \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}^*(y_i^*; \boldsymbol{\theta})^\top h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)^\alpha. \end{aligned}$$

Agora, precisamos calcular $(\partial/\partial \boldsymbol{\theta}) \mathbf{U}^*(y_i^*; \boldsymbol{\theta})$. Após alguma álgebra, podemos verificar que

$$\begin{aligned} \frac{\partial}{\partial \beta^\top} \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta}) &= -\frac{\phi_i}{[g'_\mu(\mu_i)]^2} \left[\phi v_i + (y_i^* - \mu_i^*) \frac{g''_\mu(\mu_i)}{g'_\mu(\mu_i)} \right] \boldsymbol{x}_i^\top \boldsymbol{x}_i, \\ \frac{\partial}{\partial \gamma^\top} \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta}) &= -\frac{(1 - \alpha)^{-1}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} \left\{ (1 - \alpha)^{-1} d_i + \frac{g''_\phi(\phi'_{i,1+\alpha})}{g'_\phi(\phi'_{i,1-\alpha})} [\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger)] \right\} \boldsymbol{z}_i^\top \boldsymbol{z}_i, \\ \frac{\partial}{\partial \beta^\top} \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta}) &= -\frac{(1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} f_i \boldsymbol{z}_i^\top \boldsymbol{x}_i, \end{aligned}$$

em que $f_i = c_i - (y_i^* - \mu_i^*)$. Além disso, precisamos encontrar $\mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}^*(y_i^*; \boldsymbol{\theta})^\top$. Desse modo,

$$\begin{aligned} \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta})^\top &= \phi_i^2 \frac{(y_i^* - \mu_i^*)^2}{[g'_\mu(\mu_i)]^2} \boldsymbol{x}_i^\top \boldsymbol{x}_i, \\ \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta})^\top &= \frac{(1 - \alpha)^{-2}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger) \right]^2 \boldsymbol{z}_i^\top \boldsymbol{z}_i, \\ \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta})^\top &= \frac{\phi_i (1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} (y_i^* - \mu_i^*) \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger) \right] \boldsymbol{z}_i^\top \boldsymbol{x}_i, \end{aligned}$$

Agora, calcularemos os valores esperados. Para tanto, utilizaremos a seguinte proposição.

Proposição G.1 *Sejam $\mathcal{Y}_1, \dots, \mathcal{Y}_n$ variáveis aleatórias independentes com $\mathcal{Y}_i \sim BI_c(\vartheta_i, \mu_i, \phi_i)$, para $i = 1, \dots, n$. Se $\mathfrak{S} : (0, 1) \rightarrow \mathbb{R}$ é uma função contínua, então*

$$\mathbb{E} \left(\sum_{i \in \wp} \mathfrak{S}(\mathcal{Y}_i) \right) = \sum_{i=1}^n (1 - \vartheta_i) \mathbb{E} \left(\mathfrak{S}(\mathcal{Y}_i) \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right).$$

Demonstração. Defina

$$\mathfrak{S}^*(y_i) = \begin{cases} 0, & y_i = c, \\ \mathfrak{S}(y_i), & y_i \in (0, 1). \end{cases}$$

Assim,

$$\mathbb{E} \left(\sum_{i \in \mathcal{P}} \mathfrak{S}(\mathcal{Y}_i) \right) = \mathbb{E} \left(\sum_{i=1}^n \mathfrak{S}^*(\mathcal{Y}_i) \right) = \sum_{i=1}^n \mathbb{E} (\mathfrak{S}^*(\mathcal{Y}_i)).$$

Agora

$$\begin{aligned} \mathbb{E} (\mathfrak{S}^*(\mathcal{Y}_i)) &= \mathbb{E} (\mathfrak{S}^*(\mathcal{Y}_i) \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0) \mathbb{P} (\mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0) + \mathbb{E} (\mathfrak{S}^*(\mathcal{Y}_i) \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 1) \mathbb{P} (\mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 1) \\ &= \mathbb{E} (\mathfrak{S}^*(\mathcal{Y}_i) \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0) \mathbb{P} (\mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0) = (1 - \vartheta_i) \mathbb{E} (\mathfrak{S}(\mathcal{Y}_i) \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0). \end{aligned}$$

■

Utilizando a Proposição G.1, temos que

$$\begin{aligned} \mathbf{J}_\alpha(\boldsymbol{\theta}) &= \mathbb{E} \left\{ \sum_{i \in \mathcal{P}} \frac{\partial}{\partial \boldsymbol{\theta}^\top} [\mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha] \right\} \\ &= \sum_{i=1}^n (1 - \vartheta_i) \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} [\mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha] \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\}, \end{aligned}$$

