



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

Diego Renato Risco Cosvalente

**Distribuição zero ajustada log-simétrica: estimação e modelagem**

Recife

2021

Diego Renato Risco Cosvalente

**Distribuição zero ajustada log-simétrica: estimação e modelagem**

Dissertação apresentada ao Curso de Mestrado em Estatística do Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de mestre em Estatística.

**Área de Concentração:** Estatística Aplicada

**Orientador:** Francisco José de Azevêdo Cysneiros

Recife

2021

Catálogo na fonte  
Bibliotecário Cristiano Cosme S. dos Anjos, CRB4-2290

C834d Cosavalente, Diego Renato Risco  
Distribuição zero ajustada log-simétrica: estimação e modelagem / Diego Renato Risco Cosavalente. – 2021.  
102 f.: il., fig., tab.

Orientador: Francisco José de Azevêdo Cysneiros.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2021.  
Inclui referências e apêndices.

1. Estatística aplicada. 2. Classe log-simétrica. 3. Distribuições zero ajustadas. 4. Estimador de máxima verossimilhança. I. Cysneiros, Francisco José de Azevêdo (orientador). II. Título.

310

CDD (23. ed.)

UFPE- CCEN 2021 - 69

**DIEGO RENATO RISCO COSAVALENTE**

**DISTRIBUIÇÃO ZERO AJUSTADA LOG-SIMÉTRICA: ESTIMAÇÃO E  
MODELAGEM**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 25 de fevereiro de 2021.

**BANCA EXAMINADORA**

---

Prof. Dr. Francisco José de Azevêdo Cysneiros  
UFPE

---

Prof. Dr. Roberto Ferreira Manghi  
UFPE

---

Prof. Dr. Manoel Ferreira dos Santos Neto  
UFCG

Aos meus pais Carlos e Rocio, com admiração,  
Aos meus avós, com saudades.

## AGRADECIMENTOS

Aos meus pais: Carlos Risco e Rocio Cosavalente, os quais eu amo profundamente, pela educação que me deram, e que me indicaram os caminhos corretos da vida dando apoio a esta importante decisão.

À minha irmã Mariana e à toda minha família, a qual eu tenho admiração e respeito, obrigado por todo apoio e pensamento positivo.

Aos meus avós Alberto Risco e Rosa Dávila (in memoriam) com muito carinho.

Ao meu orientador Prof. Francisco José de Azevedo Cysneiros, pela oportunidade de demonstrar a minha capacidade através da bolsa de estudo, sempre dedicado a me incentivar e melhorar profissionalmente assim como pessoalmente neste caminho da pesquisa. Eu o admiro muito como estatístico.

À minha amiga Lara Carvalho, a qual eu respeito e admiro como pessoa. Obrigado pelo carinho, apoio e bons momentos juntos.

À minhas amigas: Cristine Oliveira, Ranah Duarte e Larissa Lima pelo apoio nos primeiros dias de chegada ao Brasil e em todos os momentos juntos. Vocês são pessoas maravilhosas.

Aos meus amigos de mestrado: Thalytta Calvacante, Lucas Araújo, Luis Felix, pelo companheirismo e bons momentos compartilhados.

Agradeço ao corpo docente do Departamento de Estatística do Programa de Pós-graduação da UFPE e, de modo geral, a todos os funcionários que fazem parte deste departamento.

À FACEPE, pelo apoio financeiro.

## RESUMO

Comumente em situações reais tais como nas áreas de seguros, ecologia e biometria, nos deparamos com a variável de interesse sendo positiva e além disso, contendo zeros e observações atípicas. Transformações da variável ou ignorar os valores zeros não é um melhor procedimento a ser utilizado ou mesmo adequado levando a uma maior dificuldade de interpretação e conclusões errôneas. Nesse sentido, distribuições zero ajustada positivas permitem ajustar a variável de interesse positiva com presença de zeros considerando duas componentes. A primeira componente, denominada contínua, é responsável de modelar a variável de interesse quando esta é positiva, e a segunda é discreta e é responsável por modelar o evento em que a variável de interesse seja igual a zero. O objetivo do presente trabalho foi propor uma nova classe de distribuições e de modelos de regressão para dados contínuos positivos com excesso de zeros em que a componente contínua segue a classe log-simétrica que contém distribuições bimodais e distribuições com caudas mais leves ou pesadas do que a log-normal. Foi considerado o método de máxima verossimilhança para a estimação dos parâmetros do modelo proposto. Além disso, estudos de simulação de Monte Carlo sob para diferentes cenários para avaliar o comportamento dos estimadores de máxima verossimilhança dos parâmetros foram realizados. Bem como, construímos intervalos de confiança assintóticos para os parâmetros e avaliamos as suas propriedades. Análise de Resíduos e métodos de diagnóstico também foram desenvolvidos. Para ilustrar nossa metodologia, conjuntos de dados reais na área de finanças e educação foram analisados.

**Palavras-chaves:** Classe Log-Simétrica. Distribuições Zero Ajustadas. Estimador De Máxima Verossimilhança. Influência Global. Influência Local. Modelos De Regressão Zero Ajustado.

## ABSTRACT

Commonly in real situations such as in the areas of insurance, ecology and biometrics we find that the response variable is positive and in addition, containing zeros and atypical observations. Transformations of the variable or ignoring the zero values is not a best procedure to be used or even appropriate leading to greater difficulty in interpretation and erroneous conclusions. In this sense, zero adjusted distributions allow to fit the positive variable of interest with excess of zero considering two components. The first component, called continuous, is responsible for modeling the variable of interest when it is positive, and the second is discrete and it is responsible for modeling the event in which the variable of interest is equal to zero. The main objective of the present work was to propose a new class of distributions and regression models for positive continuous data with excess of zeros in which the continuous component follows the log-symmetric class that contains bimodal distributions and distributions with lighter or heavier tails than the log-normal. The maximum likelihood method was used to estimate the model parameters proposed. In addition, Monte Carlo simulation studies for different scenarios to evaluate the maximum likelihood estimators of the parameters were performed. As well, we build asymptotic confidence intervals for the parameters and evaluate their properties. Residual analysis and diagnostic methods were developed. To illustrate our methodology, real data sets in the area of finance and education were analyzed.

**Keywords:** Log-Symmetric Class. Zero Adjusted Distributions. Maximum Likelihood Estimator. Global Influence. Local Influence. Zero Adjusted Regression Models.

## LISTA DE FIGURAS

Figura 1 – Funções de densidade para algumas distribuições padrão da classe simétrica, t-Student(a) e exponencial-potência(b). . . . .	19
Figura 2 – Funções de densidade para algumas distribuições padrão da classe simétrica, logística-tipo-I(a) e logística-tipo-II(b). . . . .	20
Figura 3 – Funções de densidade para algumas distribuições da classe log-simétrica, log-normal ( $\eta = 2, \phi$ )(a), log-t-Student ( $\eta = 2, \phi, \nu = 4$ )(b), log-exponencial-potência ( $\eta = 2, \phi, \nu = -0.5$ )(c). . . . .	22
Figura 4 – Funções de densidade para algumas distribuições da classe zero ajustada log-simétrica, zero ajustado log-normal ( $\eta = 2, \pi = 0, 4$ )(a), zero ajustado log-t-Student ( $\eta = 2, \pi = 0, 4, \nu = 4$ )(b), zero ajustado log-exponencial-potência ( $\eta = 2, \pi = 0, 4, \nu = -0, 5$ )(c). . . . .	25
Figura 5 – Histograma para os EMV dos parâmetros $\log(\eta)$ , $\log(\phi)$ , $\text{logit}(\pi)$ com o respectivo intervalo de confiança médio para valores de $n$ e níveis $100X[1 - \alpha]\%$ de confiança para a distribuição ZALN com $\theta = (2, 4, 0, 4)^\top$ . . . . .	37
Figura 6 – Histograma para os EMV dos parâmetros $\log(\eta)$ , $\log(\phi)$ , $\text{logit}(\pi)$ com o respectivo intervalo de confiança médio para valores de $n$ e níveis $100X[1 - \alpha]\%$ de confiança para a distribuição ZALSt com $\theta = (2, 4, 0, 4)^\top$ e $\nu = 4$ . . . . .	38
Figura 7 – Histograma para os EMV dos parâmetros $\log(\eta)$ , $\log(\phi)$ , $\text{logit}(\pi)$ com o respectivo intervalo de confiança médio para valores de $n$ e níveis $100X[1 - \alpha]\%$ de confiança para a distribuição ZALPE com $\theta = (2, 4, 0, 4)^\top$ e $\nu = -0, 5$ . . . . .	39
Figura 8 – Histograma e boxplot ajustado para as despesas especiais dos vereadores eleitos no período 2012/2013 na cidade de Leicester. . . . .	41
Figura 9 – Histograma e gráfico Q-Q para as distribuições ajustadas para o conjunto de dados de despesas especiais dos vereadores eleitos no período 2012/2013 na cidade de Leicester. . . . .	43
Figura 10 – Gráfico de frequência (a) e boxplot ajustado (b) aos gastos totais em educação em adolescentes do Peru no ano 2009. . . . .	67
Figura 11 – Boxplots para gastos em educação de acordo ao sexo (a), capital (b), menores que estudam (c) e anos de escolaridade (d). . . . .	68

Figura 12 – Diagramas de dispersão em relação aos gastos em educação de acordo ao índice de moradia (a) e índice de consumo (b). . . . .	69
Figura 13 – Resíduo do tipo quantílico do modelo ZAGA ajustado aos dados de gastos em educação. . . . .	74
Figura 14 – Resíduo do tipo quantílico do modelo ZAIG ajustado aos dados de gastos em educação. . . . .	75
Figura 15 – Resíduo do tipo quantílico do modelo ZALN ajustado aos dados de gastos em educação. . . . .	76
Figura 16 – Resíduo do tipo quantílico do modelo ZALSt ajustado aos dados de gastos em educação. . . . .	77
Figura 17 – Resíduo do tipo quantílico do modelo ZALPE ajustado aos dados de gastos em educação. . . . .	78
Figura 18 – Gráfico de valores ajustados contra resíduo componente do desvio (a), (c), (e) e gráfico normal de probabilidade com envelope (b), (d), (f) para a componente contínua e discreto ajustados aos dados de gastos em educação. . . . .	80
Figura 19 – Gráfico de valores ajustados contra resíduos do tipo ZAR (a) e gráfico normal de probabilidade com envelope (b) para a componente contínua ajustado aos dados de gastos em educação. . . . .	81
Figura 20 – Gráfico de influência global para a componente discreta (a), componente contínua (b) e geral (c) para o modelo ajustado aos dados dos gastos totais em educação em adolescentes do Peru no ano 2009. . . . .	82
Figura 21 – Gráfico de influência local ponderação de casos para a componente discreta (a), componente contínua (b) e geral (c) para o modelo ajustado aos dados dos gastos totais em educação em adolescentes do Peru no ano 2009. . . . .	82
Figura 22 – Gráfico de pontos de alavanca generalizados para $\pi$ (a), $\eta$ (b) e $\phi$ (c) para o modelo ajustado aos dados dos gastos totais em educação em adolescentes do Peru no ano 2009. . . . .	83

## LISTA DE TABELAS

Tabela 1 – Valores para $d_g$ e $f_g$ para algumas distribuições da classe simétrica $Y$ . . .	27
Tabela 2 – Resultados do estudo de simulação para os EMV dos parâmetros $\eta = 2$ , $\phi$ , $\pi$ de acordo a $n = 10, 30, 50$ para a distribuição zero ajustada log-normal. .	31
Tabela 3 – Resultados do estudo de simulação para os EMV dos parâmetros $\eta = 2$ , $\phi$ , $\pi$ de acordo a $n = 10, 30, 50$ para a distribuição zero ajustada log-t-Student com $\nu = 4$ . . . . .	32
Tabela 4 – Resultados do estudo de simulação para os EMV dos parâmetros $\eta = 2$ , $\phi$ , $\pi$ de acordo a $n = 10, 30, 50$ para a distribuição zero ajustada log-exponencial-potência com $\nu = -0, 5$ . . . . .	33
Tabela 5 – Cobertura de probabilidade para os EMV dos parâmetros $\log(\eta)$ , $\log(\phi)$ e $\log(\pi/(1-\pi))$ de acordo a $n = 10, 30, 50$ e $100 \times [1-\alpha]\%$ para a distribuição zero ajustada log-normal. . . . .	34
Tabela 6 – Cobertura de probabilidade para os EMV dos parâmetros $\log(\eta)$ , $\log(\phi)$ e $\log(\pi/(1-\pi))$ de acordo a $n = 10, 30, 50$ e $100 \times [1-\alpha]\%$ para a distribuição zero ajustada log-t-Student com $\nu = 4$ . . . . .	35
Tabela 7 – Cobertura de probabilidade para os EMV dos parâmetros $\log(\eta)$ , $\log(\phi)$ e $\log(\pi/(1-\pi))$ de acordo a $n = 10, 30, 50$ e $100 \times [1-\alpha]\%$ para a distribuição zero ajustada log-exponencial-potência com $\nu = -0, 5$ . . . . .	36
Tabela 8 – Estimativa de máxima verossimilhança para os parâmetros e estatísticas para as distribuições consideradas ajustadas ao conjunto de dados de despesas especiais dos vereadores eleitos no período 2012/2013 na cidade de Leicester. . . . .	42
Tabela 9 – Medidas de resumo das variáveis contínuas.(IIQ = Intervalo interquartil) . . . . .	67
Tabela 10 – Estimativas de máxima verossimilhança para a componente discreta ajustado aos dados de gastos em educação. . . . .	70
Tabela 11 – Estimativas de máxima verossimilhança para a componente discreta com as variáveis selecionadas ajustadas aos dados de gastos em educação. . . . .	70
Tabela 12 – Estimativas de máxima verossimilhança (erro padrão) para o submodelo da componente contínua $\eta$ e $\mu$ dos modelos ajustados aos dados de gastos em educação. . . . .	71

Tabela 13 – Estimativas de máxima verossimilhança (erro padrão) para o submodelo final da componente contínua $\eta$ e $\mu$ dos modelos ajustados aos dados de gastos em educação. . . . .	72
Tabela 14 – Valor da estatística de razão de verossimilhança e $p$ -valor para o respectivo teste de hipótese nos modelos de regressão ajustados ao dados de gastos em educação. . . . .	72
Tabela 15 – Estimativas de máxima verossimilhança (erro padrão) para o submodelo da componente contínua $\phi$ e $\sigma$ dos modelos ajustados aos dados de gastos em educação. . . . .	73
Tabela 16 – Valores do $-2\log(\hat{L})$ , AIC e BIC para os modelos considerados ajustados aos dados de gastos em educação. . . . .	79
Tabela 17 – Estimativas dos parâmetros do modelo ZALPE final ajustado aos dados de gastos em educação e impacto porcentual (entre parênteses) após retirar as observações influentes consideradas. . . . .	85

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	MOTIVAÇÃO	14
1.2	OBJETIVOS	16
1.3	APRESENTAÇÃO DOS CAPÍTULOS	16
1.4	SUPORTE COMPUTACIONAL	16
1.5	PRELIMINARES	17
<b>1.5.1</b>	<b>Classe simétrica de distribuições</b>	<b>17</b>
<b>1.5.2</b>	<b>Classe log-simétrica de distribuições</b>	<b>21</b>
<b>2</b>	<b>DISTRIBUIÇÕES ZERO AJUSTADO LOG-SIMÉTRICAS</b>	<b>23</b>
2.1	DEFINIÇÃO	23
<b>2.1.1</b>	<b>Estimação de parâmetros</b>	<b>26</b>
2.2	ESTUDO DE SIMULAÇÃO	28
2.3	APLICAÇÃO	40
<b>3</b>	<b>MODELOS DE REGRESSÃO LINEARES ZERO AJUSTADO LOG-SIMÉTRICAS</b>	<b>44</b>
3.1	MODELO DE REGRESSÃO	45
<b>3.1.1</b>	<b>Função escore</b>	<b>47</b>
<b>3.1.2</b>	<b>Informação de Fisher</b>	<b>48</b>
<b>3.1.3</b>	<b>Processo de estimação</b>	<b>49</b>
<b>3.1.4</b>	<b>Distribuição assintótica</b>	<b>51</b>
<b>3.1.5</b>	<b>Teste de razão de verosimilhança</b>	<b>52</b>
<b>3.1.6</b>	<b>Seleção de variáveis explanatórias</b>	<b>52</b>
3.2	RESÍDUOS E MÉTODOS DE DIAGNÓSTICO	53
<b>3.2.1</b>	<b>Resíduos</b>	<b>53</b>
3.2.1.1	<i>Resíduo componente do desvio</i>	54
3.2.1.2	<i>Resíduo quantílico</i>	55
3.2.1.3	<i>Resíduo ZAR</i>	56
<b>3.2.2</b>	<b>Influência Global</b>	<b>56</b>
<b>3.2.3</b>	<b>Influência Local</b>	<b>58</b>
3.2.3.1	<i>Perturbação de casos ponderados</i>	60

3.2.3.2	<i>Perturbação nas variáveis explanatórias</i> . . . . .	60
3.2.4	<b>Alavancagem Generalizada</b> . . . . .	64
3.3	APLICAÇÃO . . . . .	65
4	<b>CONCLUSÕES</b> . . . . .	86
	<b>REFERÊNCIAS</b> . . . . .	87
	<b>APÊNDICE A – PRIMEIRAS E SEGUNDAS DERIVADAS DOS MODELOS LINEARES ZERO AJUSTADO LOG- SIMÉTRICAS</b> . . . . .	91
	<b>APÊNDICE B – INFLUÊNCIA LOCAL</b> . . . . .	94
	<b>APÊNDICE C – ALAVANCAGEM GENERALIZADA</b> . . . . .	100

# 1 INTRODUÇÃO

## 1.1 MOTIVAÇÃO

Em situações reais, é comum encontrar dados positivos com a presença de zeros verdadeiros, dissemos zeros verdadeiros quando o valor zero é realmente observado, que comumente ocorrem em conjuntos de dados biométricos, ecológicos ou em de dados sobre despesas de seguros. Por exemplo, em ecologia é comum encontrar uma grande proporção de valores zeros causados por efeitos ecológicos reais como contagens de abundância, taxas de ocupação proporcionais ou densidades populacionais contínuas dos quais não se ajustam a distribuições usualmente utilizadas como a distribuição normal, Poisson ou beta (Martin et al. 2005). Outro exemplo é o “Medical Expenditure Panel Survey” (MEPS), que contém despesas com saúde de adultos nos EUA, nos quais muitos indivíduos não registraram gastos médicos (resposta zero) ao longo dos anos. Para lidar com tais situações, métodos usuais como transformar a variável de resposta como, por exemplo, a transformação logarítmica não podem ser usados na presença de zeros ou ignorar os valores zeros pode ser uma estratégia não adequada, tornando impossível prever a probabilidade de zero e levando a uma inferência incorreta sobre os parâmetros. Uma estratégia é de considerar uma distribuição mista de duas componentes: uma distribuição contínua cujo suporte é o intervalo  $(0, \infty)$  e uma distribuição discreta que assume zero com uma certa probabilidade.

Aitchison & Brown (1957) introduziu uma distribuição de mistura entre uma distribuição degenerada no zero e a distribuição log-normal, denominada distribuição delta. Duan et al. (1983) propôs o modelo de duas partes. A primeira parte é um modelo probit para eventos binários de despesas zero e a segunda etapa é um modelo log-linear para despesas positivas. Heller et al. (2006) apresentou a distribuição zero ajustada gaussiana inversa para modelar o valor pago de uma seguradora. Rodrigues-Motta et al. (2015) propõem um “framework” para distribuições positivas com a presença de zeros considerando a família exponencial de dois parâmetros como a componente contínua, nos quais inclui distribuições tais como log-normal Weibull, gama e gaussiana inversa. Leiva et al. (2016) propuseram a distribuição zero ajustada Birnbaum-Saunders reparametrizada para modelar dados de fadiga com a presença de zeros verdadeiros. Recentemente, Hashimoto et al. (2019) apresentaram o modelo de regressão gama-Weibull com zeros.

A parte contínua dessas distribuições mencionadas são algumas das distribuições mais

flexíveis encontradas na literatura, mas de acordo com Lawless (2003), a distribuição log-normal tem sido usada em diversas aplicações na engenharia, medicina e outras áreas, tornando a distribuição log-normal uma das mais utilizadas para modelar variáveis de interesse que podem incluir observações atípicas. Jones (2008) investigou a classe log-simétrica, que é uma generalização da distribuição log-normal. Esta classe inclui distribuições bimodais e distribuições de caudas mais leves/pesadas do que a log-normal, por exemplo as distribuições log-Student-t, tipo-I-log-Logístico, tipo-II-log-Logístico, log-contaminado-normal e log-exponencial-potência. Além disso, de acordo com Puig (2008) as distribuições log-simétricas têm propriedades muito desejáveis, fechada sob mudança de escala e propriedade recíproca, que são utilizadas para descrever dados com razões de magnitudes positivas. A classe envolve como parâmetros a mediana e assimetria, implicitamente, tornando-a uma classe muito flexível; também de acordo com Vanegas e Paula (2016), a função de risco das distribuições log-simétricas é mais flexível em comparação com outras distribuições como as distribuições Gama ou Gaussiana Inversa e pode assumir várias formas, por exemplo, crescente, decrescente, em forma de banheira e banheira invertida. Existem vários estudos sobre a classe log-simétrica, Vanegas e Paula (2016) estudaram e discutiram algumas propriedades da classe log-simétrica, Medeiros e Ferrari (2017) desenvolvem testes de hipóteses para modelos de regressão linear simétricos e log-simétricos, enquanto Ventura et al. (2019) conduzem um estudo de simulação de Monte Carlo para investigar a precisão dos critérios de informação usuais nos modelos de regressão log-simétricos.