mas, para $i = 1, \dots, n$,

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} [\mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha] \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \right] h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)^\alpha h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i) dy_i^* \\ &\quad + \alpha \int_{-\infty}^{\infty} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}^*(y_i^*; \boldsymbol{\theta})^\top h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)^\alpha h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i) dy_i^* \\ &= b'_{i,1} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \right] h_{\boldsymbol{\theta}}^*(Y_i^*; \mu_i, \phi_i) dy_i^* \\ &\quad + \alpha b'_{i,1} \int_{-\infty}^{\infty} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}^*(y_i^*; \boldsymbol{\theta})^\top h_{\boldsymbol{\theta}}^*(Y_i^*; \mu_i, \phi_i) dy_i^* \\ &= b'_{i,1} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &\quad + b'_{i,1} \alpha \mathbb{E} \left\{ \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\}, \end{aligned}$$

em que a segunda igualdade decorre de que, como $h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i) = h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi'_{i,1+\alpha})$, podemos escrever

$$h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)^\alpha h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i) = b'_{i,1} h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i),$$

sendo

$$b'_{i,1} = \frac{B(\mu_i \phi_i, (1 - \mu_i) \phi_i)^{1-\alpha}}{B(\mu_i \phi'_{i,1+\alpha}, (1 - \mu_i) \phi'_{i,1+\alpha})}.$$

Para obter a submatriz $\mathbf{J}_\alpha^{\beta\beta}$, temos que

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^\top} \mathbf{U}_\beta^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= \mathbb{E} \left\{ \phi_i^2 \frac{(\mathcal{Y}_i^* - \mu_i^*)^2}{[g'_\mu(\mu_i)]^2} \boldsymbol{x}_i^\top \boldsymbol{x}_i \mid \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= - \frac{\phi_i^2 v_i}{[g'_\mu(\mu_i)]^2} \boldsymbol{x}_i^\top \boldsymbol{x}_i \end{aligned}$$

e

$$\begin{aligned} \mathbb{E} \left\{ \mathbf{U}_{\beta}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_{\beta}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= \mathbb{E} \left\{ \frac{\phi_i^2 (\mathcal{Y}_i^* - \mu_i^*)^2}{[g'_\mu(\mu_i)]^2} \boldsymbol{\mathcal{X}}_i^\top \boldsymbol{\mathcal{X}}_i \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= \frac{\phi_i^2 v_i}{[g'_\mu(\mu_i)]^2} \boldsymbol{\mathcal{X}}_i^\top \boldsymbol{\mathcal{X}}_i. \end{aligned}$$

Assim,

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[\mathbf{U}_{\beta}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha \right] \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= b'_{i,1} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^\top} \mathbf{U}_{\beta}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &\quad + b'_{i,1} \alpha \mathbb{E} \left\{ \mathbf{U}_{\beta}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_{\beta}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= -(1 - \alpha) \frac{b'_{i,1} \phi_i^2 v_i}{[g'_\mu(\mu_i)]^2} \boldsymbol{\mathcal{X}}_i^\top \boldsymbol{\mathcal{X}}_i. \end{aligned}$$

Da mesma forma, para obtermos a submatriz $\mathbf{J}_\alpha^{\gamma\beta}$, temos que

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^\top} \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= \mathbb{E} \left\{ -\frac{(1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} f_i \boldsymbol{\mathcal{Z}}_i^\top \boldsymbol{\mathcal{X}}_i \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= -\frac{(1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} c_i \boldsymbol{\mathcal{Z}}_i^\top \boldsymbol{\mathcal{X}}_i \end{aligned}$$

e

$$\begin{aligned} \mathbb{E} \left\{ \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_\beta^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= \mathbb{E} \left\{ \frac{\phi_i (1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} (\mathcal{Y}_i^* - \mu_i^*) [\mu_i (\mathcal{Y}_i^* - \mu_i^*) \right. \\ &\quad \left. + (\mathcal{Y}_i^\dagger - \mu_i^\dagger)] \boldsymbol{\mathcal{Z}}_i^\top \boldsymbol{\mathcal{X}}_i \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= \frac{(1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} c_i \boldsymbol{\mathcal{Z}}_i^\top \boldsymbol{\mathcal{X}}_i. \end{aligned}$$

De modo que

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left[\mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha \right] \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= b'_{i,1} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^\top} \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &\quad + b'_{i,1} \alpha \mathbb{E} \left\{ \mathbf{U}_\beta^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= -\frac{b'_{i,1} c_i}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} \boldsymbol{\mathcal{Z}}_i^\top \boldsymbol{\mathcal{X}}_i. \end{aligned}$$