Motivados pelos trabalhos de distribuições semicontínuas e a vantagem da classe log-simétrica em acomodar outliers, propomos uma distribuição semicontínua usando a classe log-simétrica, com seu suporte no intervalo  $[0, \infty)$ . Na literatura, existem diferentes nomes para este tipo de distribuição, o “zero inflacionado” como em Lambert (1992), o “zero ajustado” como em Heller et al. (2006) ou o “zero aumentado” como em Rodrigues-Motta et al. (2015), mas de acordo com Lambert (1992), os modelos zero inflacionados adicionam uma massa de probabilidade adicional ao resultado de zero, por isso, é mais comum para usá-lo quando envolve distribuições discretas. No presente trabalho, zero ajustado é usado, assim a nova classe é chamada por zero ajustado log-simétrica. A classe de distribuições zero ajustada log-simétrica têm várias propriedades estatísticas desejáveis. Por exemplo, a função Quantil é fácil de obter, os parâmetros são ortogonais e em alguns casos pode envolver um parâmetro extra ou um vetor de parâmetros extras que permitem por exemplo, generalizar distribuições, controlar a assimetria e curtose da distribuição.

## 1.2 OBJETIVOS

O objetivo geral deste trabalho é propor uma nova classe de distribuições e de modelos de regressão baseado em uma mistura entre a classe log-simétrica e uma distribuição discreta. Podemos descrever como objetivos específicos, a saber

- i - Propor a classe de distribuições zero ajustada log-simétrica assim como apresentar a estimação pontual e intervalar.
- ii - Desenvolver os modelos lineares zero ajustados log-simétricos, bem como apresentar um processo iterativo de estimação para os parâmetros.
- iii - Obter resíduos para a verificação das suposições feitas.
- iv - Obter medidas de influência tal como influência global, local e alavancagem generalizada para a utilização como diagnósticos dos modelos propostos.

## 1.3 APRESENTAÇÃO DOS CAPÍTULOS

O presente trabalho encontra-se dividido em cinco capítulos. No segundo capítulo é apresentado a classe zero ajustada log-simétrica, bem como a estimação pontual por máxima verossimilhança e intervalar dos parâmetros, seguido de um estudo de simulação para avaliar as propriedades dos estimadores de máxima verossimilhança dos parâmetros. Uma aplicação é apresentada em que comparamos as distribuições propostas com outras já existentes na literatura. No terceiro capítulo é abordado o modelo de regressão linear zero ajustado log-simétrico em que é desenvolvido um processo iterativo para a estimação por máxima verossimilhança dos parâmetros. Técnicas de análise de resíduos e métodos de diagnósticos sob o enfoque de influência global, influência local e alavancagem generalizada seguido de uma ilustração são apresentadas. Por fim, no quarto capítulo apresentamos as conclusões finais sobre a proposta apresentada.

## 1.4 SUPORTE COMPUTACIONAL

As avaliações numéricas realizadas ao longo do presente trabalho foram feitas num computador inter core i5 de 12 gb de ram sob o sistema operacional Windows 10 através da ambiente

de programação R na versão 4.0.2, encontrando-se disponível de forma gratuita no endereço [www.r-project.org/](http://www.r-project.org/). Todas as rotinas computacionais desenvolvidas pode ser solicitada ao autor. Por fim, a dissertação foi digitada usando o sistema de tipografia  $\text{\LaTeX}$  para o desenvolvimento e edição de texto científico.

## 1.5 PRELIMINARES

Nesta seção pretende-se apresentar e caracterizar a classe simétrica e log-simétrica de distribuições.

### 1.5.1 Classe simétrica de distribuições

Seja a variável aleatória  $Y$  com suporte em  $\mathbb{R}$  e que segue a classe simétrica com função de densidade de probabilidade dada por

$$f_Y(y) = \frac{g[(y - \mu)^2 \phi]}{\sqrt{\phi}}, \quad y \in \mathbb{R}, \quad (1.1)$$

em que  $g(\cdot)$  é uma função geradora de densidade que satisfaz  $g(u) > 0$  e  $\int_0^\infty u^{-\frac{1}{2}} g(u) \partial u = 1$  para  $u > 0$ ,  $\mu \in \mathbb{R}$  é o parâmetro de locação e  $\phi > 0$  de escala. Denotamos a variável  $Y \sim S(\mu, \phi, g(\cdot))$ . A classe simétrica contém distribuições com caudas mais leves/pesadas que da distribuição normal podendo modelar dados que contém mais observações atípicas. Algumas distribuições tal como normal, t-Student, t-Student generalizada, logística tipo I, logística tipo I, exponencial-potencia, normal-contaminada, entre outras pertencem a essa classe.

Algumas propriedades da distribuição normal podem ser estendidas para a classe simétrica de distribuições. Se  $Y \sim S(\mu, \phi, g(\cdot))$  então  $a + bY \sim S(a + b\mu, b^2\phi, g(\cdot))$ , em que  $a, b \in \mathbb{R}$  com  $b \neq 0$ . Da mesma forma  $\bar{Y} = (Y - \mu)/\sqrt{\phi} \sim S(0, 1, g(\cdot))$ , chamaremos de  $\bar{Y}$  de uma variável aleatória com distribuição simétrica padrão. A função característica de  $Y$ ,  $\varphi_Y(r) = E(e^{irY})$  é dada por  $\varphi_Y(r) = e^{ir\mu} \delta(r^2\phi)$ ,  $r \in \mathbb{R}$  para alguma função  $\delta$  com  $\delta(u) \in \mathbb{R}$  para  $u > 0$ . Se existir,  $E(Y) = \mu$  e  $\text{Var}(Y) = \xi\phi$ , em que  $\xi > 0$  é uma constante dada por  $\xi = -2\varphi'_Y(0)$ , com  $\varphi'_Y(0) = \partial\varphi_Y(u)/\partial u|_{u=0}$  e que não depende dos parâmetros  $\mu$  e  $\phi$  (Fang et al. 1990).

Como dito, membros da classe simétrica são induzidos por  $g(\cdot)$  assim, uma função de peso pode-se obter  $v(y) = -2g'(y^2)/g(y^2)$  sendo uma função importante para estimar os parâmetros mediante o método de máxima verossimilhança. Em algumas situações  $g(\cdot)$  envolve

um parâmetro extra ou um vetor de parâmetros indicado aqui por  $\nu$ . Apresentamos algumas distribuições pertencentes à classe simétrica, a saber

### ***Distribuição normal***

$$g(u) = \frac{1}{\sqrt{2\pi}} \exp[-u/2], \quad v(y) = 1$$

### ***Distribuição t-Student***

$$g(u) = \frac{\nu^{v/2}}{B(1/2, \nu/2)} (\nu + u)^{-\frac{\nu+1}{2}}, \quad \nu > 0 \quad \text{e} \quad v(y) = \frac{\nu + 1}{\nu + y^2}$$

em que  $B(.,.)$  é a função Beta.

### ***Distribuição exponencial-potência***

$$g(u) = C(\nu) \exp \left[ -\frac{1}{2} u^{1/(1+\nu)} \right], \quad -1 < \nu \leq 1 \quad \text{e} \quad v(y) = \frac{|y|^{-(2\nu)/(\nu+1)}}{1 + \nu}$$

em que  $C(\nu)^{-1} = \Gamma(1 + \frac{1+\nu}{2}) 2^{1+(1+\nu)/2}$ . Como casos especiais temos as distribuições normal ( $\nu = 0$ ) e Laplace ( $\nu = 1$ ).

### ***Distribuição Logística tipo-I***

$$g(u) = c \frac{e^{-u}}{(1 + e^{-u})^2}, \quad v(y) = 2 \tanh(y^2/2)$$

em que  $c \approx 1.484300029$  é uma constante normalizadora que satisfaz  $\int_0^\infty u^{-\frac{1}{2}} g(u) \partial u = 1$ .

### ***Distribuição Logística tipo-II***

$$g(u) = \frac{e^{u^{1/2}}}{(1 + e^{-u^{1/2}})^2}, \quad v(y) = (e^{|y|^{-1}})/(|y|(1 + e^y))$$

### Distribuição normal-contaminada

$$g(u) = \sqrt{\nu_2} \exp\left[-\frac{1}{2}\nu_2 u\right] + \frac{(1-\nu_1)}{\nu_1} \exp\left[-\frac{1}{2}u\right], \quad v(y) = \frac{\nu_2^{3/2}\nu_1 \exp[(1-\nu_2)(y^2/2)] + (1-\nu_1)}{\nu_2^{1/2}\nu_1 \exp[(1-\nu_2)(y^2/2)] + (1-\nu_1)}$$

em que  $0 < \nu_1 < 1, 0 < \nu_2 < 1$ .

Ilustramos nas Figuras 1 e 2, funções de densidades para algumas distribuições padrão da classe simétrica, em todos os casos, comparamos com a função de densidade de uma normal padrão. Podemos observar que as diferentes funções de densidade apresentam características particulares, a distribuição t-Student tende a distribuição normal à medida que os graus de liberdade aumenta. Distribuições platicúrticas como é o caso da distribuição exponencial-potência com  $\nu = -0.5$  e Logística tipo-I, e distribuições leptocúrticas como é o caso da distribuição exponencial-potência com  $\nu = 0.5$  e Logística tipo-II são apresentadas.

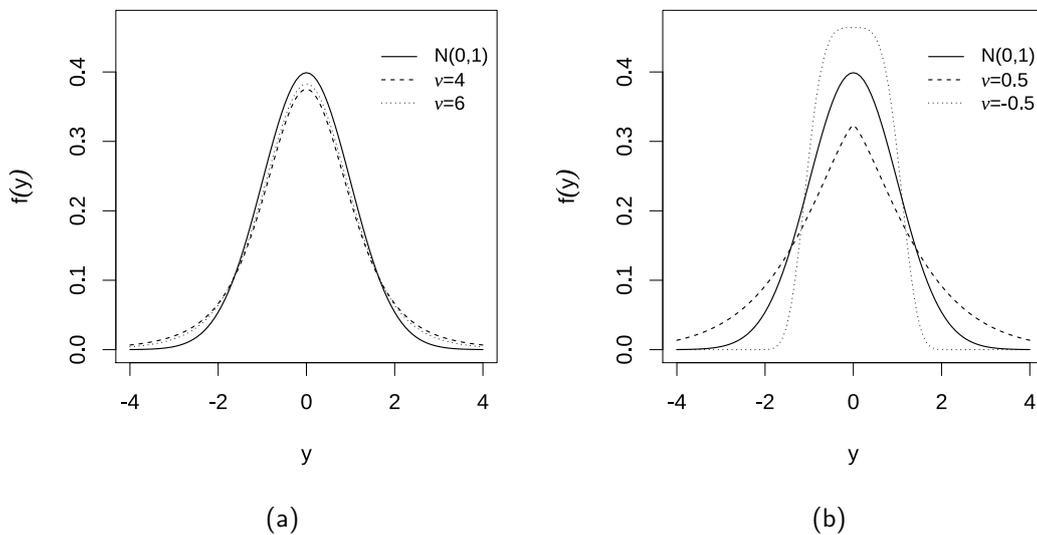


Figura 1 – Funções de densidade para algumas distribuições padrão da classe simétrica, t-Student(a) e exponencial-potência(b).

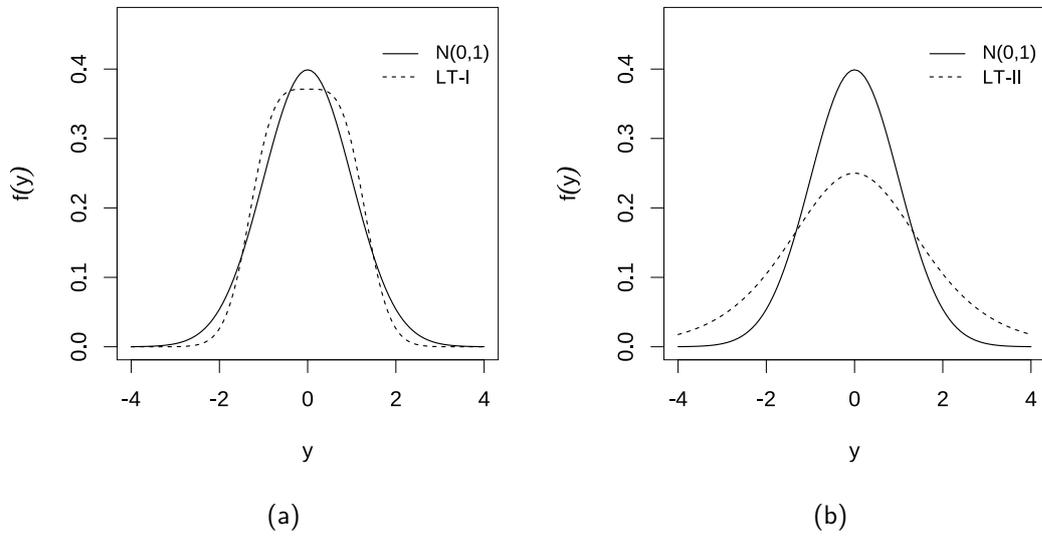


Figura 2 – Funções de densidade para algumas distribuições padrão da classe simétrica, logística-tipo-I(a) e logística-tipo-II(b).

### 1.5.2 Classe log-simétrica de distribuições

Seja  $T$  uma variável aleatória que segue uma classe de distribuições chamada log-simétrica, em que o suporte é definido no intervalo  $(0, \infty)$ . Uma variável aleatória pertence a classe log-simétrica, com parâmetros  $\eta > 0$ ,  $\phi > 0$  e tem função de densidade de probabilidade dada por

$$f_T(t) = \frac{g(\tilde{t}^2)}{t\sqrt{\phi}}, \quad t > 0, \quad (1.2)$$

em que  $\tilde{t} = \log \left[ \left( \frac{t}{\eta} \right)^{\frac{1}{\sqrt{\phi}}} \right]$ .  $T$  é uma transformação da variável aleatória  $Y$  como  $T = \exp(Y)$  cuja distribuição segue a classe simétrica. Se uma variável aleatória segue a classe log-simétrica então é denotada por  $T \sim LS(\eta, \phi, g(\cdot))$ , em que  $\eta = \exp(\mu)$  e  $\phi$  são os parâmetros de escala e potência. Algumas distribuições como log-normal, log-slash, log-Student-t, log-power-exponencial, tipo-I-log-Logistic, tipo-II-log-Logistic, log-normal-contaminada e Birnbaum-Saunders (1969) quando  $\phi = 4$  estão incluídos nesta classe.

Algumas propriedades da classe log-simétrica de distribuições são consequente da classe simétrica. Se  $T^* = (T/\eta)^{\frac{1}{\sqrt{\phi}}}$  então  $T^* \sim LS(1, 1, g(\cdot))$ , uma distribuição log-simétrica padrão. Para  $a > 0$ , segue que  $aT \sim LS(a\eta, \phi, g(\cdot))$ . Para  $b \neq 0$ , segue que  $T^b \sim LS(\eta^b, b^2\phi, g(\cdot))$ .  $(T/\eta)$  e  $(\eta/T)$  são duas variáveis identicamente distribuídas, a função geratriz de momentos de  $Y$ , quando existir, logo  $E(T^r) = M_Y(y)$ . Porque os momentos não são finitos ou podem chegar a ser difícil de obter, Vanegas e Paula (2016) derivam algumas medidas frequentemente usadas tais como locação, dispersão, dispersão relativa, assimetria e curtose com base nos quantis. A mediana de  $T$  é dada por  $\eta$ , o parâmetro de dispersão é definido por a amplitude interquartílica e é dado por  $2\eta \sinh(\sqrt{\phi} \bar{Y}_\xi^{(0,75)})$ , a dispersão relativa é dada por  $\tanh(\sqrt{\phi} \bar{Y}_\xi^{(0,75)})$ , o coeficiente de assimetria é dada por  $\operatorname{cosech}(\sqrt{\phi} \bar{Y}_\xi^{(q)}) - \operatorname{cotanh}(\sqrt{\phi} \bar{Y}_\xi^{(q)})$  e o coeficiente de curtose é dado por  $[\sinh(\sqrt{\phi} \bar{Y}_\xi^{(7/8)}) - \sinh(\sqrt{\phi} \bar{Y}_\xi^{(5/8)})] / \sinh(\sqrt{\phi} \bar{Y}_\xi^{(6/8)})$ , em que  $\bar{Y}_\xi^{(a)}$  é o  $100(a)\%$  quantil de  $\bar{Y} \sim S(0, 1, g(\cdot))$ . Observe que o coeficiente de dispersão relativa e assimetria envolve só o parâmetro  $\phi$ .

Como ilustração, apresentamos as funções de densidades na Figura 3 com  $\eta$  fixo e para diferentes valores de  $\phi$  para as distribuições log-normal, log-t-student com  $\nu = 4$  e log-exponencial-potência com  $\nu = -0.5$ .

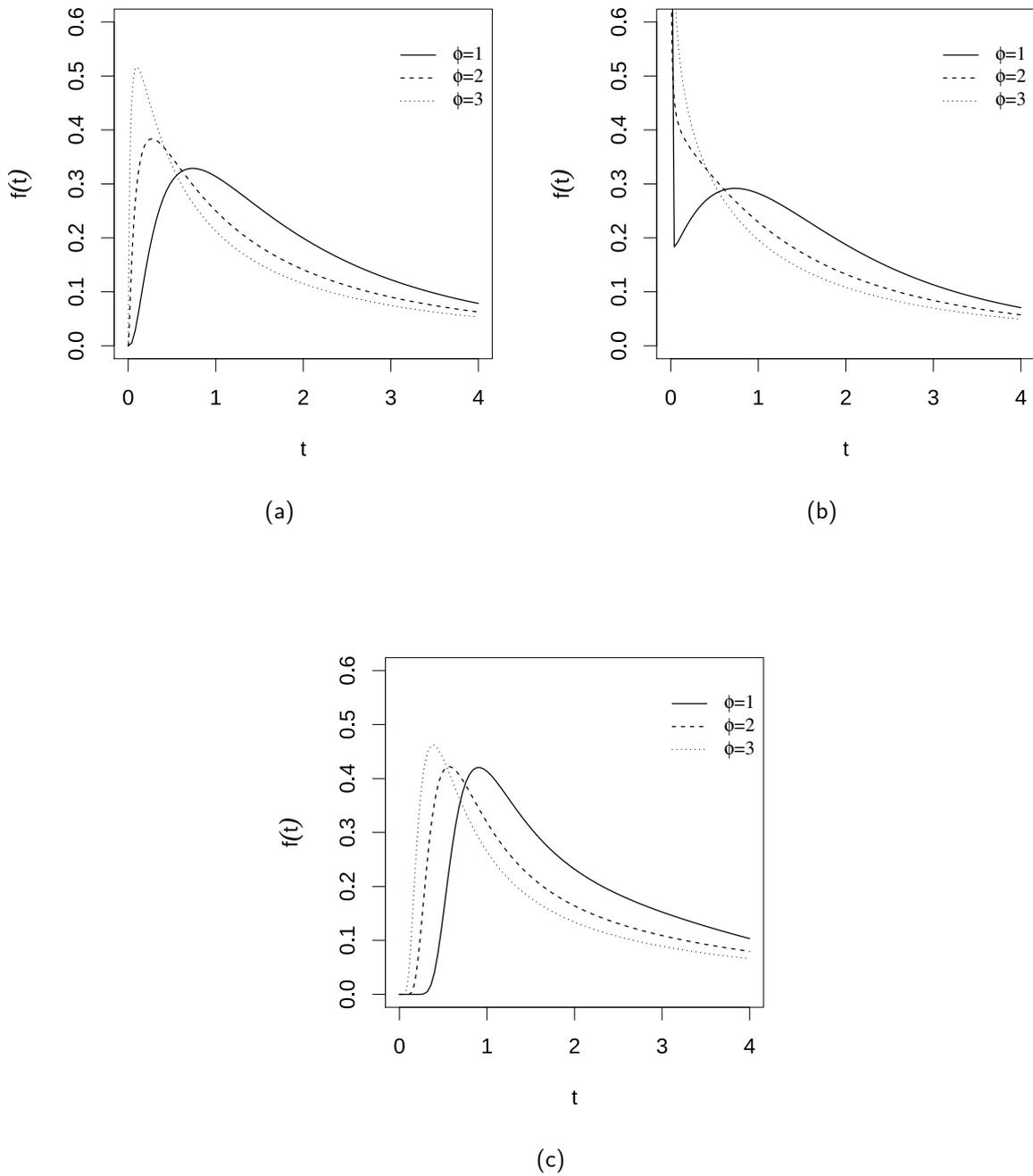


Figura 3 – Funções de densidade para algumas distribuições da classe log-simétrica, log-normal ( $\eta = 2, \phi$ )(a), log-t-Student ( $\eta = 2, \phi, \nu = 4$ )(b), log-exponencial-potência ( $\eta = 2, \phi, \nu = -0.5$ )(c).

## 2 DISTRIBUIÇÕES ZERO AJUSTADO LOG-SIMÉTRICAS

No presente capítulo apresentamos uma nova classe de distribuições semicontínuas baseada na classe log-simétrica como componente contínua e uma distribuição Bernoulli como componente discreta. Algumas propriedades e também os estimadores de máxima verossimilhança, estudos de simulação com a finalidade de avaliar o comportamento dos estimadores propostos foram derivados. Por fim, é apresentada uma aplicação utilizando um conjunto de dados reais sobre despesas de vereadores no Reino Unido.

O objetivo do presente capítulo é propor uma nova ferramenta, no sentido de uma nova classe de distribuição, para a análise de variáveis semicontínuas que tem um comportamento assimétrico na presença de observações atípicas e com excesso de valores zeros.