Finalmente, para obtermos $\mathbf{J}_\alpha^{\gamma\gamma}$, temos que

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= -\frac{(1 - \alpha)^{-1}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} \mathbb{E} \left\{ (1 - \alpha)^{-1} d_i + \frac{g''_\phi(\phi'_{i,1+\alpha})}{g'_\phi(\phi'_{i,1-\alpha})} [\mu_i (\mathcal{Y}_i^* - \mu_i^*) \right. \\ &\quad \left. + (\mathcal{Y}_i^\dagger - \mu_i^\dagger)] \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \boldsymbol{\mathcal{Z}}_i^\top \boldsymbol{\mathcal{Z}}_i \\ &= -\frac{(1 - \alpha)^{-2}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} d_i \boldsymbol{\mathcal{Z}}_i^\top \boldsymbol{\mathcal{Z}}_i \end{aligned}$$

e

$$\begin{aligned} \mathbb{E} \left\{ \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= \mathbb{E} \left\{ \frac{(1-\alpha)^{-2}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} \left[\mu_i(\mathcal{Y}_i^* - \mu_i^*) \right. \right. \\ &\quad \left. \left. + (\mathcal{Y}_i^\dagger - \mu_i^\dagger) \right]^2 \boldsymbol{z}_i^\top \boldsymbol{z}_i \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= \frac{(1-\alpha)^{-2}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} d_i \boldsymbol{z}_i^\top \boldsymbol{z}_i. \end{aligned}$$

De modo que

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \left[\mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) h_\theta^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^\alpha \right] \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} &= b'_{i,1} \mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &\quad + b'_{i,1} \alpha \mathbb{E} \left\{ \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= -\frac{(1-\alpha)^{-1}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} b'_{i,1} d_i \boldsymbol{z}_i^\top \boldsymbol{z}_i. \end{aligned}$$

Para obtermos a matriz $\mathbf{K}_\alpha(\boldsymbol{\theta})$, temos que

$$\begin{aligned} \mathbf{K}_\alpha(\boldsymbol{\theta}) &= \mathbb{E} \left\{ \sum_{i \in \wp} \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top h_\theta^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha} \right\} \\ &= \sum_{i=1}^n (1 - \vartheta_i) \mathbb{E} \left\{ \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top h_\theta^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha} \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\}, \end{aligned}$$

em que

$$\begin{aligned} \mathbb{E} \left\{ \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top h_\theta^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha} \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= \int_{-\infty}^{\infty} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}^*(y_i^*; \boldsymbol{\theta})^\top h_\theta^*(y_i^*; \mu_i, \phi_i)^{2\alpha} h_\theta(y_i^*; \mu_i, \phi_i) dy_i^* \\ &= b'_{i,2} \int_{-\infty}^{\infty} \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}^*(y_i^*; \boldsymbol{\theta})^\top h_\theta^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \end{aligned}$$

e a última igualdade decorre do fato de que

$$h_\theta^*(y_i^*; \mu_i, \phi_i)^{2\alpha} h_\theta(y_i^*; \mu_i, \phi_i) = b'_{i,2} h_\theta^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}),$$

sendo

$$b'_{i,2} = \frac{B(\mu_i \phi'_{i,1+\alpha}, (1-\mu_i) \phi'_{i,1+\alpha})}{B(\mu_i \phi_i, (1-\mu_i) \phi_i)^{2\alpha} B(\mu_i \phi'_{i,1+\alpha}, (1-\mu_i) \phi'_{i,1+\alpha})}.$$

No que segue, calculamos as esperanças necessárias para as submatrizes $\mathbf{K}_\alpha^{\beta\beta}$, $\mathbf{K}_\alpha^{\beta\gamma}$ e $\mathbf{K}_\alpha^{\gamma\gamma}$. Temos que

$$\begin{aligned} \mathbb{E} \left\{ \mathbf{U}_\beta^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_\beta^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top h_\theta^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha} \middle| \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\ &= b'_{i,2} \int_{-\infty}^{\infty} \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta})^\top h_\theta^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \\ &= \frac{b'_{i,2} \phi_i^2}{[g'_\mu(\mu_i)]^2} \boldsymbol{x}_i^\top \boldsymbol{x}_i \int_{-\infty}^{\infty} (y_i^* - \mu_i^*)^2 h_\theta^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \\ &= \frac{b'_{i,2} \phi_i^2 v'_{i,1+\alpha}}{[g'_\mu(\mu_i)]^2} \boldsymbol{x}_i^\top \boldsymbol{x}_i, \end{aligned}$$

$$\begin{aligned}
 & \mathbb{E} \left\{ \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_\beta^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top h_{\boldsymbol{\theta}}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha} \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\
 &= b'_{i,2} \int_{-\infty}^{\infty} \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}_\beta^*(y_i^*; \boldsymbol{\theta})^\top h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \\
 &= \frac{b'_{i,2} \phi_i (1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} \mathbf{Z}_i^\top \boldsymbol{\mathcal{X}}_i \int_{-\infty}^{\infty} (y_i^* - \mu_i^*) \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger) \right]^2 h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \\
 &= \frac{b'_{i,2} \phi_i (1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} \mathbf{Z}_i^\top \boldsymbol{\mathcal{X}}_i \left\{ \mu_i \int_{-\infty}^{\infty} (y_i^* - \mu_i^*)^2 h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \right. \\
 &\quad \left. + \int_{-\infty}^{\infty} (y_i^* - \mu_i^*) (y_i^\dagger - \mu_i^\dagger) h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \right\} \\
 &= \frac{b'_{i,2} \phi_i (1 - \alpha)^{-1}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} \mathbf{Z}_i^\top \boldsymbol{\mathcal{X}}_i \left[\mu_i v'_{i,1+\alpha} - \psi'((1 - \mu_i) \phi'_{i,1+\alpha}) \right] \\
 &= \frac{b'_{i,2} (1 - \alpha)^{-1} c'_{i,1+\alpha}}{g'_\mu(\mu_i) g'_\phi(\phi'_{i,1-\alpha})} \mathbf{Z}_i^\top \boldsymbol{\mathcal{X}}_i,
 \end{aligned}$$

e

$$\begin{aligned}
 & \mathbb{E} \left\{ \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta}) \mathbf{U}_\gamma^*(\mathcal{Y}_i^*; \boldsymbol{\theta})^\top h_{\boldsymbol{\theta}}^*(\mathcal{Y}_i^*; \mu_i, \phi_i)^{2\alpha} \mathbb{I}_{\{c\}}(\mathcal{Y}_i) = 0 \right\} \\
 &= b'_{i,2} \int_{-\infty}^{\infty} \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta}) \mathbf{U}_\gamma^*(y_i^*; \boldsymbol{\theta})^\top h^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \\
 &= \frac{b'_{i,2} (1 - \alpha)^{-2}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} \mathbf{Z}_i^\top \mathbf{Z}_i \int_{-\infty}^{\infty} \left[\mu_i (y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger) \right]^2 h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi'_{i,1+\alpha}) dy_i^* \\
 &= \frac{b'_{i,2} (1 - \alpha)^{-2} d'_{i,1+\alpha}}{[g'_\phi(\phi'_{i,1-\alpha})]^2} \mathbf{Z}_i^\top \mathbf{Z}_i.
 \end{aligned}$$

Dessa forma, $\mathbf{J}_\alpha(\boldsymbol{\theta})$ e $\mathbf{K}_\alpha(\boldsymbol{\theta})$ podem ser escritas nas formas matriciais dadas em (6.21) e (6.22).

Referências Bibliográficas

- Akaike(1973)** H. Akaike. Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, 1:199–213. Citado na pág. [26](#)
- Albert e Anderson(1984)** A. Albert e J. Anderson. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71:1–10. Citado na pág. [22](#)
- Aljarrah et al.(2014)** M. A. Aljarrah, C. Lee e F. Famoye. On generating T-X family of distributions using quantile functions. *Journal of Statistical Distributions and Applications*, 1:1–17. Citado na pág. [8](#)
- Alzaatreh et al.(2013)** A. Alzaatreh, C. Lee e F. Famoye. A new method for generating families of continuous distributions. *Metron*, 71:63–79. Citado na pág. [7](#)
- Atkinson(1981)** A. C. Atkinson. Two graphical displays for outlying and influential observations in regression. *Biometrika*, 68:13–20. Citado na pág. [26](#), [62](#)
- Azzalini e Arellano-Valle(2013)** A. Azzalini e R. B. Arellano-Valle. Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 143:419–433. Citado na pág. [23](#)
- Azzalini e Capitanio(1999)** A. Azzalini e A. Capitanio. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society B*, 61:579–602. Citado na pág. [22](#)
- Barndorff-Nielsen(1977)** O. E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 353:401–419. Citado na pág. [15](#)
- Barndorff-Nielsen(1995)** O. E. Barndorff-Nielsen. Quasi profile and directed likelihoods from estimating functions. *Annals of the Institute of Statistical Mathematics*, 47:461–464. Citado na pág. [24](#)
- Barndorff-Nielsen e Jørgensen(1991)** O. E. Barndorff-Nielsen e B. Jørgensen. Some parametric models on the simplex. *Journal of Multivariate Analysis*, 39:106–116. Citado na pág. [7](#)
- Barros et al.(2008)** M. Barros, G. A. Paula e V. Leiva. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, 14:316–332. Citado na pág. [59](#)
- Basu et al.(1998)** A. Basu, I. R. Harris, N. L. Hjort e M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85:549–559. Citado na pág. [108](#), [112](#), [113](#)
- Bayer et al.(2021)** F. M. Bayer, F. Cribari-Neto e J. Santos. Inflated Kumaraswamy regressions with application to water supply and sanitation in Brazil. *Statistica Neerlandica*, 75:453–481. Citado na pág. [2](#)
- Bayes et al.(2012)** C. L. Bayes, J. L. Bazan e Garcia C. A new robust regression model for proportions. *Bayesian Analysis*, 7:771–796. Citado na pág. [1](#), [81](#)