### 2.1 DEFINIÇÃO

Quando houver a presença de zeros nos dados, as distribuições log-simétricas não são apropriadas. Uma alternativa é usar uma distribuição de mistura entre uma distribuição discreta (Bernoulli) e outra contínua. Seja  $W$  uma variável de mistura entre duas variáveis aleatórias, uma seguindo uma distribuição Bernoulli, componente discreta, e uma que segue a classe log-simétrica, componente contínua; denotamos por  $W \sim ZALS(\eta, \phi, \pi, g(\cdot))$ . A função de distribuição acumulada (FDA) de  $W$  é dada por

$$F_W(w) = \begin{cases} \pi, & \text{se } w = 0, \\ \pi + (1 - \pi)F_T(w), & \text{se } w > 0, \end{cases} \quad (2.1)$$

em que  $F_T(w)$  é a função de distribuição acumulada da classe log-simétrica e  $0 < \pi < 1$  é o parâmetro de mistura. Como  $T$  é uma transformação para a variável aleatória  $Y$ , a função de distribuição acumulada de  $T$  pode ser expressada como  $F_T(t) = F_{\bar{Y}}(\tilde{t})$ , onde  $\bar{Y} = (Y - \mu)/\sqrt{\phi} \sim S(0, 1, g(\cdot))$ . Vamos denotar esta nova classe como zero ajustada log-simétrica sendo a função de densidade correspondente dada por

$$f_W(w) = \begin{cases} \pi, & \text{se } w = 0, \\ (1 - \pi)f_T(w), & \text{se } w > 0, \end{cases} \quad (2.2)$$

em que  $f_T(w)$  é a função de densidade fornecida na Equação (1.2).

Algumas propriedades para a classe zero ajustada log-simétrica são apresentadas

(P1) Se  $W \sim ZALS(\eta, \phi, \pi, g(\cdot))$  logo segue que  $cW \sim ZALS(c\eta, \phi, \pi, g(\cdot))$  para  $c > 0$ .

(P2) A função quantílica de  $W \sim ZALS(\eta, \phi, \pi, g(\cdot))$  pode ser obtida mediante Castellacci (2012)

$$F_W^{-1}(q; \eta, \phi, \pi, g(\cdot)) = \begin{cases} 0, & \text{se } \pi \geq q, \\ \eta \exp\left(\sqrt{\phi} \bar{Y}_\xi^{\left(\frac{q-\pi}{1-\pi}\right)}\right), & \text{se } \pi < q, \end{cases} \quad (2.3)$$

em que  $0 < q < 1$  e  $\bar{Y}_\xi^{(a)}$  é o 100(a)% quantil de  $\bar{Y} \sim S(0, 1, g(\cdot))$ .

(P3) O  $r$ -th momento de  $W$  existe se a função geradora de momentos de  $Y$ ,  $M_Y(r)$  existe. Logo  $E(W^r) = (1 - \pi)M_Y(r)$ .

(P4) Se  $M_Y(r)$  existe, logo podemos expandir a função característica de  $W$  mediante uma serie de Taylor

$$\varphi(r) = 1 + (1 - \pi) \sum_{k=1}^{\infty} M_Y(r) \frac{(ir)^k}{k!}. \quad (2.4)$$

(P5) A função de risco de  $W$  é 1 para  $w = 0$  e  $r_T(w) = f_Y(w)/[1 - F_T(w)]$  para  $w > 0$ . De acordo com Vanegas e Paula (2016), o  $r_T(w)$  da classe log-simétrica é muito flexível; em comparação com as funções de risco da distribuição Gama com forma crescente, decrescente ou constante, ou a distribuição Gaussiana Inversa com forma de  $\cap$ , a função de risco para a classe log-simétrica pode tomar diferentes formas. Por exemplo, crescente, decrescente, de forma de banheira invertida e de banheira.

(P6) A entropia de Shannon,  $ET(w)$ , é definida por  $E[-\log f_W(w)]$ . Para obter a entropia de shannon podemos utilizar a propriedade  $E[-\log f_W(w)] = E[E[-\log f_W(w)|I_{\{0\}}(w)]]$ . Depois de algumas manipulações,  $ET(w)$  pode ser expressada como  $ET(w) = (1 - \pi)[\log(\eta\sqrt{\pi}) + ET(\bar{Y})]$  somente se  $ET(\bar{Y})$  tem forma fechada.

Na Figura 4 é apresentado diferentes funções de densidade da classe zero ajustada log-simétrica com  $\eta = 2$  e  $\pi = 0, 4$ , a saber, zero ajustada log-normal, zero ajustada log-t-Student com  $\nu = 4$  e zero ajustada log-exponencial-potência com  $\nu = -0, 5$  para diferentes valores de  $\phi$ .

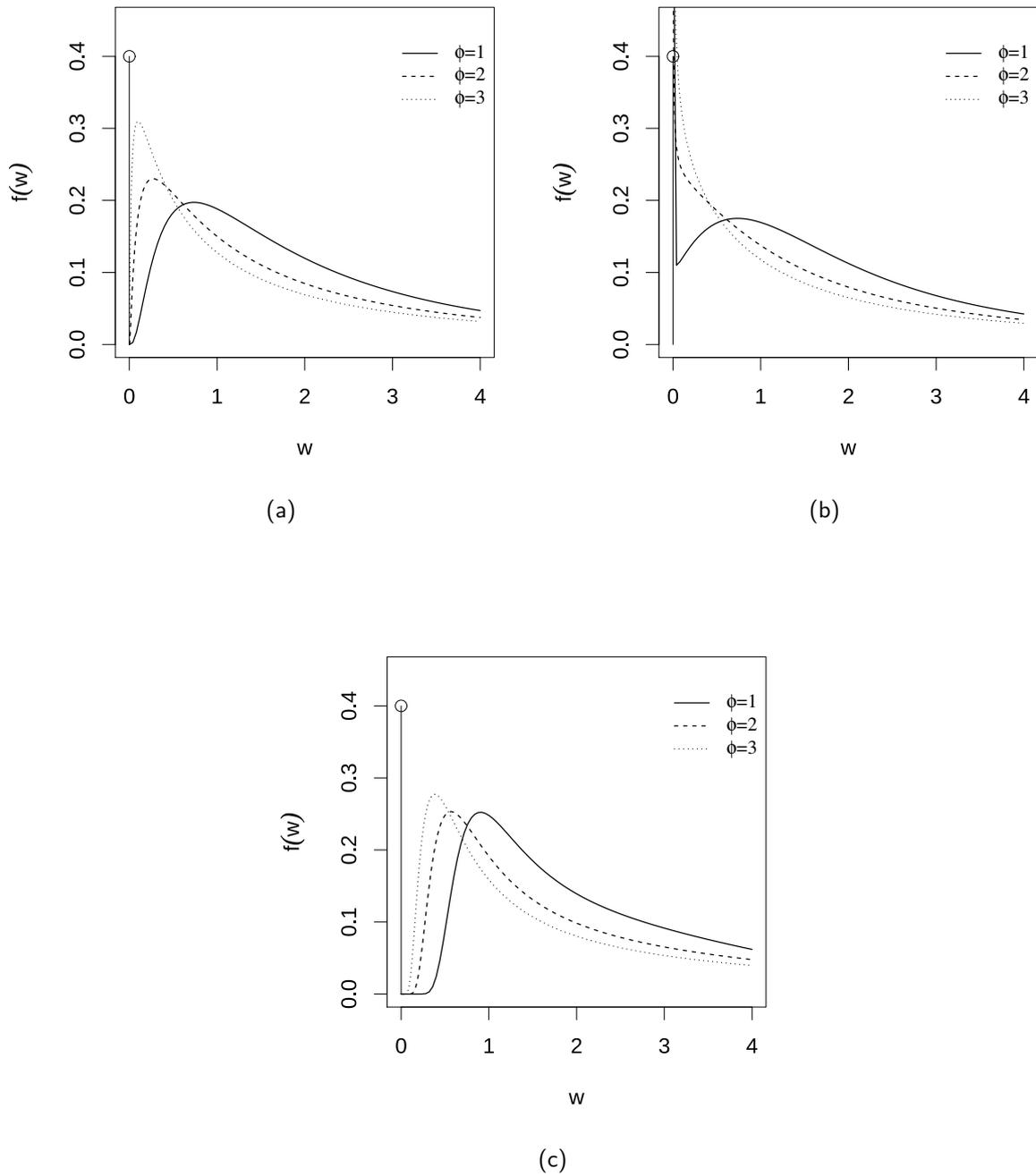


Figura 4 – Funções de densidade para algumas distribuições da classe zero ajustada log-simétrica, zero ajustado log-normal ( $\eta = 2, \pi = 0, 4$ )(a), zero ajustado log-t-Student ( $\eta = 2, \pi = 0, 4, \nu = 4$ )(b), zero ajustado log-exponencial-potência ( $\eta = 2, \pi = 0, 4, \nu = -0, 5$ )(c).

Apresentamos um algoritmo para gerar números aleatórios de uma  $W \sim ZALS(\eta, \phi, \pi, g(\cdot))$  baseado no método da transformada inversa.

**Algoritmo 1 - Geração de números aleatórios:**

Passo 1: Gerar um valor aleatório  $u$  de  $U \sim Uniforme(0, 1)$ .

Passo 2: Baseado na propriedade (P2) fazer

$$w = 0, \text{ se } u \leq \pi;$$

$$w = \eta \exp \left( \sqrt{\phi Y_{\xi}^{\left(\frac{q-\pi}{1-\pi}\right)}} \right), \text{ se } u > \pi.$$

Passo 3: Fazer passo 1 e 2 até chegar ao número requerido de amostra de tamanho  $n$ .

**2.1.1 Estimação de parâmetros**

Sejam  $W_1, \dots, W_n$   $n$  variáveis aleatórias independentes em que seguem  $W_i \sim ZALS(\eta, \phi, \pi, g(\cdot))$ . A função de verossimilhança correspondente para  $\theta = (\eta, \phi, \pi)^T$  é dada por

$$L(\theta) = \prod_{i=1}^n f_W(w_i) = L_1(\pi) L_2(\eta, \phi),$$

em que

$$L_1(\pi) = \prod_{i=1}^n \pi^{I_{\{0\}}(w_i)} (1 - \pi)^{1 - I_{\{0\}}(w_i)}, \quad L_2(\eta, \phi) = \prod_{i=1}^n f_T(w_i)^{1 - I_{\{0\}}(w_i)}.$$

O logaritmo da função de verossimilhança pode ser expressada como  $\ell(\theta) = \ell_{(1)}(\pi) + \ell_{(2)}(\eta, \phi)$ , em que  $\ell_{(1)}(\pi) = W^* \log(\pi) + (n - W^*) \log(1 - \pi)$  e  $\ell_{(2)}(\eta, \phi) = (n - W^*) \left( -\frac{1}{2} \log(\phi) \right) + \sum_{i:w_i > 0} \log[g(\tilde{w}_i^2)] + \sum_{i:w_i > 0} \log(w_i)$ .  $W^*$  é o número de zeros na amostra de tamanho  $n$ ,  $W^* = \sum_{i=1}^n I_{\{0\}}(w_i)$ . O logaritmo da função de verossimilhança pode ser fatorada em dois termos, um que depende apenas de  $\pi$  e outro que depende dos parâmetros da componente contínua  $Y \sim LS(\eta, \phi, g(\cdot))$ . Assim, o método de máxima verossimilhança pode ser aplicado separadamente de  $(\eta, \phi)^T$  e  $\pi$  (Pace e Salvani 1997).

Derivando o logaritmo da função de verossimilhança com respeito a cada parâmetro obtemos o vetor escore em que

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_\eta(\boldsymbol{\theta}) \\ \mathbf{U}_\phi(\boldsymbol{\theta}) \\ \mathbf{U}_\pi(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\eta\phi} \sum_{i:w_i>0} \log(w_i/\eta)^{v(\tilde{w}_i)} \\ \left(-\frac{n-W^*}{2\phi}\right) + \frac{1}{2\phi} \sum_{i:w_i>0} v(\tilde{w}_i)\tilde{w}_i^2 \\ \frac{W^*-n\pi}{\pi(1-\pi)} \end{bmatrix}. \quad (2.5)$$

Seja  $\ddot{L}_{\theta\theta} = \partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$  a matriz de segundas derivadas em relação aos parâmetros, tomando o valor esperado da matriz de segundas derivadas obtemos a matriz de informação de Fisher em que

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{bmatrix} k_{\eta\eta} & 0 & 0 \\ 0 & k_{\phi\phi} & 0 \\ 0 & 0 & k_{\pi\pi} \end{bmatrix} \quad (2.6)$$

sendo  $k_{\eta\eta} = n(1-\pi)d_g/(\phi\eta^2)$ ,  $k_{\phi\phi} = n(1-\pi)(f_g-1)/(4\phi^2)$  e  $k_{\pi\pi} = n/[\phi(1-\phi)]$  com  $d_g = E[v^2(\tilde{W})\tilde{W}^2|I_{\{0\}}(W)=0]$  e  $f_g = E[v^2(\tilde{W})\tilde{W}^4|I_{\{0\}}(W)=0]$  para  $\tilde{W}|I_{\{0\}}(W)=0 \sim S(0,1,g(\cdot))$ .

Tabela 1 – Valores para  $d_g$  e  $f_g$  para algumas distribuições da classe simétrica  $Y$ .

Distribuição	$d_g$	$f_g$
Normal	1	3
t-Student	$\frac{(\nu+1)}{(\nu+3)}$	$\frac{3(\nu+1)}{(\nu+3)}$
Exponencial potencia	$\frac{2^{1-\nu}\Gamma[(3-\nu)/2]}{(1+\nu)^2\Gamma[(1+\nu)/2]}$	$\frac{(\nu+3)}{(\nu+1)}$
Logística tipo-I	0,09232751	0,2508615

Para obter as estimativas de máxima verossimilhança (EMV) de  $\boldsymbol{\theta}$ , cada elemento do vetor escore da Equação (2.5) deve ser igualado a zero, em que o EMV para  $\pi$  é  $W^*/n$ , representa a proporção de zeros na amostra. Para os parâmetros correspondentes a componente contínua,  $\eta$  e  $\phi$ , não existe uma forma fechada para obter os EMV, requerendo um algoritmo de otimização não-linear tais como Newton-Raphson, escore de Fisher, BHHH, entre outros.

Pela separabilidade entre os parâmetros  $\pi$  e  $(\eta, \phi)^T$  e dado que as condições de regularidade são satisfeitas para as distribuições log-simétrica exceto para algumas distribuições como por exemplo, a distribuição log-Laplace (Vanegas e Paula 2016), e também para o EMV de  $\pi$  temos que  $\hat{\theta}$  é um estimador consistente de  $\theta$  e segue que assintoticamente temos  $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}_3(0, \mathbf{K}(\theta)^{-1})$ , em que  $\mathbf{K}(\theta)^{-1}$  é a matriz de variâncias e covariâncias de  $\hat{\theta}$ . Como em situações reais não conhecemos os verdadeiros valores dos parâmetros, segue que um estimador consistente de  $\mathbf{K}(\theta)^{-1}$  é  $\mathbf{K}(\hat{\theta})^{-1}$ . Outro ponto a mencionar-se é que podemos ver que os EMV são assintoticamente independentes como pode ser observado na Equação (2.6).

Intervalos de confiança assintóticos (IC) para  $\theta$  podem ser obtidos da distribuição assintótica, mas se nosso interesse é de estimar uma função de  $\theta$ , i.e,  $h(\theta)$ , logo, o método delta (Lehmann & Casella 1998, Sect. 1.9) pode ser usado. Assim, usando o método delta para obter intervalos de confiança para  $\log(\eta)$ ,  $\log(\phi)$  e  $\text{logit}(\pi) = \log(\pi/(1 - \pi))$  com um nível  $100 \times [1 - \xi]\%$  de confiança são dados por

$$\begin{aligned} \text{IC}(\log(\eta); [1 - \xi] \times 100\%) &= \log(\hat{\eta}) \pm z_{(1-\xi/2)} \sqrt{\frac{\hat{\phi}}{nd_g(1 - \hat{\pi})}} \\ \text{IC}(\log(\phi); [1 - \xi] \times 100\%) &= \log(\hat{\phi}) \pm z_{(1-\xi/2)} \sqrt{\frac{4}{n(f_g - 1)(1 - \hat{\pi})}} \\ \text{IC}(\log(\pi/(1 - \pi)); [1 - \xi] \times 100\%) &= \text{logit}(\hat{\pi}) \pm z_{(1-\xi/2)} \sqrt{\frac{1}{n\hat{\pi}(1 - \hat{\pi})}} \end{aligned}$$

em que  $\text{logit}(\hat{\pi}) = \log(\hat{\pi}/(1 - \hat{\pi}))$ ,  $0 < \xi < 1/2$  e  $z_{(1-\xi/2)}$  representa o quantil  $1 - \xi/2$  da distribuição normal padrão.

## 2.2 ESTUDO DE SIMULAÇÃO

Para avaliar o desempenho dos estimadores de máxima verossimilhança dos parâmetros das distribuições da classe ZALS, um estudo de simulação de Monte Carlo é proposto. Neste estudo consideramos algumas distribuições com caudas mais leves/pesadas que a distribuição zero ajustada log-normal tal como as distribuições zero ajustada log-t-Student (ZALSt) e a zero ajustada log-exponencial-potência (ZALPE) com  $\nu = 4$  e  $\nu = -0,5$  como parâmetro extra, respectivamente. Em cada cenário é considerado tamanhos amostrais,  $n = 10, 30$  e  $50$ , parâmetro de assimetria (ou dispersão relativa) para a componente contínua,  $\phi \in \{1, 2, 3\}$  e o parâmetro da proporção,  $\pi \in \{0, 1, 0, 2, 0, 3, 0, 4\}$ . Para o caso de  $\eta$  foi considerado fixo

para um único valor,  $\eta = 2$ , e o número de réplicas para o estudo de Monte Carlo foi de 5000. Para gerar amostras das distribuições consideradas, foi utilizado o Algoritmo 1.

As Tabelas 2-4 apresentam os resultados das simulações para as distribuições ZALN, ZALSt e ZALPE contendo algumas medidas estatísticas para avaliar os estimadores de máxima verossimilhança como média empírica, mediana, desvio padrão (DP), Erro Quadrático Médio (EQM),  $\sqrt{\text{EQM}}$  e viés de estimação.

Podemos notar nas Tabela 2-4 que a média empírica de  $\hat{\eta}$  incrementa ao momento que  $\phi$  toma valores maiores, não entanto, para a mediana de  $\hat{\eta}$ , apresenta um comportamento similar para os diferentes valores de  $\phi$ . Mas note que quando a proporção de zeros na amostra incrementa, o valor da média empírica e mediana de  $\hat{\eta}$  são afetados, apresentando desempenhos inferiores. Como era de se esperar, a medida que o tamanho amostral crescer o valor da média empírica assim como a mediana ficam muito próximo do valor real de  $\hat{\eta}$ . Em particular, o menor viés para  $\hat{\eta}$  é observado para a distribuição ZALPE (ver Tabela 4). Para o caso de  $\hat{\phi}$ , para os três estudos de simulação, o valor da média empírica e mediana são afetados, diminuem, por o incremento de  $\pi$ , porém incrementando o tamanho amostral, esses valores ficam aproximando-se ao valor real de  $\phi$ . Já para o caso do estimador de máxima verossimilhança de  $\pi$ , os valores da média empírica assim como da mediana estiveram muito próximo do valor real.

O viés do EMV para  $\eta$  está superestimado, isto é, a média empírica é maior que o valor real, em cada cenário para as distribuições consideradas. Como esperado e já antes mencionado, o viés decresce ficando muito próximo de zero a medida que o tamanho amostral aumenta. Pode-se notar que para o caso da distribuição ZALPE, o estimador apresenta melhores resultados em relação ao viés de  $\hat{\eta}$  em comparação para as outras duas distribuições. Nas Tabelas 2 e 4, o viés dos EMV de  $\phi$  subestimam o valor verdadeiro mas isso não acontece no caso da distribuição ZALSt, e no caso do EMV de  $\pi$ , também subestima o valor verdadeiro de  $\pi$ . Vale a pena dizer que o estimador de máxima verossimilhança de  $\pi$  não depende da distribuição já que a estimação da componente discreta é feito separadamente da componente contínua. Em relação ao desvio padrão e o  $\sqrt{\text{EQM}}$ , a distribuição ZALPE apresenta melhores propriedades que as distribuições ZALN e ZALSt para os EMV de  $\eta$  e  $\phi$ .

Nas Tabelas 5-7 mostram os resultados das simulações para os intervalos de confiança assintóticos para os parâmetros  $\log(\eta)$ ,  $\log(\phi)$  e  $\text{logit}(\pi)$  com as respectivas taxas de cobertura assim como para o limite inferior(CI) e superior(CS) considerando diferentes níveis de confiança de 0,90, 0,95 e 0,99. Em geral para cada distribuição podemos ver que a cobertura para o limite inferior do intervalo de confiança assintótico para  $\log(\eta)$  é similar a seu respectivo

limite superior. Porém, para  $\log(\phi)$ , a cobertura para o limite inferior é menor que o do limite superior. Já para  $\log(\pi)$ , a cobertura do limite inferior é maior que o do limite superior, mas para valores maiores de  $\pi$ , as coberturas dos limites inferiores e superiores são similares, isto é, o intervalo confidenciais tendem a ser balanceados. Para o caso das coberturas de probabilidade, a distribuição ZALSt apresenta um melhor desempenho, isto é, tendem a taxa de confiança verdadeira, para os intervalos assintóticos confidenciais para os parâmetros  $\log(\eta)$  e  $\log(\phi)$  que em comparação para as distribuições ZALN e ZALPE para diferentes valores de  $\pi$ . A distribuição ZALPE apresenta um menor desempenho nas coberturas de probabilidade especialmente para o intervalo assintótico confidencial para  $\log(\phi)$  quando a proporção de zeros é maior que 0,1. Outro a mencionar que com o aumento do valor de  $\pi$ , as taxas de cobertura destes parâmetros tendem a diminuir do valor de confiança estipulado. No caso do  $\text{logit}(\pi)$ , a taxa de cobertura tem as mesmas propriedades independente da distribuição por razões já antes mencionadas. Como esperado, para esse parâmetro, a taxa de cobertura aumenta a medida que o tamanho amostral aumenta.