- Berkane e Bentler(1986)** M. Berkane e P. M. Bentler. Moments of elliptically distributed random variates. *Statistics and Probability Letters*, 4:333–335. Citado na pág. [16](#)
- Bianco e Martínez(2009)** A. M. Bianco e E. Martínez. Robust testing in the logistic regression model. *Computational Statistics and Data Analysis*, 53:4095–4105. Citado na pág. [124](#)
- Bianco e Yohai(1996)** A. M. Bianco e V. J. Yohai. Robust estimation in the logistic regression model. Em *Robust Statistics, Data Analysis, and Computer Intensive Methods*, páginas 17–34. Springer, London. Citado na pág. [3](#), [108](#)
- Billor e Loynes(1993)** N. Billor e R. M. Loynes. Local influence: a new approach. *Communications in Statistics-Theory and Methods*, 22:1595–1611. Citado na pág. [68](#)
- Bondell(2005)** H. D. Bondell. Minimum distance estimation for the logistic regression model. *Biometrika*, 92:724–731. Citado na pág. [108](#)
- Box e Cox(1964)** G. E. P. Box e D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society B*, 26:211–243. Citado na pág. [114](#)
- Box e Tiao(2011)** G. E. P. Box e G. C. Tiao. *Bayesian Inference in Statistical Analysis*, volume 40. Reading, MA: Addison-Wesely Publishing Company. Citado na pág. [12](#)
- Bryson e Johnson(1981)** M. Bryson e M. Johnson. The incidence of monotone likelihood in the Cox model. *Technometrics*, 23:381–383. Citado na pág. [22](#)
- Cantoni e Ronchetti(2001)** E. Cantoni e E. Ronchetti. Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96:1022–1030. Citado na pág. [108](#)
- Carrasco et al.(2014)** J. M. F. Carrasco, S. L. P. Ferrari e R. B. Arellano-Valle. Errors-in-variables beta regression models. *Journal of Applied Statistics*, 41:1530–1547. Citado na pág. [1](#)
- Carroll e Pederson(1993)** R. J. Carroll e S. Pederson. On robust estimation in the logistic regression model. *Journal of the Royal Statistical Society B*, 55:693–706. Citado na pág. [108](#), [125](#)
- Cheng et al.(2019)** T. L. Cheng, A. Gerson, M. S. Moore, J. D. Reichard, J. DeSimone, C. K. Willis, W. F. Frick e A. M. Kilpatrick. Higher fat stores contribute to persistence of little brown bat populations with white-nose syndrome. *Journal of Animal Ecology*, 88:591–600. Citado na pág. [39](#), [81](#)
- Cook et al.(2008)** D. O. Cook, R. Kieshnick e B. D. McCullough. Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance*, 15:860–867. Citado na pág. [2](#)
- Cook(1986)** R. D. Cook. Assessment of local influence. *Journal of the Royal Statistical Society B*, 48:133–155. Citado na pág. [67](#), [98](#)
- Copas(1988)** J. B. Copas. Binary regression models for contaminated data. *Journal of the Royal Statistical Society B*, 50:225–253. Citado na pág. [2](#), [107](#)
- Cordeiro e de Castro(2011)** G. M. Cordeiro e M. de Castro. A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81:883–898. Citado na pág. [7](#)
- Cordeiro et al.(2000)** G. M. Cordeiro, S. L. P. Ferrari, M. A. Uribe-Opazo e K. L. P. Vasconcellos. Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters*, 46:317–328. Citado na pág. [59](#)
- Cox e Snell(1968)** D. R. Cox e E. J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society B*, 30:248–265. Citado na pág. [59](#), [60](#)

- Croux e Haesbroeck(2003)** C. Croux e G. Haesbroeck. Implementing the bianco and yohai estimator for logistic regression. *Computational Statistics and Data Analysis*, 44:273–295. Citado na pág. [108](#), [124](#), [127](#)
- Croux et al.(2002)** C. Croux, C. Flandre e G. Haesbroeck. The breakdown behavior of the maximum likelihood estimator in the logistic regression model. *Statistics and Probability Letters*, 60:377–386. Citado na pág. [3](#), [108](#), [132](#)
- Cysneiros e Vanegas(2008)** F. J. A. Cysneiros e L. H. Vanegas. Residuals and their statistical properties in symmetrical nonlinear models. *Statistics and Probability Letters*, 78:3269–3273. Citado na pág. [59](#)
- Cysneiros et al.(2010)** F. J. A. Cysneiros, G. M. Cordeiro e A. H. M. A. Cysneiros. Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*, 80:451–461. Citado na pág. [60](#)
- da Paz et al.(2019)** R. F. da Paz, N. Balakrishnan e J. L. Bazán. L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. *Brazilian Journal of Probability and Statistics*, 33:455–479. Citado na pág. [6](#), [11](#), [17](#), [53](#)
- Dickey(1967)** J. M. Dickey. Multivariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *The Annals of Mathematical Statistics*, 38:511–518. Citado na pág. [16](#)
- Dunn e Smyth(1996)** P. K. Dunn e G. K. Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5:236–244. Citado na pág. [26](#), [59](#), [61](#), [98](#)
- Espinheira et al.(2008)** P. L. Espinheira, S. L. P. Ferrari e F. Cribari-Neto. On beta regression residuals. *Journal of Applied Statistics*, 35:407–419. Citado na pág. [59](#), [67](#), [101](#)
- Espinheira et al.(2017)** P. L. Espinheira, E. G. Santos e F. Cribari-Neto. On nonlinear beta regression residuals. *Biometrical Journal*, 59:445–461. Citado na pág. [59](#)
- Eugene et al.(2002)** N. Eugene, C. Lee e F. Famoye. Beta-normal distribution and its applications. *Communications in Statistics -Theory and Methods*, 31:497–512. Citado na pág. [7](#)
- Fang et al.(1990)** K. T. Fang, S. Kotz e K. W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London. Citado na pág. [6](#)
- Ferrari e La Vecchia(2012)** D. Ferrari e D. La Vecchia. On robust estimation via pseudo-additive information. *Biometrika*, 99:238–244. Citado na pág. [114](#), [115](#)
- Ferrari e Yang(2010)** D. Ferrari e Y. Yang. Maximum L_q -likelihood estimation. *The Annals of Statistics*, páginas 753–783. Citado na pág. [114](#)
- Ferrari e Cribari-Neto(2004)** S. L. P. Ferrari e F. Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31:799–815. Citado na pág. [1](#), [6](#), [105](#), [108](#)
- Ferrari et al.(2011)** S. L. P. Ferrari, P. L. Espinheira e F. Cribari-Neto. Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica*, 65:337–351. Citado na pág. [1](#)
- Firth(1993)** D. Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, 80:27–38. Citado na pág. [22](#)
- Firth(1992)** D. Firth. Bias reduction, the Jeffreys prior and GLIM. Em *Advances in GLIM and Statistical Modelling*, páginas 91–100. Springer, New York. Citado na pág. [23](#)
- Fonseca et al.(2008)** T. C. Fonseca, M. A. Ferreira e H. S. Migon. Objective Bayesian analysis for the Student-t regression model. *Biometrika*, 95:325–333. Citado na pág. [27](#)

- Fonseca et al.(2012)** T. C. Fonseca, H. S. Migon e M. A. Ferreira. Bayesian analysis based on the Jeffreys prior for the hyperbolic distribution. *Brazilian Journal of Probability and Statistics*, 26:327–343. Citado na pág. 23, 27
- Galea et al.(2005)** M. Galea, G. A. Paula e F. J. A. Cysneiros. On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters*, 73:459–467. Citado na pág. 59, 60
- Ghosh(2019)** A. Ghosh. Robust inference under the beta regression model with application to health care studies. *Statistical Methods in Medical Research*, 28:871–888. Citado na pág. 3, 109, 112, 116, 124
- Ghosh e Basu(2016)** A. Ghosh e A. Basu. Robust estimation in generalized linear models: the density power divergence approach. *Test*, 25:269–290. Citado na pág. 3, 108, 112, 124
- Gómez-Déniz et al.(2014)** E. Gómez-Déniz, M. A. Sordo e E. Calderín-Ojeda. The log-Lindley distribution as an alternative to the beta regression model with application in insurance. *Insurance: Mathematics and Economics*, 54:49–57. Citado na pág. 1
- Hampel(1974)** F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69:383–393. Citado na pág. 106
- Hampel et al.(2011)** F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw e W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*,. John Wiley and Sons, New York. Citado na pág. 106
- Hinkley(1975)** D. V. Hinkley. On power transformations to symmetry. *Biometrika*, 62:101–111. Citado na pág. 17
- Huber(1964)** P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35:73–101. Citado na pág. 106, 107, 108
- Hubert e Vandervieren(2008)** M. Hubert e E. Vandervieren. An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, 52:5186–5201. Citado na pág. 39
- Jeffreys(1946)** H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 186:453–461. Citado na pág. 23
- Johnson(1949)** N. L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36:149–176. Citado na pág. 2, 8, 9, 17
- Johnson et al.(1995)** N. L. Johnson, S. Kotz e N. Balakrishnan. *Continuous Univariate Distributions*. John Wiley, New York. Citado na pág. 16
- Jones(2009)** M. C. Jones. Kumaraswamy’s distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6:70–81. Citado na pág. 7
- Jørgensen(2012)** B. Jørgensen. *Statistical Properties of the Generalized Inverse Gaussian Distribution*, volume 9. New York: Springer Science and Business Media, New York. Citado na pág. 15
- Kallianpur e Rao(1955)** G. Kallianpur e C. R. Rao. On fisher’s lower bound to asymptotic variance of a consistent estimate. *Sankhya*, 15:331–342. Citado na pág. 106
- Korkmaz(2020)** M. Korkmaz. A new heavy-tailed distribution defined on the bounded interval: the logit slash distribution and its application. *Journal of Applied Statistics*, 47:2097–2119. Citado na pág. 7, 11