Nas Figuras 5 a 7 são apresentadas os respectivos histogramas das simulações para distribuição assintótica dos estimadores  $\log(\hat{\eta})$ ,  $\log(\hat{\phi})$  e  $\text{logit}(\hat{\pi})$  para cada tamanho amostral e tomando o cenário com maior proporção de zeros e  $\phi = 3$ ,  $\theta = (2, 3, 0, 4)^\top$ . Decidiu-se tomar esse cenário por ser interessante, já que apresentou os piores desempenhos dos estimadores, isto é maior vies e EQM. Adicionalmente, os seguimentos de retas representam os intervalos assintóticos para cada nível confiança, e a linha vertical representa o valor real do parâmetro. Assim, podemos notar um comportamento simétrico para o estimador de máxima verossimilhança de  $\log(\eta)$ , já para o caso dos EMV de  $\log(\phi)$  e  $\text{logit}(\pi)$ , eles tem uma leve assimetria.

Tabela 2 – Resultados do estudo de simulação para os EMV dos parâmetros  $\eta = 2, \phi, \pi$  de acordo a  $n = 10, 30, 50$  para a distribuição zero ajustada log-normal.

	$\pi$	$\phi$	$n$	Média			Mediana			SD			EQM			$\sqrt{\text{EQM}}$			Viés		
				$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$
	10	2,098	0,889	0,101	1,996	0,808	0,100	0,728	0,450	0,094	0,540	0,215	0,009	0,735	0,464	0,094	0,098	-0,111	0,001	0,001	
	1	30	2,044	0,966	0,100	1,994	0,942	0,100	0,401	0,267	0,056	0,163	0,072	0,003	0,403	0,269	0,056	0,044	-0,034	0,000	
	50	2,017	0,979	0,100	1,995	0,962	0,100	0,300	0,210	0,043	0,090	0,044	0,002	0,301	0,211	0,043	0,017	-0,021	-0,000	0,000	
	10	2,266	1,784	0,102	2,030	1,637	0,100	1,160	0,892	0,096	1,416	0,843	0,009	1,190	0,918	0,096	0,266	-0,216	0,002	0,000	
0,1	2	30	2,080	1,936	0,100	2,000	1,893	0,100	0,578	0,537	0,054	0,341	0,293	0,003	0,584	0,541	0,054	0,080	-0,064	0,000	
	50	2,043	1,951	0,099	2,010	1,913	0,100	0,432	0,412	0,043	0,188	0,172	0,002	0,434	0,415	0,043	0,043	-0,049	-0,001	0,000	
	10	2,346	2,664	0,101	2,008	2,450	0,100	1,474	1,351	0,095	2,291	1,939	0,009	1,513	1,392	0,095	0,346	-0,336	0,001	0,001	
	3	30	2,126	2,885	0,101	2,011	2,808	0,100	0,734	0,785	0,055	0,554	0,629	0,003	0,745	0,793	0,055	0,126	-0,115	0,001	
	50	2,054	2,931	0,100	2,001	2,890	0,100	0,540	0,619	0,043	0,295	0,388	0,002	0,543	0,623	0,043	0,054	-0,069	0,000	0,000	
	10	2,146	0,876	0,199	2,030	0,794	0,200	0,778	0,466	0,121	0,627	0,233	0,015	0,792	0,483	0,121	0,146	-0,124	-0,001	0,001	
	1	30	2,043	0,956	0,199	1,999	0,929	0,200	0,420	0,284	0,073	0,179	0,082	0,005	0,423	0,287	0,073	0,043	-0,044	-0,001	
	50	2,021	0,973	0,199	1,997	0,955	0,200	0,319	0,223	0,056	0,102	0,051	0,003	0,320	0,225	0,056	0,021	-0,027	-0,001	0,001	
	10	2,294	1,754	0,198	1,998	1,586	0,200	1,269	0,950	0,122	1,697	0,962	0,015	1,303	0,981	0,122	0,294	-0,246	-0,002	0,000	
0,2	2	30	2,079	1,915	0,202	2,004	1,848	0,200	0,617	0,565	0,074	0,388	0,326	0,005	0,623	0,571	0,074	0,079	-0,085	0,002	
	50	2,047	1,954	0,199	1,998	1,918	0,200	0,476	0,442	0,057	0,229	0,198	0,003	0,478	0,444	0,057	0,047	-0,046	-0,001	0,001	
	10	2,391	2,642	0,199	1,942	2,410	0,200	1,665	1,416	0,122	2,925	2,133	0,015	1,710	1,461	0,122	0,391	-0,358	-0,001	0,001	
	3	30	2,137	2,852	0,201	2,013	2,772	0,200	0,790	0,846	0,074	0,643	0,738	0,005	0,802	0,859	0,074	0,137	-0,148	0,001	
	50	2,077	2,932	0,199	2,013	2,878	0,200	0,572	0,671	0,056	0,333	0,455	0,003	0,577	0,674	0,056	0,077	-0,068	-0,001	0,001	
	10	2,163	0,852	0,283	2,002	0,747	0,300	0,864	0,499	0,129	0,774	0,271	0,017	0,880	0,520	0,130	0,163	-0,148	-0,017	0,001	
	1	30	2,040	0,958	0,299	1,999	0,924	0,300	0,443	0,303	0,081	0,198	0,094	0,007	0,445	0,306	0,081	0,040	-0,042	-0,001	
	50	2,031	0,972	0,300	2,001	0,952	0,300	0,348	0,237	0,064	0,122	0,057	0,004	0,349	0,238	0,064	0,031	-0,028	-0,000	0,000	
	10	2,321	1,704	0,282	2,016	1,524	0,300	1,352	0,986	0,130	1,930	1,060	0,017	1,389	1,030	0,131	0,321	-0,296	-0,018	0,001	
0,3	2	30	2,095	1,899	0,297	1,987	1,839	0,300	0,669	0,600	0,081	0,457	0,370	0,007	0,676	0,608	0,081	0,095	-0,101	-0,003	
	50	2,053	1,940	0,300	1,994	1,901	0,300	0,490	0,477	0,064	0,243	0,231	0,004	0,493	0,481	0,064	0,053	-0,060	-0,000	0,000	
	10	2,496	2,565	0,287	2,034	2,300	0,300	1,895	1,502	0,127	3,834	2,444	0,016	1,958	1,563	0,127	0,496	-0,435	-0,013	0,001	
	3	30	2,143	2,858	0,297	2,000	2,763	0,300	0,847	0,909	0,083	0,738	0,847	0,007	0,859	0,920	0,083	0,143	-0,142	-0,003	
	50	2,075	2,920	0,300	1,987	2,866	0,300	0,622	0,714	0,066	0,392	0,516	0,004	0,626	0,718	0,066	0,075	-0,080	-0,000	0,000	
	10	2,165	0,833	0,348	2,003	0,730	0,400	0,913	0,509	0,118	0,861	0,287	0,017	0,928	0,535	0,129	0,165	-0,167	-0,052	0,001	
	1	30	2,046	0,941	0,383	1,988	0,907	0,400	0,492	0,318	0,076	0,244	0,105	0,006	0,494	0,323	0,078	0,046	-0,059	-0,017	
	50	2,038	0,970	0,392	2,001	0,948	0,400	0,376	0,257	0,061	0,143	0,067	0,004	0,378	0,259	0,062	0,038	-0,030	-0,008	0,000	
	10	2,379	1,682	0,353	2,017	1,455	0,400	1,482	1,061	0,116	2,340	1,226	0,016	1,530	1,107	0,125	0,379	-0,318	-0,047	0,001	
0,4	2	30	2,124	1,876	0,382	2,012	1,800	0,400	0,714	0,650	0,076	0,525	0,437	0,006	0,725	0,661	0,078	0,124	-0,124	-0,018	
	50	2,068	1,935	0,391	2,002	1,893	0,400	0,540	0,505	0,061	0,296	0,259	0,004	0,544	0,509	0,062	0,068	-0,065	-0,009	0,000	
	10	2,554	2,553	0,351	1,993	2,278	0,400	2,081	1,539	0,118	4,634	2,568	0,016	2,153	1,603	0,128	0,554	-0,447	-0,049	0,001	
	3	30	2,164	2,819	0,383	2,001	2,716	0,400	0,889	0,951	0,074	0,818	0,937	0,006	0,904	0,968	0,076	0,164	-0,181	-0,017	
	50	2,097	2,907	0,392	2,002	2,842	0,400	0,681	0,761	0,061	0,473	0,587	0,004	0,688	0,766	0,062	0,097	-0,093	-0,008	0,000	

Tabela 3 – Resultados do estudo de simulação para os EMV dos parâmetros  $\eta = 2, \phi, \pi$  de acordo a  $n = 10, 30, 50$  para a distribuição zero ajustada log-t-Student com  $\nu = 4$ .

	$\pi$	$\phi$	$n$	Média			Mediana			SD			EQM			$\sqrt{\text{EQM}}$			Viés			
				$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	
0,1	1	30	10	2,183	1,007	0,100	1,985	0,827	0,100	1,012	0,717	0,094	1,058	0,513	0,009	1,029	0,716	0,094	0,183	0,007	0,000	
			30	2,057	1,009	0,101	2,012	0,948	0,100	0,467	0,376	0,055	0,221	0,141	0,003	0,470	0,376	0,055	0,057	0,009	0,001	
			50	2,037	1,006	0,100	2,010	0,967	0,100	0,368	0,286	0,043	0,137	0,082	0,002	0,370	0,286	0,043	0,037	0,006	0,000	
	3	30	10	2,321	2,004	0,098	1,983	1,657	0,100	1,469	1,402	0,094	2,261	1,966	0,009	1,504	1,402	0,094	0,321	0,004	-0,002	
			30	2,106	1,998	0,100	1,998	1,893	0,100	0,716	0,747	0,055	0,523	0,557	0,003	0,723	0,746	0,055	0,106	-0,002	0,000	
			50	2,060	2,007	0,100	1,980	1,933	0,100	0,523	0,574	0,042	0,277	0,329	0,002	0,526	0,574	0,042	0,060	0,007	0,000	
	3	30	10	2,561	3,057	0,101	1,996	2,477	0,100	2,123	2,331	0,095	4,819	5,438	0,009	2,195	2,332	0,095	0,561	0,057	0,001	
			30	2,150	2,997	0,101	1,989	2,842	0,100	0,894	1,109	0,055	0,821	1,229	0,003	0,906	1,108	0,055	0,150	-0,003	0,001	
			50	2,087	3,012	0,099	1,993	2,892	0,100	0,653	0,855	0,043	0,434	0,731	0,002	0,659	0,855	0,043	0,087	0,012	-0,001	
	0,2	1	30	10	2,183	0,998	0,196	1,988	0,812	0,200	1,042	0,758	0,124	1,119	0,575	0,015	1,058	0,758	0,124	0,183	-0,002	-0,004
				30	2,049	1,001	0,200	1,989	0,935	0,200	0,508	0,399	0,074	0,260	0,159	0,005	0,510	0,399	0,074	0,049	0,001	0,000
				50	2,029	1,002	0,200	1,991	0,959	0,200	0,381	0,306	0,056	0,146	0,094	0,003	0,382	0,306	0,056	0,029	0,002	-0,000
		3	30	10	2,447	2,014	0,197	2,016	1,589	0,200	1,901	1,609	0,123	3,812	2,588	0,015	1,952	1,609	0,123	0,447	0,014	-0,003
				30	2,130	1,999	0,199	2,009	1,865	0,200	0,755	0,801	0,072	0,587	0,641	0,005	0,766	0,801	0,072	0,130	-0,001	-0,001
				50	2,084	2,006	0,202	2,013	1,924	0,200	0,581	0,612	0,057	0,344	0,374	0,003	0,586	0,611	0,057	0,084	0,006	0,002
3		30	10	2,741	3,003	0,197	1,998	2,431	0,200	2,817	2,422	0,123	8,480	5,863	0,015	2,912	2,421	0,123	0,741	0,003	-0,003	
			30	2,179	3,011	0,200	1,972	2,798	0,200	0,990	1,218	0,072	1,011	1,483	0,005	1,006	1,218	0,072	0,179	0,011	0,000	
			50	2,121	3,026	0,199	1,999	2,891	0,200	0,737	0,921	0,057	0,557	0,850	0,003	0,747	0,922	0,057	0,121	0,026	-0,001	
0,3		1	30	10	2,226	1,006	0,281	1,998	0,788	0,300	1,127	0,849	0,127	1,322	0,721	0,016	1,150	0,849	0,128	0,226	0,006	-0,019
				30	2,081	0,993	0,297	2,012	0,919	0,300	0,570	0,425	0,082	0,332	0,181	0,007	0,576	0,425	0,082	0,081	-0,007	-0,003
				50	2,033	0,999	0,300	1,991	0,947	0,300	0,408	0,328	0,066	0,167	0,107	0,004	0,409	0,328	0,066	0,033	-0,001	0,000
		3	30	10	2,455	1,992	0,286	1,976	1,555	0,300	1,934	1,679	0,128	3,946	2,817	0,017	1,986	1,678	0,129	0,455	-0,008	-0,014
				30	2,137	2,012	0,299	1,995	1,864	0,300	0,833	0,862	0,082	0,712	0,743	0,007	0,844	0,862	0,082	0,137	0,012	-0,001
				50	2,076	2,004	0,299	1,996	1,906	0,300	0,608	0,664	0,064	0,376	0,440	0,004	0,613	0,663	0,064	0,076	0,004	-0,001
	3	30	10	2,824	3,046	0,284	2,002	2,383	0,300	3,072	2,487	0,128	10,111	6,188	0,017	3,180	2,488	0,129	0,824	0,046	-0,016	
			30	2,203	3,002	0,299	1,971	2,797	0,300	1,091	1,297	0,082	1,230	1,681	0,007	1,109	1,296	0,082	0,203	0,002	-0,001	
			50	2,131	3,007	0,300	2,004	2,883	0,300	0,774	0,988	0,063	0,616	0,977	0,004	0,785	0,988	0,063	0,131	0,007	-0,000	
	0,4	1	30	10	2,277	1,002	0,349	2,018	0,760	0,400	1,260	0,892	0,120	1,663	0,795	0,017	1,290	0,892	0,131	0,277	0,002	-0,051
				30	2,070	1,007	0,381	1,980	0,928	0,400	0,603	0,470	0,075	0,368	0,221	0,006	0,607	0,470	0,077	0,070	0,007	-0,019
				50	2,049	1,005	0,392	1,995	0,956	0,400	0,453	0,351	0,061	0,207	0,123	0,004	0,455	0,351	0,061	0,049	0,005	-0,008
		3	30	10	2,527	2,053	0,356	1,987	1,541	0,400	2,025	1,840	0,117	4,377	3,386	0,016	2,092	1,840	0,125	0,527	0,053	-0,044
				30	2,158	2,010	0,383	1,999	1,847	0,400	0,913	0,923	0,074	0,859	0,852	0,006	0,927	0,923	0,076	0,158	0,010	-0,017
				50	2,117	1,997	0,390	2,027	1,892	0,400	0,663	0,714	0,061	0,453	0,510	0,004	0,673	0,714	0,062	0,117	-0,003	-0,010
3		30	10	2,894	3,032	0,352	1,998	2,268	0,400	3,363	3,026	0,119	12,107	9,155	0,017	3,480	3,026	0,129	0,894	0,032	-0,048	
			30	2,259	2,990	0,382	1,996	2,735	0,400	1,221	1,384	0,076	1,557	1,914	0,006	1,248	1,384	0,078	0,259	-0,010	-0,018	
			50	2,150	3,009	0,392	1,999	2,855	0,400	0,841	1,052	0,061	0,729	1,107	0,004	0,854	1,052	0,061	0,150	0,009	-0,008	

Tabela 4 – Resultados do estudo de simulação para os EMV dos parâmetros  $\eta = 2, \phi, \pi$  de acordo a  $n = 10, 30, 50$  para a distribuição zero ajustada log-exponencial-potência com  $\nu = -0, 5$ .

$\pi$	$\phi$	$n$	Média			Mediana			SD			EQM			$\sqrt{\text{EQM}}$			Viés		
			$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$
1	0,1	10	2,052	0,820	0,099	2,008	0,792	0,100	0,450	0,316	0,094	0,205	0,133	0,009	0,453	0,364	0,094	0,052	-0,180	-0,001
		30	2,013	0,944	0,100	1,998	0,939	0,100	0,239	0,194	0,054	0,057	0,041	0,003	0,240	0,202	0,054	0,013	-0,056	-0,000
	50	2,007	0,964	0,100	1,999	0,961	0,100	0,183	0,147	0,042	0,034	0,023	0,002	0,183	0,152	0,042	0,007	-0,036	0,000	
	10	2,095	1,658	0,101	1,992	1,608	0,100	0,644	0,641	0,095	0,424	0,527	0,009	0,651	0,726	0,095	0,095	-0,342	0,001	
	30	2,022	1,889	0,100	1,997	1,872	0,100	0,337	0,384	0,054	0,114	0,160	0,003	0,337	0,400	0,054	0,022	-0,111	-0,000	
	50	2,014	1,936	0,100	2,000	1,937	0,100	0,255	0,299	0,042	0,065	0,093	0,002	0,256	0,305	0,042	0,014	-0,064	-0,000	
3	0,2	10	2,142	2,482	0,102	1,997	2,425	0,100	0,828	0,963	0,097	0,705	1,196	0,009	0,840	1,094	0,097	0,142	-0,518	0,002
		30	2,034	2,825	0,100	1,994	2,802	0,100	0,415	0,567	0,054	0,173	0,353	0,003	0,416	0,594	0,054	0,034	-0,175	-0,000
	50	2,023	2,906	0,100	1,996	2,902	0,100	0,317	0,440	0,043	0,101	0,202	0,002	0,318	0,450	0,043	0,023	-0,094	-0,000	
	10	2,053	0,800	0,198	1,997	0,776	0,200	0,476	0,337	0,124	0,229	0,153	0,015	0,478	0,391	0,124	0,053	-0,200	-0,002	
	30	2,019	0,938	0,199	2,001	0,931	0,200	0,251	0,206	0,073	0,063	0,046	0,005	0,252	0,215	0,073	0,019	-0,062	-0,001	
	50	2,008	0,966	0,202	2,000	0,964	0,200	0,189	0,157	0,056	0,036	0,026	0,003	0,189	0,161	0,057	0,008	-0,034	0,002	
3	0,3	10	2,110	1,618	0,196	2,010	1,565	0,200	0,705	0,681	0,120	0,509	0,609	0,014	0,714	0,781	0,120	0,110	-0,382	-0,004
		30	2,027	1,864	0,199	1,999	1,852	0,200	0,357	0,406	0,073	0,128	0,184	0,005	0,358	0,428	0,073	0,027	-0,136	-0,001
	50	2,014	1,929	0,199	1,995	1,919	0,200	0,270	0,314	0,057	0,073	0,104	0,003	0,270	0,322	0,057	0,014	-0,071	-0,001	
	10	2,162	2,424	0,197	2,019	2,353	0,200	0,873	0,989	0,123	0,788	1,309	0,015	0,888	1,144	0,123	0,162	-0,576	-0,003	
	30	2,052	2,793	0,200	2,002	2,761	0,200	0,454	0,609	0,074	0,209	0,414	0,005	0,457	0,643	0,074	0,052	-0,207	0,000	
	50	2,030	2,894	0,200	2,000	2,876	0,200	0,334	0,481	0,057	0,113	0,242	0,003	0,336	0,492	0,057	0,030	-0,106	0,000	
3	0,4	10	2,058	0,776	0,285	1,995	0,743	0,300	0,518	0,354	0,127	0,272	0,175	0,016	0,521	0,419	0,127	0,058	-0,224	-0,015
		30	2,011	0,927	0,299	1,990	0,921	0,300	0,270	0,213	0,082	0,073	0,051	0,007	0,270	0,225	0,082	0,011	-0,073	-0,001
	50	2,009	0,951	0,301	1,997	0,946	0,300	0,206	0,169	0,064	0,042	0,031	0,004	0,206	0,176	0,064	0,009	-0,049	0,001	
	10	2,106	1,560	0,288	1,989	1,495	0,300	0,749	0,722	0,129	0,573	0,715	0,017	0,757	0,845	0,129	0,106	-0,440	-0,012	
	30	2,032	1,859	0,298	1,997	1,843	0,300	0,383	0,432	0,082	0,147	0,207	0,007	0,384	0,455	0,082	0,032	-0,141	-0,002	
	50	2,022	1,914	0,300	2,005	1,905	0,300	0,292	0,342	0,064	0,086	0,124	0,004	0,293	0,352	0,064	0,022	-0,086	-0,000	
3	0,5	10	2,175	2,323	0,283	1,983	2,240	0,300	0,989	1,048	0,127	1,009	1,555	0,016	1,004	1,247	0,128	0,175	-0,677	-0,017
		30	2,063	2,768	0,301	2,011	2,741	0,300	0,488	0,647	0,081	0,242	0,473	0,006	0,492	0,688	0,081	0,063	-0,232	0,001
	50	2,034	2,873	0,298	2,006	2,852	0,300	0,359	0,507	0,064	0,130	0,273	0,004	0,361	0,523	0,064	0,034	-0,127	-0,002	
	10	2,068	0,763	0,351	1,998	0,726	0,400	0,553	0,368	0,119	0,310	0,191	0,017	0,557	0,438	0,129	0,068	-0,237	-0,049	
	30	2,019	0,918	0,385	2,000	0,909	0,400	0,294	0,231	0,074	0,087	0,060	0,006	0,295	0,245	0,076	0,019	-0,082	-0,015	
	50	2,013	0,946	0,392	1,998	0,941	0,400	0,223	0,179	0,061	0,050	0,035	0,004	0,224	0,187	0,062	0,013	-0,054	-0,008	
3	0,4	10	2,142	1,513	0,351	2,000	1,434	0,400	0,834	0,736	0,117	0,715	0,779	0,016	0,846	0,883	0,127	0,142	-0,487	-0,049
		30	2,036	1,826	0,384	1,998	1,808	0,400	0,423	0,461	0,075	0,180	0,243	0,006	0,424	0,493	0,076	0,036	-0,174	-0,016
	50	2,025	1,894	0,390	2,007	1,883	0,400	0,308	0,364	0,063	0,095	0,144	0,004	0,309	0,379	0,064	0,025	-0,106	-0,010	
	10	2,228	2,278	0,353	1,997	2,168	0,400	1,094	1,103	0,119	1,249	1,738	0,016	1,118	1,318	0,127	0,228	-0,722	-0,047	
	30	2,062	2,762	0,381	2,006	2,739	0,400	0,515	0,690	0,074	0,269	0,533	0,006	0,518	0,730	0,076	0,062	-0,238	-0,019	
	50	2,030	2,851	0,392	1,990	2,826	0,400	0,387	0,547	0,061	0,151	0,321	0,004	0,388	0,567	0,061	0,030	-0,149	-0,008	

Tabela 5 – Cobertura de probabilidade para os EMV dos parâmetros  $\log(\eta)$ ,  $\log(\phi)$  e  $\log(\pi/(1-\pi))$  de acordo a  $n = 10, 30, 50$  e  $100 \times [1 - \alpha]\%$  para a distribuição zero ajustada log-normal.