- Kotz(1975)** S. Kotz. Multivariate distributions at a cross road. Em *A Modern Course on Statistical Distributions in Scientific Work*, páginas 247–270. Springer, Dordrecht. Citado na pág. 16
- Kumaraswamy(1980)** P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46:79–88. Citado na pág. 6
- Leiva et al.(2007)** V. Leiva, M. Barros, G. A. Paula e M. Galea. Influence diagnostics in log-birnbaum–saunders regression models with censored data. *Computational Statistics and Data Analysis*, 51:5694–5707. Citado na pág. 67
- Lemonte e Bazán(2016)** A. J. Lemonte e J. L. Bazán. New class of Johnson SB distributions and its associated regression model for rates and proportions. *Biometrical Journal*, 58:727–746. Citado na pág. 1, 2, 8, 9, 11, 53, 67, 151
- Lemonte e Moreno-Arenas(2019)** A. J. Lemonte e G. Moreno-Arenas. On residuals in generalized johnson sb regressions. *Applied Mathematical Modelling*, 67:62–73. Citado na pág. 59
- Lesaffre e Verbeke(1998)** E. Lesaffre e G. Verbeke. Local influence in linear mixed models. *Biometrics*, 54:570–582. Citado na pág. 67
- Lima(2018)** F. A. Lima. Distribuições de probabilidade no intervalo unitário. Dissertação de Mestrado, Universidade de São Paulo. Citado na pág. 6
- Lima e Cribari-Neto(2019)** V. M. Lima e F. Cribari-Neto. Penalized maximum likelihood estimation in the modified extended Weibull distribution. *Communications in Statistics–Simulation and Computation*, 48:334–349. Citado na pág. 22
- Liu et al.(2020)** P. Liu, K. C. Yuen, L. Wu, G. Tian e T. Li. Zero-one inflated simplex regression models for the analysis of continuous proportion data. *Statistics and its Interface*, 13:193–208. Citado na pág. 2
- Maluf et al.(2022)** Y. S. Maluf, S. L. P. Ferrari e F. F. Queiroz. Robust beta regression through the logit transformation. *Em preparação*. Citado na pág. 3, 109, 112, 115, 116, 117, 119, 125, 135
- McCullagh e Nelder(1989)** P. McCullagh e J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, London. Citado na pág. 53, 61, 96, 111
- Mitnik e Baek(2013)** P.A. Mitnik e S. Baek. The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54:177–192. Citado na pág. 6, 149
- Muscarella et al.(2020)** R. Muscarella, T. Emilio, O. L. Phillips, S. L. Lewis, F. Slik, W. J. Baker, T. L. P. Couvreur, W. L. Eiserhardt, J. C. Svenning, K. Affum-Baffoe et al. The global abundance of tree palms. *Global Ecology and Biogeography*, 29:1495–1514. Citado na pág. 100, 102
- Nelson(1991)** D. B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 57:347–370. Citado na pág. 12, 13
- Ospina e Ferrari(2010)** R. Ospina e S. L. P. Ferrari. Inflated beta distributions. *Statistical Papers*, 51:111–126. Citado na pág. 2, 110
- Ospina e Ferrari(2012)** R. Ospina e S. L. P. Ferrari. A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis*, 56:1609–1623. Citado na pág. 2, 101, 105, 146
- Pace e Salvan(1997)** Luigi Pace e Alessandra Salvan. *Principles of Statistical Inference: from a Neo-Fisherian Perspective*, volume 4. World Scientific, Singapore. Citado na pág. 95, 111

- Pereira et al.(2012)** G. H. A. Pereira, D. A. Botter e M. C. Sandoval. The truncated inflated beta distribution. *Communications in Statistics-Theory and Methods*, 41:907–919. Citado na pág. [2](#)
- Pianto e Cribari-Neto(2011)** D. M. Pianto e F. Cribari-Neto. Dealing with monotone likelihood in a model for speckled data. *Computational Statistics and Data Analysis*, 55:1394–1409. Citado na pág. [23](#)
- Pregibon(1982)** D. Pregibon. Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, 38:485–498. Citado na pág. [3](#), [108](#)
- Queiroz(2018)** F. F. Queiroz. O modelo de regressão GJS inflacionado em zero ou um. Dissertação de Mestrado, Universidade Federal do Rio Grande do Norte. Citado na pág. [67](#)
- Queiroz e Lemonte(2021)** F. F. Queiroz e A. J. Lemonte. A broad class of zero-or-one inflated regression models for rates and proportions. *The Canadian Journal of Statistics*, 49:566–590. Citado na pág. [2](#), [94](#), [146](#)
- R Core Team(2019)** R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL <https://www.R-project.org/>. Citado na pág. [27](#)
- Rao(1948)** C. R. Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. Em *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 44, páginas 50–57. Cambridge University Press. Citado na pág. [57](#)
- Ribeiro(2020)** T. K. A Ribeiro. *Regressão Beta Robusta*. Tese de Doutorado, Universidade de São Paulo. Citado na pág. [105](#)
- Ribeiro e Ferrari(2022)** T. K. A. Ribeiro e S. L. P. Ferrari. Robust estimation in beta regression via maximum L_q -likelihood. *Statistical Papers*, 46:1–33. Citado na pág. [3](#), [72](#), [74](#), [75](#), [109](#), [112](#), [114](#), [115](#), [117](#), [119](#), [124](#), [125](#), [126](#), [135](#), [146](#)
- Rieck e Nedelman(1991)** J. R. Rieck e J. R. Nedelman. A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33:51–60. Citado na pág. [14](#)
- Rogers e Tukey(1972)** W. H. Rogers e J. W. Tukey. Understanding some long-tailed symmetrical distributions. *Statistica Neerlandica*, 26:211–226. Citado na pág. [7](#), [14](#)
- Sartori(2003)** N. Sartori. Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, 90:533–549. Citado na pág. [22](#)
- Sartori(2006)** N. Sartori. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, 136:4259–4275. Citado na pág. [22](#)
- Schmid et al.(2009)** M. Schmid, F. Wickler, K. O. Maloney, Mitchell R., N. Fenske e A. Mayr. Boosted beta regression. *Plos One*, 8. doi: e61623. Citado na pág. [1](#)
- Schmit e Roth(1990)** J. T. Schmit e K. Roth. Cost effectiveness of risk management practices. *Journal of Risk and Insurance*, 57:455–470. Citado na pág. [72](#), [81](#)
- Schneider et al.(2020)** L. Schneider, R. P. Chalmers, R. Debelak e E. C. Merkle. Model selection of nested and non-nested item response models using Vuong tests. *Multivariate Behavioral Research*, 55:664–684. Citado na pág. [146](#)
- Smithson e Merkle(2013)** M. Smithson e E. C. Merkle. *Generalized Linear Models for Categorical and Continuous Limited Dependent Variables*. CRC Press, New York. Citado na pág. [6](#), [8](#)

- Smithson e Shou(2017)** M. Smithson e Y. Shou. Cdf-quantile distributions for modelling random variables on the unit interval. *Journal of Mathematical and Statistical Psychology*, 70:412–438. Citado na pág. 1, 8, 11
- Smithson e Verkuilen(2006)** M. Smithson e J. Verkuilen. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11:54–71. Citado na pág. 1
- Stasinopoulos et al.(2022)** M. Stasinopoulos, B. Rigby, V. Voudouris, C. Akantziliotou, M. Enea e D. Kiose. *gamlss: Generalised Additive Models for Location Scale and Shape*, 2022. URL <https://CRAN.R-project.org/package=gamlss>. R package version 5.4-3. Citado na pág. 38
- Tadikamalla e Johnson(1982)** P. R. Tadikamalla e N. L. Johnson. Systems of frequency curves generated by transformation of logistic variables. *Biometrika*, 69:461–465. Citado na pág. 6, 11
- Terrell(2002)** G. R. Terrell. The gradient statistic. *Computing Science and Statistics*, 34:206–215. Citado na pág. 57
- Thomas e Cook(1989)** W. Thomas e R. D. Cook. Assessing influence on regression coefficients in generalized linear models. *Biometrika*, 76:741–749. Citado na pág. 69, 70
- Townsend e Colonius(2005)** J. Townsend e H. Colonius. Variability of the max and min statistic: a theory of the quantile spread as a function of sample size. *Psychometrika*, 70:759–772. Citado na pág. 149
- Vanegas e Paula(2016)** L. H. Vanegas e G. A. Paula. Log-symmetric distributions: statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics*, 30:196–220. Citado na pág. 26
- Vuong(1989)** Q. H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 57:307–333. Citado na pág. 146
- Wald(1943)** A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54:426–482. Citado na pág. 57
- Wang et al.(2014)** X. F Wang, B. Hu, B. Wang e K. Fang. Bayesian generalized varying coefficient models for longitudinal proportional data with errors-in-covariates. *Journal of Applied Statistics*, 41:1342–1357. Citado na pág. 1
- Warwick e Jones(2005)** J. Warwick e M. C. Jones. Choosing a robustness tuning parameter. *Journal of Statistical Computation and Simulation*, 75:581–588. Citado na pág. 124
- Wei et al.(1998)** B. C. Wei, Y. Q. Hu e W. K. Fung. Generalized leverage and its applications. *Scandinavian Journal of Statistics*, 25:25–37. Citado na pág. 66
- Wilks(1938)** S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9:60–62. Citado na pág. 57
- Zeileis e Croissant(2020)** A. Zeileis e Y. Croissant. *Formula: Extended Model Formulas*, 2020. URL <https://CRAN.R-project.org/package=Formula>. R package version 1.2-4. Citado na pág. 80
- Zeileis et al.(2020)** A. Zeileis, F. Cribari-Neto, B. Gruen, I. Kosmidis, A. B. Simas e A. V. Rocha. *betareg: Beta Regression*, 2020. URL <https://CRAN.R-project.org/package=betareg>. R package version 3.1-4. Citado na pág. 38, 77
- Zhang et al.(2016)** P. Zhang, Z. Qiu e C. Shi. *simplexreg: An r package for regression analysis of proportional data using the simplex distribution*. *Journal of Statistical Software*, 71:1–21. Citado na pág. 77