$\pi$	$\phi$	$n$	90%															95%															99%																																																					
			CI(%)					CS(%)					Cobertura(%)					CI(%)					CS(%)					Cobertura(%)					CI(%)					CS(%)					Cobertura(%)																																											
			$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$																														
1	30	10	7,70	1,40	6,80	8,88	15,36	0,00	83,42	83,24	93,20	93,20	4,78	0,48	1,06	5,74	10,68	0,00	89,48	88,84	98,94	1,68	0,06	1,06	2,00	4,12	0,00	96,32	95,82	98,94	7,70	1,40	6,80	8,88	15,36	0,00	83,42	83,24	93,20	93,20	4,78	0,48	1,06	5,74	10,68	0,00	89,48	88,84	98,94	1,68	0,06	1,06	2,00	4,12	0,00	96,32	95,82	98,94	7,70	1,40	6,80	8,88	15,36	0,00	83,42	83,24	93,20	93,20	4,78	0,48	1,06	5,74	10,68	0,00	89,48	88,84	98,94	1,68	0,06	1,06	2,00	4,12	0,00	96,32	95,82	98,94
		30	6,20	2,42	7,92	5,66	9,86	0,00	88,14	87,72	92,08	92,08	3,48	0,78	2,88	2,80	6,04	0,00	93,72	93,18	97,12	1,00	0,04	1,00	0,84	2,14	0,00	98,16	97,82	99,00	6,20	2,42	7,92	5,66	9,86	0,00	88,14	87,72	92,08	92,08	3,48	0,78	2,88	2,80	6,04	0,00	93,72	93,18	97,12	1,00	0,04	1,00	0,84	2,14	0,00	98,16	97,82	99,00	6,20	2,42	7,92	5,66	9,86	0,00	88,14	87,72	92,08	92,08	3,48	0,78	2,88	2,80	6,04	0,00	93,72	93,18	97,12	1,00	0,04	1,00	0,84	2,14	0,00	98,16	97,82	99,00
		50	5,16	2,98	5,90	5,50	7,72	2,98	89,34	89,30	91,12	91,12	2,58	1,08	2,70	3,26	4,36	0,00	94,16	94,56	97,30	0,50	0,18	1,16	0,82	1,46	0,00	98,68	98,36	98,84	5,16	2,98	5,90	5,50	7,72	2,98	89,34	89,30	91,12	91,12	2,58	1,08	2,70	3,26	4,36	0,00	94,16	94,56	97,30	0,50	0,18	1,16	0,82	1,46	0,00	98,68	98,36	98,84	5,16	2,98	5,90	5,50	7,72	2,98	89,34	89,30	91,12	91,12	2,58	1,08	2,70	3,26	4,36	0,00	94,16	94,56	97,30	0,50	0,18	1,16	0,82	1,46	0,00	98,68	98,36	98,84
0,1	2	10	9,10	1,14	7,64	8,06	15,28	0,00	82,84	83,58	92,36	92,36	5,48	0,44	1,26	4,92	10,68	0,00	89,60	88,88	98,74	2,02	0,00	1,26	1,72	4,54	0,00	96,26	95,46	98,74	9,10	1,14	7,64	8,06	15,28	0,00	82,84	83,58	92,36	92,36	5,48	0,44	1,26	4,92	10,68	0,00	89,60	88,88	98,74	2,02	0,00	1,26	1,72	4,54	0,00	96,26	95,46	98,74	9,10	1,14	7,64	8,06	15,28	0,00	82,84	83,58	92,36	92,36	5,48	0,44	1,26	4,92	10,68	0,00	89,60	88,88	98,74	2,02	0,00	1,26	1,72	4,54	0,00	96,26	95,46	98,74
		30	5,74	2,66	6,56	6,32	9,12	0,00	87,94	88,22	93,44	93,44	3,10	0,96	2,10	3,16	5,46	0,00	93,74	93,58	97,90	0,94	0,02	0,64	0,96	1,78	0,00	98,10	98,20	99,36	5,74	2,66	6,56	6,32	9,12	0,00	87,94	88,22	93,44	93,44	3,10	0,96	2,10	3,16	5,46	0,00	93,74	93,58	97,90	0,94	0,02	0,64	0,96	1,78	0,00	98,10	98,20	99,36	5,74	2,66	6,56	6,32	9,12	0,00	87,94	88,22	93,44	93,44	3,10	0,96	2,10	3,16	5,46	0,00	93,74	93,58	97,90	0,94	0,02	0,64	0,96	1,78	0,00	98,10	98,20	99,36
		50	5,22	2,84	5,28	6,00	8,12	3,42	88,78	89,04	91,30	91,30	2,60	1,16	2,38	3,22	4,70	0,00	94,18	94,14	97,62	0,68	0,08	1,08	0,74	1,08	0,00	98,58	98,84	98,92	5,22	2,84	5,28	6,00	8,12	3,42	88,78	89,04	91,30	91,30	2,60	1,16	2,38	3,22	4,70	0,00	94,18	94,14	97,62	0,68	0,08	1,08	0,74	1,08	0,00	98,58	98,84	98,92	5,22	2,84	5,28	6,00	8,12	3,42	88,78	89,04	91,30	91,30	2,60	1,16	2,38	3,22	4,70	0,00	94,18	94,14	97,62	0,68	0,08	1,08	0,74	1,08	0,00	98,58	98,84	98,92
3	10	10	8,14	1,20	6,54	7,56	15,30	0,00	84,30	83,50	93,46	93,46	5,18	0,52	1,26	4,86	10,76	0,00	89,96	88,72	98,74	2,32	0,02	1,26	2,02	4,78	0,00	95,66	95,20	98,74	8,14	1,20	6,54	7,56	15,30	0,00	84,30	83,50	93,46	93,46	5,18	0,52	1,26	4,86	10,76	0,00	89,96	88,72	98,74	2,32	0,02	1,26	2,02	4,78	0,00	95,66	95,20	98,74	8,14	1,20	6,54	7,56	15,30	0,00	84,30	83,50	93,46	93,46	5,18	0,52	1,26	4,86	10,76	0,00	89,96	88,72	98,74	2,32	0,02	1,26	2,02	4,78	0,00	95,66	95,20	98,74
		30	5,82	1,84	7,28	5,82	10,62	0,00	88,36	87,54	92,72	92,72	3,08	0,84	2,18	3,16	6,64	0,00	93,76	92,52	97,82	0,90	0,08	0,78	1,08	2,16	0,00	98,02	97,76	99,22	5,82	1,84	7,28	5,82	10,62	0,00	88,36	87,54	92,72	92,72	3,08	0,84	2,18	3,16	6,64	0,00	93,76	92,52	97,82	0,90	0,08	0,78	1,08	2,16	0,00	98,02	97,76	99,22	5,82	1,84	7,28	5,82	10,62	0,00	88,36	87,54	92,72	92,72	3,08	0,84	2,18	3,16	6,64	0,00	93,76	92,52	97,82	0,90	0,08	0,78	1,08	2,16	0,00	98,02	97,76	99,22
		50	5,02	2,58	5,98	6,08	8,12	2,82	88,90	89,30	91,20	91,20	2,76	1,12	2,56	2,86	4,82	0,00	94,38	94,06	97,44	0,60	0,10	0,94	0,72	1,36	0,00	98,02	97,76	99,06	5,02	2,58	5,98	6,08	8,12	2,82	88,90	89,30	91,20	91,20	2,76	1,12	2,56	2,86	4,82	0,00	94,38	94,06	97,44	0,60	0,10	0,94	0,72	1,36	0,00	98,02	97,76	99,06	5,02	2,58	5,98	6,08	8,12	2,82	88,90	89,30	91,20	91,20	2,76	1,12	2,56	2,86	4,82	0,00	94,38	94,06	97,44	0,60	0,10	0,94	0,72	1,36	0,00	98,02	97,76	99,06
1	30	10	8,84	1,04	3,52	7,50	16,26	0,00	83,66	82,70	96,48	96,48	5,74	0,34	3,52	4,90	11,28	0,00	89,36	88,38	96,48	2,60	0,00	0,74	2,40	5,28	0,00	95,00	94,72	99,26	8,84	1,04	3,52	7,50	16,26	0,00	83,66	82,70	96,48	96,48	5,74	0,34	3,52	4,90	11,28	0,00	89,36	88,38	96,48	2,60	0,00	0,74	2,40	5,28	0,00	95,00	94,72	99,26	8,84	1,04	3,52	7,50	16,26	0,00	83,66	82,70	96,48	96,48	5,74	0,34	3,52	4,90	11,28	0,00	89,36	88,38	96,48	2,60	0,00	0,74	2,40	5,28	0,00	95,00	94,72	99,26
		30	6,06	2,08	5,96	5,96	10,28	4,30	87,98	87,64	89,74	89,74	3,22	0,74	2,36	3,36	6,18	0,00	93,42	93,08	97,64	0,94	0,06	0,86	0,88	2,10	0,00	98,18	97,84	99,14	6,06	2,08	5,96	5,96	10,28	4,30	87,98	87,64	89,74	89,74	3,22	0,74	2,36	3,36	6,18	0,00	93,42	93,08	97,64	0,94	0,06	0,86	0,88	2,10	0,00	98,18	97,84	99,14	6,06	2,08	5,96	5,96	10,28	4,30	87,98	87,64	89,74	89,74	3,22	0,74	2,36	3,36	6,18	0,00	93,42	93,08	97,64	0,94	0,06	0,86	0,88	2,10	0,00	98,18	97,84	99,14
		50	5,88	2,50	6,44	4,94	8,68	5,40	89,18	88,82	88,16	88,16	3,14	0,96	3,22	2,80	5,10	1,96	94,06	93,94	94,82	0,82	0,06	0,76	0,62	1,42	0,00	98,56	98,52	99,24	5,88	2,50	6,44	4,94	8,68	5,40	89,18	88,82	88,16	88,16	3,14	0,96	3,22	2,80	5,10	1,96	94,06	93,94	94,82	0,82	0,06	0,76	0,62	1,42	0,00	98,56	98,52	99,24	5,88	2,50	6,44	4,94	8,68	5,40	89,18	88,82	88,16	88,16	3,14	0,96	3,22	2,80	5,10	1,96	94,06	93,94	94,82	0,82	0,06	0,76	0,62	1,42	0,00	98,56	98,52	99,24
0,2	2	10	8,64	1,02	3,10	8,06	17,06	0,00	83,30	81,92	96,90	96,90	5,62	0,40	3,10	5,04	11,92	0,00	89,34	87,68	96,90	2,32	0,06	0,62	2,30	5,56	0,00	95,38	94,38	99,38	8,64	1,02	3,10	8,06	17,06	0,00	83,30	81,92	96,90	96,90	5,62	0,40	3,10	5,04	11,92	0,00	89,34	87,68	96,90	2,32	0,06	0,62	2,30	5,56	0,00	95,38	94,38	99,38	8,64	1,02	3,10	8,06	17,06	0,00	83,30	81,92	96,90	96,90	5,62	0,40	3,10	5,04	11,92	0,00	89,34	87,68	96,90	2,32	0,06	0,62	2,30	5,56	0,00	95,38	94,38	99,38
		30	5,70	2,60	6,28	5,62	10,60	4,74	88,68	86,80	88,98	88,98	3,24	1,06	2,84	2,92	6,64	0,00	93,84	92,30	97,16	0,88	0,14	1,04	0,68	2,04	0,00	98,44	97,82	98,96	5,70	2,60	6,28	5,62	10,60	4,74	88,68	86,80	88,98	88,98	3,24	1,06	2,84	2,92	6,64	0,00	93,84	92,30	97,16	0,88	0,14	1,04	0,68	2,04	0,00	98,44	97,82	98,96	5,70	2,60	6,28	5,62	10,60	4,74	88,68	86,80	88,98	88,98	3,24	1,06	2,84	2,92	6,64	0,00	93,84	92,30	97,16	0,88	0,14	1,04	0,68	2,04	0,00	98,44	97,8	

Tabela 6 – Cobertura de probabilidade para os EMV dos parâmetros  $\log(\eta)$ ,  $\log(\phi)$  e  $\log(\pi/(1 - \pi))$  de acordo a  $n = 10, 30, 50$  e  $100 \times [1 - \alpha]\%$  para a distribuição zero ajustada log-t-Student com  $\nu = 4$ .

$\pi$	$\phi$	$n$	90%															95%															99%																							
			CI(%)					Cobertura(%)					CI(%)					CS(%)					Cobertura(%)					CI(%)					CS(%)					Cobertura(%)																		
			$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$																		
0.1	1	10	8.05	2.94	6.86	8.43	11.53	0.00	83.53	85.53	93.14	5.40	1.18	1.24	5.24	6.85	0.00	93.35	91.97	98.76	2.50	0.20	1.24	2.24	2.70	0.00	95.26	97.10	98.76	8.05	2.94	6.86	8.43	11.53	0.00	83.53	85.53	93.14	5.40	1.18	1.24	5.24	6.85	0.00	93.35	91.97	98.76	2.50	0.20	1.24	2.24	2.70	0.00	95.26	97.10	98.76
		30	6.08	3.32	6.60	8.14	0.00	88.06	88.54	93.40	3.38	1.56	2.16	3.28	4.94	0.00	93.34	93.50	97.84	0.74	0.34	0.62	0.78	1.52	0.00	98.48	98.14	99.38	6.08	3.32	6.60	8.14	0.00	88.06	88.54	93.40	3.38	1.56	2.16	3.28	4.94	0.00	93.34	93.50	97.84	0.74	0.34	0.62	0.78	1.52	0.00	98.48	98.14	99.38		
		50	6.04	3.70	6.06	5.30	7.10	3.26	88.66	89.20	90.68	2.98	1.76	2.44	3.20	4.08	0.00	93.82	94.16	97.56	0.66	0.28	0.94	0.80	1.04	0.00	98.54	98.68	99.06	6.04	3.70	6.06	5.30	7.10	3.26	88.66	89.20	90.68	2.98	1.76	2.44	3.20	4.08	0.00	93.82	94.16	97.56	0.66	0.28	0.94	0.80	1.04	0.00	98.54	98.68	99.06
		10	8.03	2.24	6.78	7.53	11.04	0.00	84.44	86.72	93.22	4.91	1.18	1.12	4.75	6.91	0.00	90.35	91.91	98.88	2.04	0.14	1.12	1.88	2.72	0.00	96.07	97.14	98.88	8.03	2.24	6.78	7.53	11.04	0.00	84.44	86.72	93.22	4.91	1.18	1.12	4.75	6.91	0.00	90.35	91.91	98.88	2.04	0.14	1.12	1.88	2.72	0.00	96.07	97.14	98.88
		30	5.81	3.40	7.70	5.75	7.29	0.00	88.45	89.31	92.30	3.06	1.50	2.52	2.90	3.86	0.00	94.03	94.63	97.48	0.76	0.12	0.62	0.66	1.16	0.00	98.58	98.72	99.38	5.81	3.40	7.70	5.75	7.29	0.00	88.45	89.31	92.30	3.06	1.50	2.52	2.90	3.86	0.00	94.03	94.63	97.48	0.76	0.12	0.62	0.66	1.16	0.00	98.58	98.72	99.38
	50	4.97	3.75	6.10	5.31	7.34	2.56	89.71	88.91	91.34	2.95	1.91	2.78	2.85	3.87	0.00	94.20	94.22	97.22	0.80	0.34	1.04	0.66	0.94	0.00	98.54	98.72	98.96	4.97	3.75	6.10	5.31	7.34	2.56	89.71	88.91	91.34	2.95	1.91	2.78	2.85	3.87	0.00	94.20	94.22	97.22	0.80	0.34	1.04	0.66	0.94	0.00	98.54	98.72	98.96	
	10	8.09	2.81	7.44	7.85	11.46	0.00	84.05	85.73	92.56	5.09	1.44	1.48	5.11	7.55	0.00	89.80	91.00	98.52	1.62	0.20	1.48	2.18	2.87	0.00	96.19	96.93	98.52	8.09	2.81	7.44	7.85	11.46	0.00	84.05	85.73	92.56	5.09	1.44	1.48	5.11	7.55	0.00	89.80	91.00	98.52	1.62	0.20	1.48	2.18	2.87	0.00	96.19	96.93	98.52	
	30	5.94	3.31	7.64	5.16	7.91	0.00	88.90	88.78	92.36	3.25	1.35	2.42	2.87	4.42	0.00	93.88	94.24	97.58	0.78	0.22	0.62	0.64	1.06	0.00	98.57	98.71	99.38	5.94	3.31	7.64	5.16	7.91	0.00	88.90	88.78	92.36	3.25	1.35	2.42	2.87	4.42	0.00	93.88	94.24	97.58	0.78	0.22	0.62	0.64	1.06	0.00	98.57	98.71	99.38	
	50	5.79	3.55	5.92	5.85	7.28	3.14	88.36	89.16	90.94	3.23	1.45	2.58	3.29	3.87	0.00	93.48	94.67	97.42	0.89	0.32	0.92	0.81	0.69	0.00	98.31	98.99	99.08	5.79	3.55	5.92	5.85	7.28	3.14	88.36	89.16	90.94	3.23	1.45	2.58	3.29	3.87	0.00	93.48	94.67	97.42	0.89	0.32	0.92	0.81	0.69	0.00	98.31	98.99	99.08	
	10	7.66	2.84	3.74	8.80	12.93	0.00	83.53	84.23	96.26	5.18	1.16	3.74	5.44	8.14	0.00	89.38	90.70	96.26	2.22	0.16	0.72	2.32	3.14	0.00	95.46	96.70	99.28	7.66	2.84	3.74	8.80	12.93	0.00	83.53	84.23	96.26	5.18	1.16	3.74	5.44	8.14	0.00	89.38	90.70	96.26	2.22	0.16	0.72	2.32	3.14	0.00	95.46	96.70	99.28	
	30	5.74	3.28	5.76	6.28	7.98	4.12	87.97	88.73	90.12	3.08	1.44	2.16	3.70	4.56	0.00	93.22	94.00	97.84	0.80	0.30	0.96	1.20	1.54	0.00	98.00	98.16	99.04	5.74	3.28	5.76	6.28	7.98	4.12	87.97	88.73	90.12	3.08	1.44	2.16	3.70	4.56	0.00	93.22	94.00	97.84	0.80	0.30	0.96	1.20	1.54	0.00	98.00	98.16	99.04	
	50	5.86	3.80	6.42	5.90	7.76	4.94	88.23	88.43	88.64	3.10	1.58	3.46	3.04	4.10	1.76	93.86	94.32	94.78	0.90	0.34	0.90	0.82	1.06	0.00	98.28	98.60	99.10	5.86	3.80	6.42	5.90	7.76	4.94	88.23	88.43	88.64	3.10	1.58	3.46	3.04	4.10	1.76	93.86	94.32	94.78	0.90	0.34	0.90	0.82	1.06	0.00	98.28	98.60	99.10	
	10	7.87	2.40	2.90	9.23	12.87	0.00	82.91	84.73	97.10	5.40	0.90	2.90	5.68	8.35	0.00	88.91	90.75	97.10	2.22	0.18	0.58	2.66	3.04	0.00	95.12	96.78	99.42	7.87	2.40	2.90	9.23	12.87	0.00	82.91	84.73	97.10	5.40	0.90	2.90	5.68	8.35	0.00	88.91	90.75	97.10	2.22	0.18	0.58	2.66	3.04	0.00	95.12	96.78	99.42	
	30	6.27	3.37	6.18	5.53	8.69	4.30	88.20	87.94	89.52	3.63	1.62	2.52	3.02	4.77	0.00	93.35	93.61	97.48	1.00	0.30	0.90	0.78	1.50	0.00	98.22	98.20	99.10	6.27	3.37	6.18	5.53	8.69	4.30	88.20	87.94	89.52	3.63	1.62	2.52	3.02	4.77	0.00	93.35	93.61	97.48	1.00	0.30	0.90	0.78	1.50	0.00	98.22	98.20	99.10	
	50	5.59	3.39	6.08	5.35	7.18	4.42	89.05	89.43	89.50	3.05	1.70	3.12	2.81	4.13	1.90	94.15	94.17	94.98	0.62	0.18	0.84	0.74	1.18	0.00	98.64	98.64	99.16	5.59	3.39	6.08	5.35	7.18	4.42	89.05	89.43	89.50	3.05	1.70	3.12	2.81	4.13	1.90	94.15	94.17	94.98	0.62	0.18	0.84	0.74	1.18	0.00	98.64	98.64	99.16	
10	7.87	2.60	3.12	8.89	12.64	0.00	83.23	84.76	96.88	4.97	1.12	3.12	5.99	8.35	0.00	89.04	90.53	96.88	2.04	0.18	0.70	2.40	2.98	0.00	95.55	96.84	99.30	7.87	2.60	3.12	8.89	12.64	0.00	83.23	84.76	96.88	4.97	1.12	3.12	5.99	8.35	0.00	89.04	90.53	96.88	2.04	0.18	0.70	2.40	2.98	0.00	95.55	96.84	99.30		
30	5.73	3.62	6.14	6.31	7.98	4.64	87.96	88.40	89.22	2.96	1.57	2.50	3.68	4.40	0.00	93.37	94.03	97.50	1.03	0.24	0.98	1.07	1.53	0.00	97.91	98.23	99.02	5.73	3.62	6.14	6.31	7.98	4.64	87.96	88.40	89.22	2.96	1.57	2.50	3.68	4.40	0.00	93.37	94.03	97.50	1.03	0.24	0.98	1.07	1.53	0.00	97.91	98.23	99.02		
50	6.06	3.76	5.96	5.50	6.36	5.30	88.45	89.88	88.74	3.32	1.81	3.08	2.98	3.22	2.06	93.70	94.97	94.86	1.03	0.28	0.70	0.89	0.79	0.00	98.09	98.93	99.30	6.06	3.76	5.96	5.50	6.36	5.30	88.45	89.88	88.74	3.32	1.81	3.08	2.98	3.22	2.06	93.70	94.97	94.86	1.03	0.28	0.70	0.89	0.79	0.00	98.09	98.93	99.30		
10	9.06	2.66	4.02	9.24	13.06	0.00	81.70	84.28	95.98	5.82	1.20	0.86	5.84	8.68	0.00	88.34	90.12	99.14	2.58	0.20	0.62	2.62	3.58	0.00	94.80	96.22	99.94	9.06	2.66	4.02	9.24	13.06	0.00	81.70	84.28	95.98	5.82	1.20	0.86	5.84	8.68	0.00	88.34	90.12	99.14	2.58	0.20	0.62	2.62	3.58	0.00	94.80	96.22	99.94		
30	6.56	3.20	3.56	5.84	8.88	3.22	87.60	87.92	93.22	3.24	1.38	1.34	3.34	4.92	0.94	93.42	93.70	97.72	0.96	0.38	0.56	1.16	1.34	0.00	97.88	98.28	99.44	6.56	3.20	3.56	5.84	8.88	3.22	87.60	87.92	93.22	3.24	1.38	1.34	3.34	4.92	0.94	93.42	93.70	97.72	0.96	0.38	0.56	1.16	1.34	0.00	97.88	98.28	99.44		
50	4.84	3.72	4.92	5.26	7.54	4.46	89.90	88.74	90.62	2.70	1.84	2.44	3.20	4.20	2.10	94.10	93.96	95.46	0.58	0.28	0.56	0.60	1.10	0.28	98.82	98.62	99.16	4.84	3.72	4.92	5.26	7.54	4.46	89.90	88.74	90.62	2.70	1.84	2.44	3.20	4.20	2.10	94.10	93.96	95.46	0.58	0.28	0.56	0.60	1.10	0.28	98.82	98.62	99.16		
10	8.81	2.46	4.82	9.82	13.14	0.00	81.37	84.40	95.18	6.15	1.26	1.00	6.73	8.49	0.00	87.12	90.25	99.00	2.74	0.20	0.14	3.12	3.50	0.00	94.13	96.29	99.86	8.81	2.46	4.82	9.82	13.14	0.00	81.37	84.40	95.18	6.15	1.26	1.00	6.73	8.49	0.00	87.12	90.25	99.00	2.74	0.20	0.14	3.12	3.50	0.00	94.13	96.29	99.86		
30	5.67	3.14	4.58	6.55	9.13	2.78	87.78	87.72	92.64	2.94</																																														

Tabela 7 – Cobertura de probabilidade para os EMV dos parâmetros  $\log(\eta)$ ,  $\log(\phi)$  e  $\log(\pi/(1-\pi))$  de acordo a  $n = 10, 30, 50$  e  $100 \times [1 - \alpha]\%$  para a distribuição zero ajustada log-exponencial-potência com  $\nu = -0, 5$ .

$\pi$	$\phi$	$n$	90%												95%												99%																																
			CI(%)				CS(%)				Cobertura(%)				CI(%)				CS(%)				Cobertura(%)				CI(%)				CS(%)				Cobertura(%)																								
			$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$	$\eta$	$\phi$	$\pi$																					
10	10,18	0,56	7,00	9,84	23,32	0,00	79,98	76,12	93,00	7,26	0,08	1,08	7,00	17,70	0,00	85,74	82,22	98,92	3,48	0,00	1,08	3,58	9,64	0,00	92,94	90,36	98,92	10	10,18	0,56	7,00	9,84	23,32	0,00	79,98	76,12	93,00	7,26	0,08	1,08	7,00	17,70	0,00	85,74	82,22	98,92	3,48	0,00	1,08	3,58	9,64	0,00	92,94	90,36	98,92				
1	30	6,70	1,36	6,90	6,38	13,76	0,00	86,92	84,88	93,10	3,98	0,48	2,36	3,98	8,88	0,00	92,04	90,64	97,64	1,44	0,04	0,56	1,44	3,54	0,00	97,12	96,42	99,44	1	30	6,70	1,36	6,90	6,38	13,76	0,00	86,92	84,88	93,10	3,98	0,48	2,36	3,98	8,88	0,00	92,04	90,64	97,64	1,44	0,04	0,56	1,44	3,54	0,00	97,12	96,42	99,44		
50	6,18	1,98	5,56	6,22	11,10	2,96	87,60	86,92	91,48	3,36	0,76	2,22	3,50	7,02	0,00	93,14	92,22	97,78	1,02	0,16	0,84	1,00	2,08	0,00	97,98	97,76	99,16	50	6,18	1,98	5,56	6,22	11,10	2,96	87,60	86,92	91,48	3,36	0,76	2,22	3,50	7,02	0,00	93,14	92,22	97,78	1,02	0,16	0,84	1,00	2,08	0,00	97,98	97,76	99,16				
10	9,54	0,66	6,90	9,66	22,96	0,00	80,80	76,38	93,10	6,40	0,10	1,36	6,60	17,04	0,00	87,00	82,86	98,64	3,42	0,00	1,36	3,00	8,90	0,00	93,58	91,10	98,64	10	9,54	0,66	6,90	9,66	22,96	0,00	80,80	76,38	93,10	6,40	0,10	1,36	6,60	17,04	0,00	87,00	82,86	98,64	3,42	0,00	1,36	3,00	8,90	0,00	93,58	91,10	98,64				
0,1	2	30	6,42	2,16	6,68	7,14	13,44	0,00	86,44	84,40	93,32	3,26	0,72	2,18	4,16	8,74	0,00	92,58	90,54	97,82	0,92	0,06	0,68	1,30	3,50	0,00	97,78	96,44	99,32	0,1	2	30	6,42	2,16	6,68	7,14	13,44	0,00	86,44	84,40	93,32	3,26	0,72	2,18	4,16	8,74	0,00	92,58	90,54	97,82	0,92	0,06	0,68	1,30	3,50	0,00	97,78	96,44	99,32
50	6,24	2,42	5,50	5,50	10,64	2,82	88,26	86,94	91,68	3,56	0,90	2,34	3,14	6,62	0,00	93,30	92,48	97,66	0,96	0,00	0,78	0,96	2,24	0,00	98,08	97,76	99,22	50	6,24	2,42	5,50	5,50	10,64	2,82	88,26	86,94	91,68	3,56	0,90	2,34	3,14	6,62	0,00	93,30	92,48	97,66	0,96	0,00	0,78	0,96	2,24	0,00	98,08	97,76	99,22				
10	9,13	0,40	7,15	10,19	23,03	0,00	80,67	76,57	92,85	6,67	0,00	1,46	7,01	17,20	0,00	86,32	82,80	98,54	3,44	0,00	1,46	3,34	8,95	0,00	93,21	91,05	98,54	10	9,13	0,40	7,15	10,19	23,03	0,00	80,67	76,57	92,85	6,67	0,00	1,46	7,01	17,20	0,00	86,32	82,80	98,54	3,44	0,00	1,46	3,34	8,95	0,00	93,21	91,05	98,54				
3	30	6,62	1,44	7,36	6,40	13,46	0,00	86,98	85,10	92,64	3,52	0,64	2,46	3,68	8,46	0,00	92,80	90,90	97,54	1,08	0,00	0,76	1,08	3,24	0,00	97,84	96,76	99,24	3	30	6,62	1,44	7,36	6,40	13,46	0,00	86,98	85,10	92,64	3,52	0,64	2,46	3,68	8,46	0,00	92,80	90,90	97,54	1,08	0,00	0,76	1,08	3,24	0,00	97,84	96,76	99,24		
50	5,96	1,88	6,04	6,02	9,78	3,36	88,02	88,34	90,60	3,20	0,72	2,40	3,16	6,40	0,00	93,64	92,88	97,60	0,78	0,00	0,84	0,90	2,20	0,00	98,32	97,72	99,16	50	5,96	1,88	6,04	6,02	9,78	3,36	88,02	88,34	90,60	3,20	0,72	2,40	3,16	6,40	0,00	93,64	92,88	97,60	0,78	0,00	0,84	0,90	2,20	0,00	98,32	97,72	99,16				
10	10,28	0,50	3,62	10,26	24,90	0,00	79,46	74,60	96,38	7,46	0,10	3,62	7,40	19,46	0,00	85,14	80,44	96,38	4,24	0,00	0,78	3,98	10,98	0,00	91,78	89,02	99,22	10	10,28	0,50	3,62	10,26	24,90	0,00	79,46	74,60	96,38	7,46	0,10	3,62	7,40	19,46	0,00	85,14	80,44	96,38	4,24	0,00	0,78	3,98	10,98	0,00	91,78	89,02	99,22				
1	30	6,74	1,72	6,06	6,26	13,68	4,48	87,00	84,60	89,46	4,08	0,58	2,64	3,44	9,14	0,00	92,48	90,28	97,36	1,48	0,02	1,00	0,96	3,78	0,00	97,56	96,20	99,00	1	30	6,74	1,72	6,06	6,26	13,68	4,48	87,00	84,60	89,46	4,08	0,58	2,64	3,44	9,14	0,00	92,48	90,28	97,36	1,48	0,02	1,00	0,96	3,78	0,00	97,56	96,20	99,00		
50	6,20	2,22	6,10	7,02	10,22	4,76	86,78	87,56	89,14	3,22	0,88	2,76	4,28	6,60	1,84	92,50	92,52	95,40	0,88	0,04	0,56	1,20	2,56	0,00	97,92	97,40	99,44	50	6,20	2,22	6,10	7,02	10,22	4,76	86,78	87,56	89,14	3,22	0,88	2,76	4,28	6,60	1,84	92,50	92,52	95,40	0,88	0,04	0,56	1,20	2,56	0,00	97,92	97,40	99,44				
10	10,34	0,60	3,06	9,90	25,48	0,00	79,76	73,92	96,94	7,24	0,12	3,06	6,90	18,88	0,00	85,86	81,00	96,94	4,02	0,00	0,64	3,46	10,60	0,00	92,52	89,40	99,36	10	10,34	0,60	3,06	9,90	25,48	0,00	79,76	73,92	96,94	7,24	0,12	3,06	6,90	18,88	0,00	85,86	81,00	96,94	4,02	0,00	0,64	3,46	10,60	0,00	92,52	89,40	99,36				
0,2	2	30	7,54	1,42	5,74	6,18	13,78	4,80	86,28	84,80	89,46	4,32	0,44	2,32	3,82	9,14	0,00	91,86	90,42	97,68	1,36	0,02	0,92	1,44	4,16	0,00	97,20	95,82	99,08	0,2	2	30	7,54	1,42	5,74	6,18	13,78	4,80	86,28	84,80	89,46	4,32	0,44	2,32	3,82	9,14	0,00	91,86	90,42	97,68	1,36	0,02	0,92	1,44	4,16	0,00	97,20	95,82	99,08
50	5,78	2,02	6,00	5,38	11,16	4,52	88,84	86,82	89,48	3,50	0,80	2,78	3,10	7,22	1,92	93,40	91,96	95,30	1,10	0,06	0,50	0,68	2,66	0,00	98,22	97,28	99,50	50	5,78	2,02	6,00	5,38	11,16	4,52	88,84	86,82	89,48	3,50	0,80	2,78	3,10	7,22	1,92	93,40	91,96	95,30	1,10	0,06	0,50	0,68	2,66	0,00	98,22	97,28	99,50				
10	9,92	0,40	2,94	9,58	24,16	0,00	80,50	75,44	97,06	6,98	0,02	2,94	6,72	18,02	0,00	86,30	81,96	97,06	3,54	0,00	0,66	3,50	10,04	0,00	92,96	89,96	99,34	10	9,92	0,40	2,94	9,58	24,16	0,00	80,50	75,44	97,06	6,98	0,02	2,94	6,72	18,02	0,00	86,30	81,96	97,06	3,54	0,00	0,66	3,50	10,04	0,00	92,96	89,96	99,34				
3	30	6,86	1,30	6,42	6,50	13,80	4,32	86,64	84,90	89,26	4,02	0,52	2,56	3,90	8,98	0,00	92,08	90,50	97,44	1,38	0,02	0,94	1,44	3,44	0,00	97,18	96,54	99,06	3	30	6,86	1,30	6,42	6,50	13,80	4,32	86,64	84,90	89,26	4,02	0,52	2,56	3,90	8,98	0,00	92,08	90,50	97,44	1,38	0,02	0,94	1,44	3,44	0,00	97,18	96,54	99,06		
50	6,14	2,18	5,76	5,60	11,14	4,94	88,26	86,68	89,30	3,08	0,86	3,16	3,02	7,02	1,92	93,90	92,12	94,92	0,88	0,08	0,62	1,00	2,68	0,00	98,12	97,24	99,38	50	6,14	2,18	5,76	5,60	11,14	4,94	88,26	86,68	89,30	3,08	0,86	3,16	3,02	7,02	1,92	93,90	92,12	94,92	0,88	0,08	0,62	1,00	2,68	0,00	98,12	97,24	99,38				
10	10,74	0,38	4,96	11,88	27,80	0,00	77,38	71,82	95,04	7,84	0,04	0,96	9,54	21,36	0,00	82,62	78,60	99,04	4,16	0,00	0,12	5,54	12,80	0,00	90,30	87,20	99,88	10	10,74	0,38	4,96	11,88	27,80	0,00	77,38	71,82	95,04	7,84	0,04	0,96	9,54	21,36	0,00	82,62	78,60	99,04	4,16	0,00	0,12	5,54	12,80	0,00	90,30	87,20	99,88				
1	30	7,12	1,30	4,18	7,02	13,92	3,36	85,86	84,78	92,46	4,34	0,46	1,52	4,38	9,24	0,90	91,28	90,30	97,58	1,72	0,04	0,64	1,66	4,04	0,00	96,62	95,92	99,36	1	30	7,12	1,30	4,18	7,02	13,92	3,36	85,86	84,78	92,46	4,34	0,46	1,52	4,38	9,24	0,90	91,28	90,30	97,58	1,72	0,04	0,64	1,66	4,04	0,00	96,62	95,92	99,36		
50	6,24	1,70	4,76	6,12	12,36	4,28	87,64	85,94	90,96	3,30	0,58	2,60	3,24	7,26	2,08	93,46	92,16	95,32	1,00	0,06	0,56	1,00	2,44	0,00	98,00	97,50	99,10	50	6,24	1,70	4,76	6,12	12,36	4,28	87,64	85,94	90,96	3,30	0,58	2,60	3,24	7,26	2,08	93,46	92,16	95,32	1,00	0,06	0,56	1,00	2,44	0,00	98,00	97,50	99,10				
10	11,06	0,32	4,70	11,36	27,20	0,00	77,58	72,48	95,30	8,44	0,02	1,12	8,48	21,20	0,00	83,08	78,78	98,88	4,98	0,00	0,08	4,78	12,90	0,00	90,24	87,10	99,92	10	11,06	0,32	4,70	11,36	27,20	0,00	77,58	72,48	95,30	8,44																					

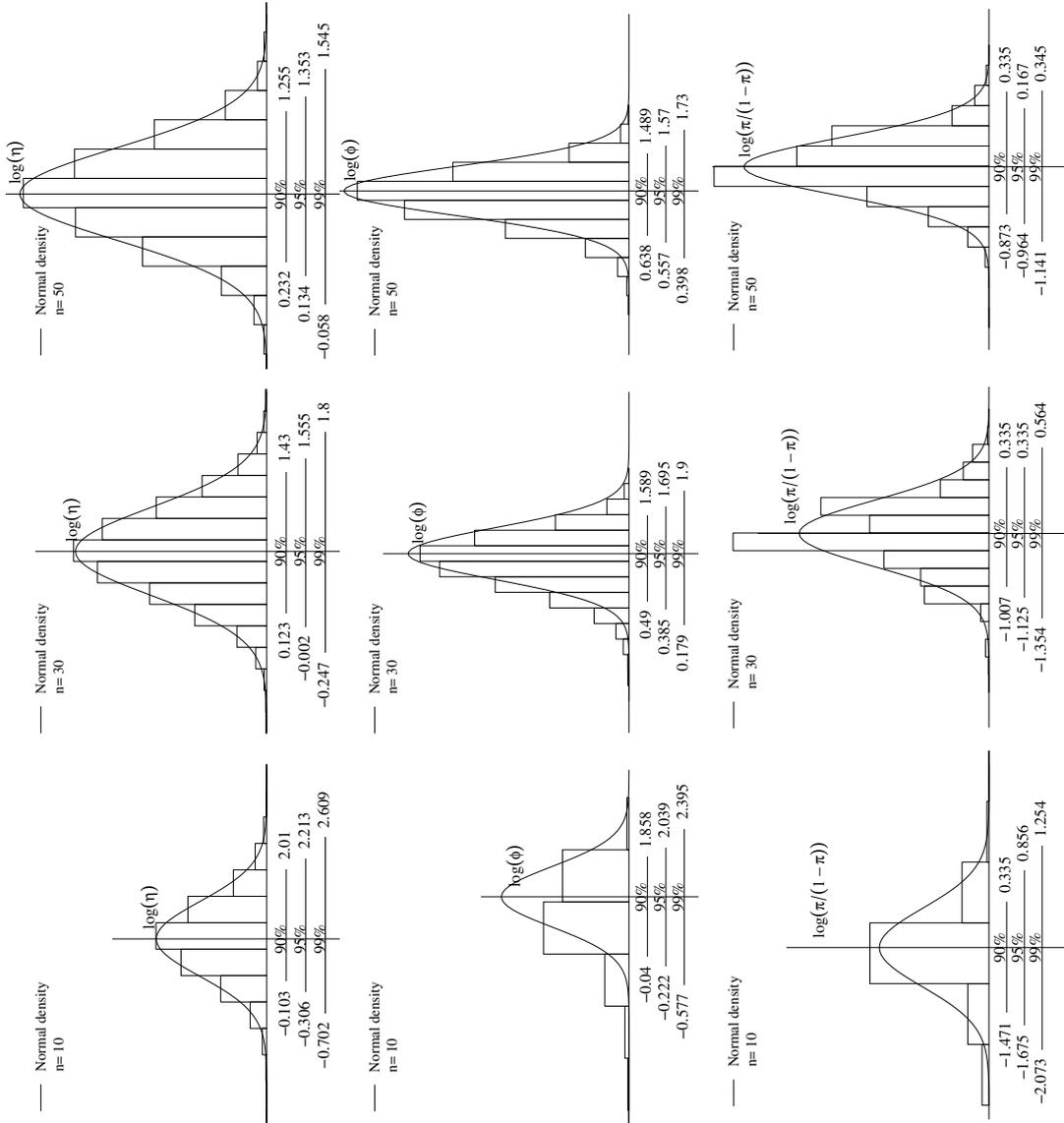


Figura 5 – Histograma para os EMV dos parâmetros  $\log(\eta)$ ,  $\log(\phi)$ ,  $\logit(\pi)$  com o respectivo intervalo de confiança médio para valores de  $n$  e níveis  $100X[1 - \alpha]\%$  de confiança para a distribuição ZALN com  $\theta = (2, 4, 0, 4)^T$ .

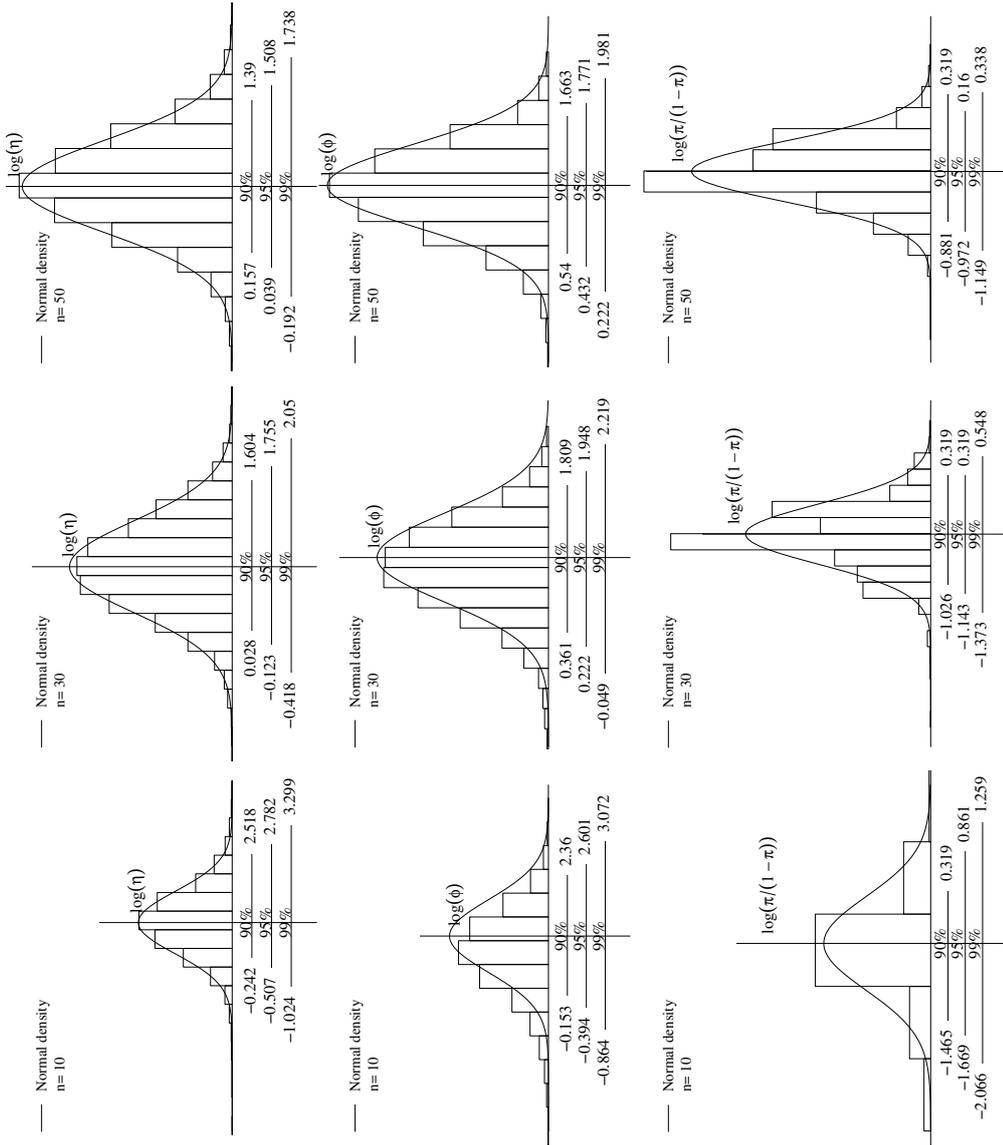


Figura 6 – Histograma para os EMV dos parâmetros  $\log(\eta)$ ,  $\log(\phi)$ ,  $\text{logit}(\pi)$  com o respectivo intervalo de confiança médio para valores de  $n$  e níveis  $100X[1 - \alpha]\%$  de confiança para a distribuição ZALSt com  $\theta = (2, 4, 0, 4)^\top$  e  $\nu = 4$ .

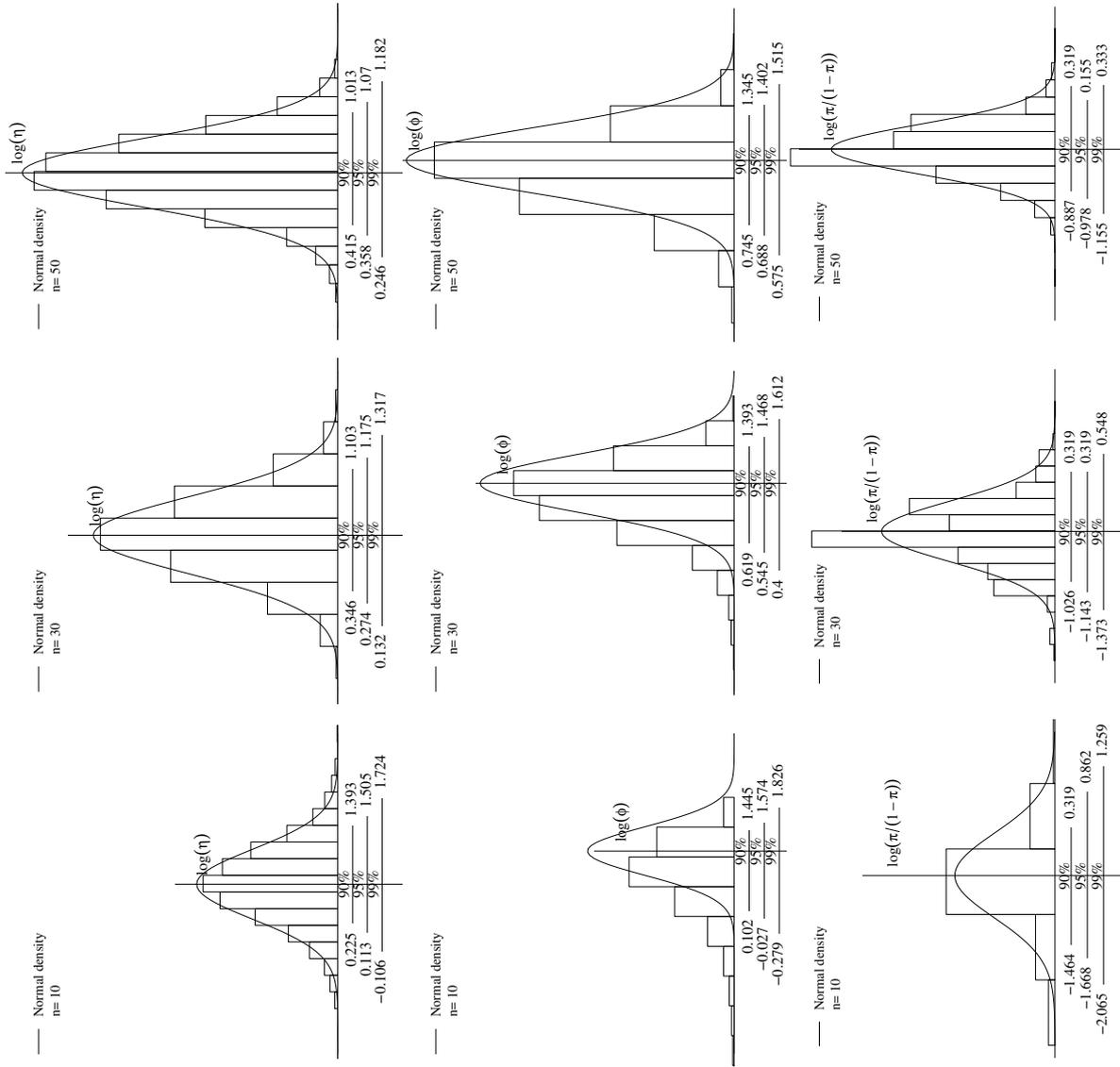


Figura 7 – Histograma para os EMV dos parâmetros  $\log(\eta)$ ,  $\log(\phi)$ ,  $\log(\pi/(1-\pi))$  com o respectivo intervalo de confiança médio para valores de  $n$  e níveis  $100X[1 - \alpha]\%$  de confiança para a distribuição ZALPE com  $\theta = (2, 4, 0, 4)^T$  e  $\nu = -0, 5$ .

## 2.3 APLICAÇÃO

Nesta seção, ajustamos as distribuições propostas e comparamos com algumas distribuições já estabelecidas na literatura em um conjunto de dados reais. Assim, consideramos o conjunto de dados sobre despesas de vereadores eleitos, incluindo o prefeito da cidade de Leicester City, no Reino Unido, do período 2012/2013. O conjunto de dados é extraído do *Leicester City Council*, disponível em <https://data.leicester.gov.uk/pages/home/>. Neste caso, estamos interessados na variável *subsídio de responsabilidade especial* (£), que consiste nos subsídios adicionais para responsabilidades específicas. A amostra contém 18 valores de zeros, que representa 32,73% das observações. Para os valores positivos foi feito um análise exploratória dos dados, em que foi observado um mínimo de 221,93, o primeiro quantil de 3879,96, o valor mediano de 6759,58, a média de 9009,571, o terceiro quantil de 9259,03 e o valor máximo de 55908,96. Para algumas medidas de variabilidade, um desvio padrão de 10553,33, a amplitude interquartil de 5379,07. O coeficiente de assimetria e curtose de 3,097 e 13,207, respectivamente. Também, a Figura 8 apresenta o respectivo histograma considerando os valores zeros na linha vertical, e um boxplot ajustado considerando só os valores positivos. Assim, podemos observar que os dados apresentam uma assimetria positiva e a presença de pontos atípicos.

Baseado na análise exploratória dos dados ajustamos três distribuições da classe ZALS aos dados de despesas a saber: zero ajustado log-normal (ZALN), zero ajustado t-Student (ZALSt) e zero ajustado exponencial-potência (ZALPE). Também ajustamos três distribuições encontradas na literatura: zero ajustado gaussiana inversa (ZAIG), zero ajustado gama (ZAGA) e zero ajustado Birnbaum-Saunders reparametrizado (ZARBS). O parâmetro extra para os casos das distribuições ZALSt e ZALPE foi considerado fixo e escolhido minimizando o AIC para um vetor de valores de  $\nu = [4; 10]$  para ZALSt e  $\nu = (\pm 0, 3; \pm 0, 2; \pm 0, 1)$  para ZALPE. Assim, baseado nessa metodologia, o parâmetro extra escolhido foi de  $\nu = 4$  e  $\nu = 0, 3$  para o caso de ZALSt e ZALPE, respectivamente.

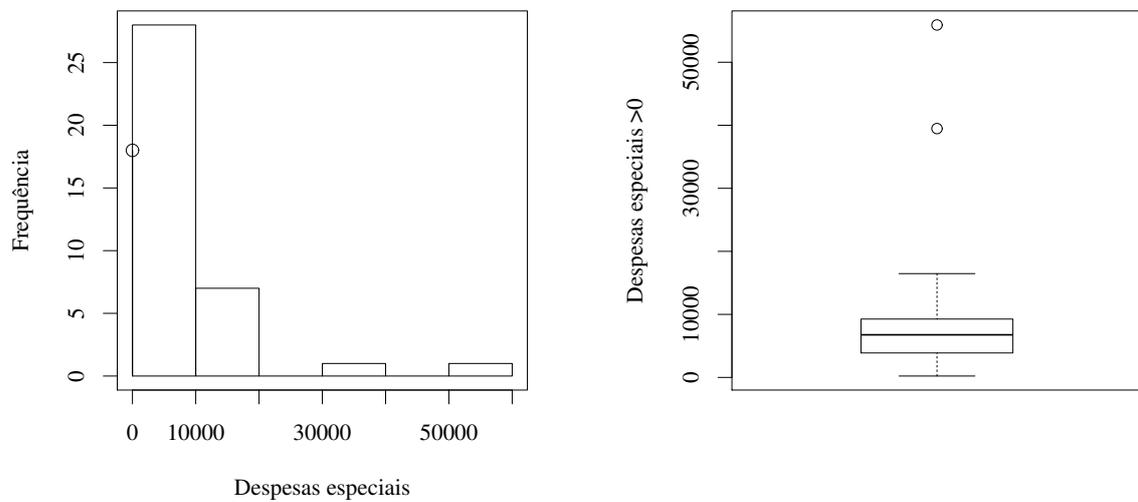


Figura 8 – Histograma e boxplot ajustado para as despesas especiais dos vereadores eleitos no período 2012/2013 na cidade de Leicester.

A Tabela 8 apresenta as estimativas de máxima verossimilhança dos parâmetros, erros padrão assintóticos (entre parênteses), o critério de informação de Akaike ( $AIC = 2p - 2\log(\hat{L})$ ) e o Critério de Informação Bayesiano ( $BIC = p\log(n) - 2\log(\hat{L})$ ), em que  $n$  é o tamanho amostral,  $p$ , o número de parâmetros e  $\hat{L} = L(\hat{\theta})$  é a verossimilhança avaliada nos parâmetros estimados para as distribuições ajustadas. Adicionalmente, o teste de Kolmogorov–Smirnov (KS) foi obtido para avaliar a bondade do ajuste. Assim, podemos observar que a distribuição ZALSt apresenta o menor AIC e BIC entre as distribuições ajustadas. Além disso, o teste KS indica para as três distribuições pertencentes ao classe ZALS e a distribuição ZAGA que não se tem evidência suficiente para rejeitar que os dados seguem a distribuição em questão com um nível de 5% de significância.

Tabela 8 – Estimativa de máxima verossimilhança para os parâmetros e estatísticas para as distribuições consideradas ajustadas ao conjunto de dados de despesas especiais dos vereadores eleitos no período 2012/2013 na cidade de Leicester.

	ZALSt	ZALPE	ZALN	ZAGA	ZARBS	ZAIG
$\eta$	5989,105 (806,091)	6023,433 (896,832)	5841,316 (929,127)	– –	– –	– –
$\mu$	– –	– –	– –	9009,571 (1294,390)	8767,315 (1694,320)	9009,571 (2097,152)
$\phi$	0,475 (0,146)	0,508 (0,135)	0,938 (0,218)	– –	– –	– –
$\sigma$	– –	– –	– –	0,879 (0,092)	1,387 (0,325)	0,015 (0,002)
$-2\log(\hat{L})$	<b>809,13</b>	811,23	813,94	815,97	822,20	825,15
AIC	<b>815,13</b>	817,23	819,94	821,97	828,20	831,15
BIC	<b>821,15</b>	823,25	825,97	827,99	834,22	837,18
KS	0,151	0,132	0,151	0,150	0,230	0,249
$p$ -valor	0,369	0,541	0,369	0,380	0,040	0,021

Por último, a Figura 9 apresenta as densidades ajustadas aos dados e gráficos Q-Q para as diferentes distribuições utilizadas para a presente análise. Como podemos ver, as densidades ajustadas das distribuições da classe ZALS se ajustam melhor ao histograma dos dados que em comparação com as distribuições ZAGA, ZAIG e ZARBS.

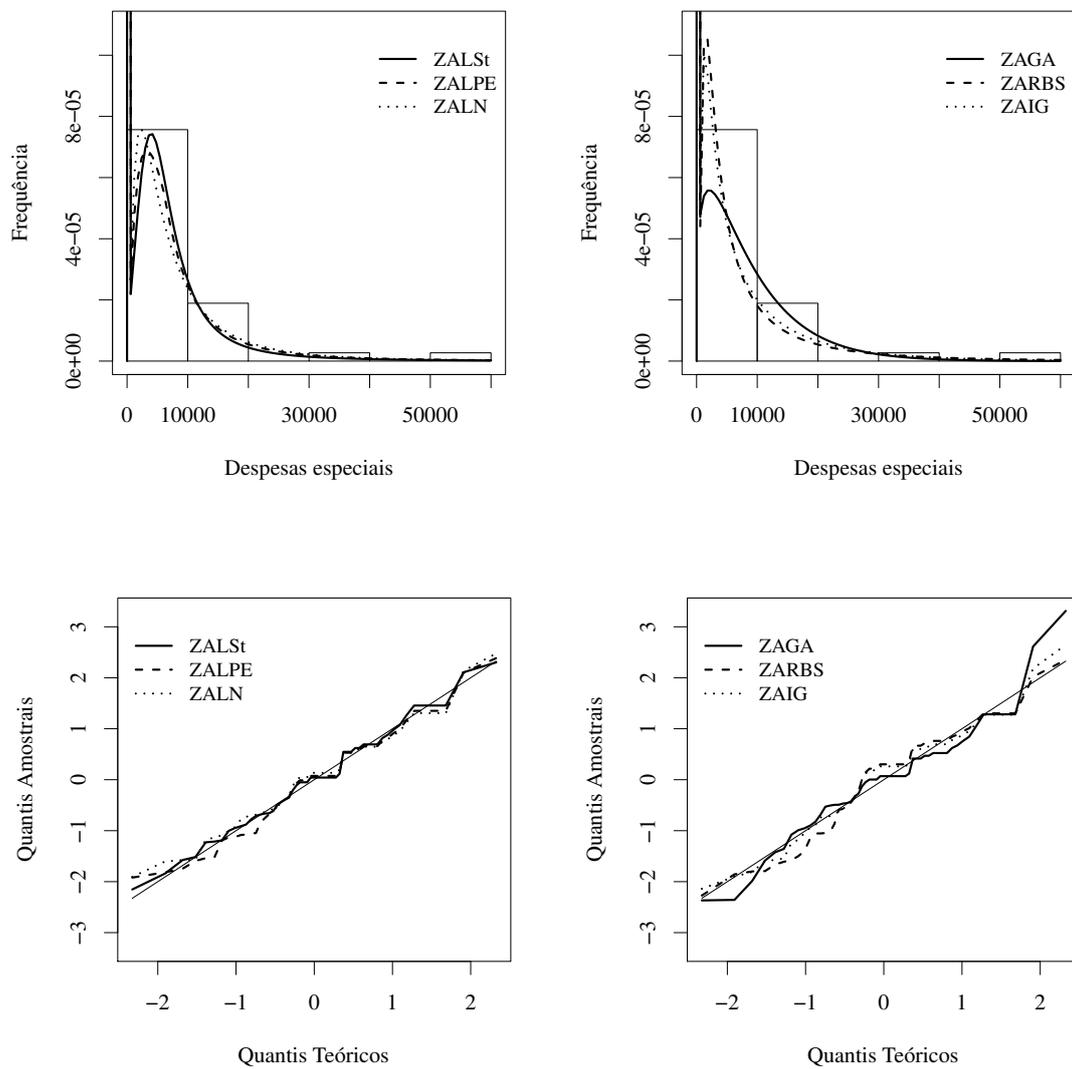


Figura 9 – Histograma e gráfico Q-Q para as distribuições ajustadas para o conjunto de dados de despesas especiais dos vereadores eleitos no período 2012/2013 na cidade de Leicester.

### 3 MODELOS DE REGRESSÃO LINEARES ZERO AJUSTADO LOG-SIMÉTRICAS

Os modelos de regressão são comumente usados para explicar o comportamento de uma variável resposta através de um conjunto de variáveis explanatórias, sendo de muito interesse por exemplo nas áreas de saúde, economia, engenharia e química. É muito comum assumir normalidade dos erros (Wei e Tanner, 1990), mas pode acontecer em que certas situações assumir normalidade não seja o mais adequado. Por exemplo, é bem conhecido que os estimadores de máxima verossimilhança do modelo de regressão normal são altamente sensíveis a observações aberrantes. Neste contexto, os modelos simétricos de regressão tem sido estudado e recebido uma crescente atenção (veja por exemplo, Fang, Kotz e Ng , 1990; Kai-Tai e Yao-Ting, 1990; Anderson e Fang, 1990; Cordeiro et al., 2000; Cysneiros et al., 2007; Maior e Cysneiros, 2018). Contudo, em situações em que a variável resposta é positiva e tem comportamento assimétrico, os modelos simétricos não são adequados. Uma possível solução é transformar a variável resposta de tal forma que a nova variável assuma valores na reta real, por exemplo, a transformação logarítmica e Box Cox (Box e Cox, 1964) foram provavelmente as mais conhecidas e sugeridas a fim de alcançar uma normalidade nos valores observados. No entanto, esta metodologia tem limitações no contexto que a parte inferencial do modelo já não é mais feito em relação à variável resposta original ocasionando a perda de interpretabilidade dos parâmetros. Uma solução viável é utilizar distribuições contínuas positivas com suporte  $(0, \infty)$  por exemplo, a distribuição gaussiana inversa, gama e gama generalizada.

As distribuições mencionadas são umas das mais atrativas e citadas na literatura, porém, os estimadores são altamente influenciados na presença de observações atípicas. Assim, Vanegas e Paula (2015) propuseram os modelos log-simétricos de regressão em que a variável resposta é estritamente positiva e assimétrica. Este tipo de modelos é interessante já que permite que a mediana e a assimetria da distribuição da variável resposta sejam modeladas por variáveis explanatórias sendo uma opção mais natural na presença de observações atípicas. Algumas das distribuições log-simétricas contém um parâmetro extra, permitindo que as distribuições log-simétricas sejam ainda mais flexíveis.

Por outro lado, haverá ocasiões em que se procura estudar ou fazer previsões de uma determinada variável do qual os valores são contínuos e não negativos. Este tipo de variáveis são comumente chamadas de semi-contínuas (Min e Agresti, 2002) com distribuição contínua exceto no ponto zero em que atribuímos uma probabilidade de ocorrência. Por exemplo, em um

estudo sobre níveis de dano corporal de pacientes com alguma doença, a medição pode assumir o valor de zero se um paciente não apresentar dano ou pode assumir valores positivos para indicar o nível de dano (Su et al., 2009), Transande e Chatterjee (2009) fizeram um estudo avaliando o impacto nas despesas de saúde em crianças de 6 a 19 anos, encontrando zeros na variável resposta.

No contexto de distribuições semi-contínuas, Duan et al. (1983) trata o modelo em duas partes como uma mistura entre uma variável com distribuição discreta Bernoulli para os casos que ocorrem zero e outra distribuição contínua, log-normal, para a parte positiva. Heller et al. (2006) apresenta o modelo de regressão zero ajustado gaussiana inversa para modelar dados de demanda de seguros em que a variável resposta apresenta uma assimetria positiva. Rodrigues-Motta et al. (2015) propõem modelos de regressão para distribuições positivas com zeros considerando a família exponencial dupla como a componente contínua. Ainda mais recente, Tomazella et al. (2019) propõem o modelo de regressão zero ajustado reparametrizado Birnbaum-Saunders.

Neste capítulo apresentamos os modelos de regressão lineares zero ajustado log-simétricas para valores positivos contínuos e assimétricos que apresentam zeros na variável resposta. O capítulo se encontra organizado na seguinte forma. Na secção 3.2 é definido o modelo de regressão lineares zero ajustado log-simétrica, abordando o processo de estimação por máxima verossimilhança baseado na função score e a matriz de informação de Fisher para apresentar o algoritmo de estimação. Estimação intervalar e método de seleção de variáveis também são apresentados. Resíduos e métodos de diagnósticos são desenvolvidos para avaliar o modelo proposto. Por último, um conjunto de dados reais é apresentado para aplicar a teoria desenvolvida.

### 3.1 MODELO DE REGRESSÃO

Seja  $w_1, \dots, w_n$   $n$  realizações independentes das variáveis aleatórias  $W_i$  com função de densidade dada na Equação (2.2), i.e.,  $W_i \sim ZALS(\eta_i, \phi_i, \pi_i, g(\cdot))$  para  $i = 1, \dots, n$ . A relação dos parâmetros de interesse  $(\eta_i, \phi_i, \pi_i)^\top$  é definida como

$$\begin{aligned}
m_1(\eta_i) &= \xi_{1i} = \sum_{j=1}^{M_1} x_{ij} \beta_j = \mathbf{X}_i^\top \boldsymbol{\beta}, \\
m_2(\phi_i) &= \xi_{2i} = \sum_{j=1}^{M_2} z_{ij} \alpha_j = \mathbf{Z}_i^\top \boldsymbol{\alpha}, \\
m_3(\pi_i) &= \xi_{3i} = \sum_{j=1}^{M_3} d_{ij} \gamma_j = \mathbf{D}_i^\top \boldsymbol{\gamma},
\end{aligned} \tag{3.1}$$

em que  $x_{i1}, \dots, x_{iM_1}$ ,  $z_{i1}, \dots, z_{iM_2}$  e  $d_{i1}, \dots, d_{iM_3}$  representa os valores de variáveis explicativas relacionados com  $\eta_i$ ,  $\phi_i$  e  $\pi_i$ , respectivamente;  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{M_1})^\top$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{M_2})^\top$  e  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{M_3})^\top$  são vetores de parâmetros de regressão desconhecidos a serem estimados tais que  $\boldsymbol{\beta} \in \mathbb{R}^{M_1}$ ,  $\boldsymbol{\alpha} \in \mathbb{R}^{M_2}$  e  $\boldsymbol{\gamma} \in \mathbb{R}^{M_3}$ . As funções de ligações  $m_k(\cdot)$  para  $k = 1, 2, 3$ , são estritamente monótonas e pelo menos duas vezes diferenciáveis em que  $m_1 : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $m_2 : \mathbb{R}_+ \rightarrow \mathbb{R}$  e  $m_3 : (0, 1) \rightarrow \mathbb{R}$ .  $g(\cdot)$  é a função geradora de densidade que satisfaz as condições que  $g(u) > 0$  e  $\int_0^\infty u^{-\frac{1}{2}} g(u) \partial u = 1$  para  $u > 0$ .

Considerando os parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\gamma}^\top)^\top$ , a função de verosimilhança é dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f_{W_i}(\eta_i, \phi_i, \pi_i, g(\cdot)) = L_1(\boldsymbol{\gamma}) L_2(\boldsymbol{\beta}, \boldsymbol{\alpha}), \tag{3.2}$$

sendo  $L_1(\boldsymbol{\gamma}) = \prod_{i=1}^n \pi_i^{I_{\{0\}}(w_i)} (1 - \pi_i)^{1 - I_{\{0\}}(w_i)}$  e  $L_2(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{i:w_i > 0} f_{T_i}(w_i)^{1 - I_{\{0\}}(w_i)}$  em que os parâmetros  $\eta_i$ ,  $\phi_i$  e  $\pi_i$  são funções dos parâmetros a serem estimados  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$  e  $\boldsymbol{\gamma}$ , respectivamente, através de (3.1). Em alguns casos, a função geradora de densidade,  $g(\cdot)$ , envolve um parâmetro extra,  $\nu$ , que será considerado fixo ou conhecido. Assim, O logaritmo da função verosimilhança é dada por

$$\ell(\boldsymbol{\theta}) = \ell_{(1)}(\boldsymbol{\gamma}) + \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}), \tag{3.3}$$

em que

$$\begin{aligned}
\ell_{(1)}(\boldsymbol{\gamma}) &= \sum_{i=1}^n \ell_i(\pi_i), \\
\ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^n \ell_i(\eta_i, \phi_i),
\end{aligned}$$

e

$$\begin{aligned} \ell_i(\pi_i) &= I_{\{0\}}(w_i)\log(\pi_i) + [1 - I_{\{0\}}(w_i)]\log(1 - \pi_i), \\ \ell_i(\eta_i, \phi_i) &= [1 - I_{\{0\}}(w_i)]\left\{-\frac{1}{2}\log(\phi_i) + \log[g(\tilde{w}_i^2)] - \log(w_i)\right\}. \end{aligned} \quad (3.4)$$

Dado que em (3.2) os vetores de parâmetros  $(\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$  e  $\boldsymbol{\gamma}$  são separáveis (Pace & Salvan, 1997, p. 128) a função de verosimilhança,  $L(\boldsymbol{\theta})$  pode ser fatorada em dois termos, uma que depende do vetor  $\boldsymbol{\gamma}$  associado a componente discreta que é usado para modelar a probabilidade de ocorrência de zero na variável resposta e outra que depende de  $(\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$  associado a componente contínua para modelar a distribuição condicional da variável resposta no suporte  $\mathbb{R}_+$ . Portanto o processo de estimação por máxima verosimilhança pode ser desenvolvido separadamente para  $(\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$  e para o vetor  $\boldsymbol{\gamma}$ .

### 3.1.1 Função escore

As funções escore de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\gamma}^\top)^\top$  são dadas por  $U(\boldsymbol{\theta}) = (U(\boldsymbol{\beta})^\top, U(\boldsymbol{\alpha})^\top, U(\boldsymbol{\gamma})^\top)^\top = (\partial\ell(\boldsymbol{\beta})/\partial\boldsymbol{\beta}, \partial\ell(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}, \partial\ell(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma})^\top$ . Assim,  $U(\boldsymbol{\theta})$  toma a forma

$$U(\boldsymbol{\theta}) = \begin{bmatrix} U_\beta(\boldsymbol{\theta}) \\ U_\alpha(\boldsymbol{\theta}) \\ U_\gamma(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{T} \mathbf{D}_{(v)} \mathbf{Q} (\mathbf{I}_n - \mathbf{W}^c) (\mathbf{w}^+ - \boldsymbol{\eta}^*) \\ \frac{1}{2} \mathbf{Z}^\top \mathbf{T} \boldsymbol{\Delta}_\phi (\mathbf{I}_n - \mathbf{W}^c) (\mathbf{s} - \mathbf{1}_n) \\ \mathbf{D}^\top \mathbf{P} \boldsymbol{\Delta}_\pi (\mathbf{w}^c - \boldsymbol{\pi}^*) \end{bmatrix},$$

em que  $\boldsymbol{\eta}^* = (\eta_1^*, \dots, \eta_n^*)^\top$ ,  $\mathbf{w}^+ = (w_1^+, \dots, w_n^+)^\top$ ,  $\mathbf{w}^c = (w_1^c, \dots, w_n^c)^\top$ ,  $(\mathbf{W}^c) = \text{diag}\{w_1^c, \dots, w_n^c\}$ ,  $\mathbf{Q} = \text{diag}\{q_1, \dots, q_n\}$ ,  $\mathbf{D}_{(v)} = \text{diag}\{v_1, \dots, v_n\}$  e a matriz  $\mathbf{T} = \text{diag}\{1/\phi_1, \dots, 1/\phi_n\}$ ,  $\eta_i^* = \log(\eta_i)$ ,  $w_i^c = I_{\{0\}}(w_i)$ ,  $q_i = (\partial\eta_i/\partial\xi_{1i})/\eta_i$  com  $i = 1, \dots, n$ ,  $v_i = v_+(w_i)$ , em que  $w^+$  e  $v_+(w)$  é uma função de  $w$  da forma

$$w^+ = \begin{cases} \log(w), & \text{se } w > 0, \\ 0, & \text{se } w = 0, \end{cases} \quad v_+(w) = \begin{cases} -2g'(\tilde{w}^2)/g(\tilde{w}^2), & \text{se } w > 0, \\ 0, & \text{se } w = 0, \end{cases}$$

em  $U_\alpha(\boldsymbol{\theta})$  temos que  $\mathbf{s} = (v_1 w_1^{+2}, \dots, v_n w_n^{+2})$  e  $\boldsymbol{\Delta}_\phi = \text{diag}\{\partial\phi_1/\partial\xi_{21}, \dots, \partial\phi_n/\partial\xi_{2n}\}$ . Para a componente discreta temos que  $\boldsymbol{\pi}^* = (\pi_1, \dots, \pi_n)^\top$ ,  $\boldsymbol{\Delta}_\pi = \text{diag}\{\partial\pi_1/\partial\xi_{31}, \dots, \partial\pi_n/\partial\xi_{3n}\}$  e

$$\mathbf{P} = \text{diag}\{1/[\pi_1(1 - \pi_1)], \dots, 1/[\pi_n(1 - \pi_n)]\}.$$

$$\tilde{w}_i^+ = \begin{cases} \tilde{w}_i^2, & \text{se } w_i > 0, \\ 0, & \text{se } w_i = 0, \end{cases}$$

### 3.1.2 Informação de Fisher

Seja  $-\ddot{L}_{\hat{\theta}\hat{\theta}}$  a matriz de informação observada de Fisher para  $\theta$  em que  $\ddot{L}_{\theta\theta} = \partial^2 l(\theta)/\partial\theta\partial\theta^\top$  é a matriz de segundas derivadas em relação aos parâmetros de interesse, assim

$$\ddot{L}_{\theta\theta} = \begin{bmatrix} \ddot{L}_{\theta_1\theta_1} & 0 \\ 0 & \ddot{L}_{\gamma\gamma} \end{bmatrix}, \quad (3.5)$$

em que  $\ddot{L}_{\theta_1\theta_1}$  e  $\ddot{L}_{\gamma\gamma}$  são matrizes de segundas derivadas de  $\theta_1 = (\beta^\top, \alpha^\top)^\top$  e  $\gamma$  respetivamente,

$$\ddot{L}_{\theta_1\theta_1} = \begin{bmatrix} \ddot{L}_{\beta\beta} & \ddot{L}_{\beta\alpha} \\ \ddot{L}_{\alpha\beta} & \ddot{L}_{\alpha\alpha} \end{bmatrix},$$

sendo que

$$\ddot{L}_{\gamma\gamma} = -\mathbf{D}^\top \boldsymbol{\psi}_1 \mathbf{D}, \quad \ddot{L}_{\beta\beta} = -\mathbf{X}^\top \boldsymbol{\psi}_2 \mathbf{X}, \quad \ddot{L}_{\beta\alpha} = \ddot{L}_{\alpha\beta}^\top = -\mathbf{X}^\top \boldsymbol{\psi}_3 \mathbf{Z}, \quad \ddot{L}_{\alpha\alpha} = -\mathbf{Z}^\top \boldsymbol{\psi}_4 \mathbf{Z},$$

em que  $\boldsymbol{\psi}_k$  para  $k = 1, 2, 3, 4$  são matrizes diagonais em que cada elemento  $i$  da sua respectiva diagonal para  $i = 1, \dots, n$ , é dado por

$$\begin{aligned} \boldsymbol{\psi}_{1i} &= \frac{1}{\pi_i(1 - \pi_i)} \left[ (I_{\{0\}}(w_i) - \pi_i) \frac{\partial^2 \pi_i}{\partial \xi_{3i}^2} + \left( \frac{\partial \pi_i}{\partial \xi_{3i}} \right)^2 \left[ \frac{2\pi_i - 1}{\pi_i(1 - \pi_i)} (I_{\{0\}}(w_i) - \pi_i) - 1 \right] \right], \\ \boldsymbol{\psi}_{2i} &= (1 - I_{\{0\}}(w_i)) \left[ \frac{1}{\phi_i} \left( \frac{\partial \eta_i}{\eta_i \partial \xi_{1i}} \right)^2 (\mathbf{v}_i + \mathbf{v}'_i \tilde{w}_i) + \frac{\mathbf{v}_i \tilde{w}_i}{\sqrt{\phi_i}} \left[ \frac{\partial^2 \eta_i}{\eta_i \partial \xi_{1i}^2} - \left( \frac{\partial \eta_i}{\eta_i \partial \xi_{1i}} \right)^2 \right] \right], \\ \boldsymbol{\psi}_{3i} &= (1 - I_{\{0\}}(w_i)) \frac{1}{\phi_i^{3/2}} \left( \frac{\partial \eta_i}{\eta_i \partial \xi_{1i}} \right) \frac{\partial \phi_i}{\partial \xi_{2i}} [\mathbf{v}_i \tilde{w}_i + \mathbf{v}'_i \tilde{w}_i^2 / 2], \\ \boldsymbol{\psi}_{4i} &= (1 - I_{\{0\}}(w_i)) \left[ \left( \frac{\partial \phi_i}{\phi_i \partial \xi_{2i}} \right)^2 [\mathbf{v}_i \tilde{w}_i + \mathbf{v}'_i \tilde{w}_i^2 / 2] \tilde{w}_i / 2 + \frac{(\mathbf{v}_i \tilde{w}_i^2 - 1)}{2\phi_i} \left[ \frac{\partial^2 \phi_i}{\partial \xi_{2i}^2} - \frac{1}{\phi_i} \left( \frac{\partial \phi_i}{\partial \xi_{2i}} \right)^2 \right] \right]. \end{aligned} \quad (3.6)$$

A matriz de informação de Fisher esperada para o vetor de parâmetros  $\theta$ ,  $\mathbf{K}(\theta)$ , tem a forma

$$\mathbf{K}(\theta) = \begin{bmatrix} \mathbf{K}_{\theta_1}(\theta_1) & 0 \\ 0 & \mathbf{K}_\gamma(\gamma) \end{bmatrix}, \quad (3.7)$$

sendo  $\mathbf{K}_{\theta_1}(\theta_1)$  a matriz de informação de Fisher para  $\theta_1 = (\beta^\top, \alpha^\top)^\top$  e  $\mathbf{K}_\gamma(\gamma) = \mathbf{K}_{\gamma\gamma} = \mathbf{D}^\top \mathbf{P} \Delta_\pi^2 \mathbf{D}$  a matriz de informação de Fisher para  $\gamma$ . Temos a matriz de informação de Fisher para  $\theta_1 = (\beta^\top, \alpha^\top)^\top$  é dada por

$$\mathbf{K}_{\theta_1}(\theta_1) = \begin{bmatrix} \mathbf{K}_{\beta\beta} & 0 \\ 0 & \mathbf{K}_{\alpha\alpha} \end{bmatrix}, \quad (3.8)$$

em que  $\mathbf{K}_{\beta\beta} = d_g \mathbf{X}^\top \mathbf{R} \mathbf{T} \mathbf{Q}^2 \mathbf{X}$ ,  $\mathbf{K}_{\alpha\alpha} = \frac{(f_g-1)}{4} \mathbf{Z}^\top \mathbf{R} \mathbf{T}^2 \Delta_\phi^2 \mathbf{Z}$  e  $\mathbf{R} = \text{diag}\{r_1, \dots, r_n\}$  com  $r_i = 1 - \pi_i$  para  $i = 1, \dots, n$ . Valores para  $d_g$  e  $f_g$  podem ser encontrados na Tabela 1. Em (3.7) e (3.8), os parâmetros  $\beta$ ,  $\alpha$  e  $\gamma$  são ortogonais, assim, deste fato temos a matriz de informação de Fisher para  $\theta$  é bloco-diagonal,  $\mathbf{K}(\theta) = \text{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\alpha\alpha}, \mathbf{K}_{\gamma\gamma}\}$ . Considere como casos particulares as funções de ligação  $m_1(\eta_i) = \log(\eta_i)$ ,  $m_2(\phi_i) = \log(\phi_i)$  e  $m_3(\pi_i) = \text{logit}(\pi_i)$  em que  $\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ ; temos a função escore de  $\theta$  em que  $U_\beta(\theta) = \mathbf{X}^\top \mathbf{T} \mathbf{D}_{(v)} (\mathbf{I}_n - \mathbf{W}^c) (\mathbf{w}^+ - \boldsymbol{\eta}^*)$ ,  $U_\alpha(\theta) = \frac{1}{2} \mathbf{Z}^\top (\mathbf{I}_n - \mathbf{W}^c) (\mathbf{s} - \mathbf{1}_n)$  e  $U_\gamma(\theta) = \mathbf{D}^\top (\mathbf{w}^c - \boldsymbol{\pi}^*)$ ; a matriz de informação de Fisher para  $\gamma$ ,  $\beta$  e  $\alpha$  tem a forma  $\mathbf{K}_\gamma(\gamma) = \mathbf{D}^\top \mathbf{P}^{-1} \mathbf{D}$ ,  $\mathbf{K}_{\beta\beta} = d_g \mathbf{X}^\top \mathbf{R} \mathbf{T} \mathbf{X}$  e  $\mathbf{K}_{\alpha\alpha} = \frac{(f_g-1)}{4} \mathbf{Z}^\top \mathbf{R} \mathbf{Z}$ , respectivamente.

### 3.1.3 Processo de estimação

A estimação de máxima verossimilhança de  $\theta$  é feita igualando a função escore a zero e resolvendo o sistema de equações não lineares,  $U(\theta) = \mathbf{0}$ . Neste caso o sistema de equações não lineares não apresenta solução analítica fechada sendo necessário algum algoritmo de otimização não linear para maximizar a função de log-verossimilhança. Um ponto importante, é que pela separabilidade dos parâmetros, o estimador de máxima verossimilhança de  $\gamma$  pode ser obtido independentemente de  $\theta_1$ . Assim, é apresentado dois algoritmos: um para a componente discreta e outro para a componente contínua.

**Algoritmo 2 - Componente discreta:**

Passo 1: Indicar o chute inicial para  $\gamma^{(0)}$

Passo 2: Iniciar o contador em  $l = 0$  e  $\gamma^{(l)} = \gamma^{(0)}$

Passo 3: Basado em  $\gamma^{(l)}$  fazer:

$$\gamma^{(l+1)} = \gamma^{(l)} + \left\{ \mathbf{D}^\top \mathbf{P}^{(l)} \mathbf{\Delta}_\pi^{2(l)} \mathbf{D} \right\}^{-1} \mathbf{D}^\top \mathbf{P}^{(l)} \mathbf{\Delta}_\pi^{(l)} (\mathbf{w}^c - \boldsymbol{\pi}^{*(l)})$$

Passo 4: Fazer  $l = (l + 1)$  e atualizar  $\gamma^{(l)}$

Passo 5: Fazer passos 3, 4 até certo critério de convergência

**Algoritmo 3 - Componente contínua:**

Passo 1: Indicar o chute inicial para  $\boldsymbol{\theta}_1^{(0)}$

Passo 2: Iniciar o contador em  $l = 0$  e  $\boldsymbol{\theta}_1^{(l)} = \boldsymbol{\theta}_1^{(0)}$

Passo 3: Basado em  $\boldsymbol{\theta}_1^{(l)}$  fazer:

$$\begin{aligned} \boldsymbol{\beta}^{(l+1)} &= \boldsymbol{\beta}^{(l)} + \frac{1}{d_g} \left\{ \mathbf{X}^\top \mathbf{R} \mathbf{T}^{(l)} \mathbf{Q}^{2(l)} \mathbf{X} \right\}^{-1} \mathbf{X}^\top \mathbf{T}^{(l)} \mathbf{D}_{(v)}^{(l)} \mathbf{Q}^{(l)} (\mathbf{I}_n - \mathbf{W}^c) (\mathbf{w}^+ - \boldsymbol{\eta}^{*(l)}) \\ \boldsymbol{\alpha}^{(l+1)} &= \boldsymbol{\alpha}^{(l)} + \frac{4}{(f_g - 1)} \left\{ \mathbf{Z}^\top \mathbf{R} \mathbf{T}^{2(l)} \mathbf{\Delta}_\phi^{2(l)} \mathbf{Z} \right\}^{-1} \frac{1}{2} \mathbf{Z}^\top \mathbf{T}^{(l)} \mathbf{\Delta}_\phi^{(l)} (\mathbf{I}_n - \mathbf{W}^c) (\mathbf{s}^{(l)} - \mathbf{1}_n) \end{aligned}$$

Passo 4: Fazer  $l = (l + 1)$  e atualizar  $\boldsymbol{\theta}_1^{(l)}$

Passo 5: Fazer passos 3, 4 até certo critério de convergência

Foi implementado utilizando as extensões do pacote GAMLSS no Software R, em que o algoritmo CG (Rigby e Stasinopoulos, 2005) reduz ao algoritmo escore de Fisher.

Como a estimação de  $\boldsymbol{\theta}$  é obtida mediante o algoritmo escore de Fisher, então é importante estabelecer valores iniciais para a inicialização desse processo o qual estão fortemente ligados ao sucesso da convergência. Baseado nos dados, podemos escolher os chutes iniciais para  $\boldsymbol{\pi}$ ,  $\boldsymbol{\eta}$  e  $\boldsymbol{\phi}$  e conseqüentemente atribuir valores iniciais aos vetores  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\beta}$  e  $\boldsymbol{\alpha}$ . Vanegas e Paula (2016) sugeriram para a classe log-simétrica utilizar como chute inicial para  $\boldsymbol{\eta}$  e  $\boldsymbol{\phi}$  os EMV

dos parâmetros de uma log-normal. Desta forma, consideramos os chutes iniciais para  $\pi$ ,  $\eta$  e  $\phi$ , respectivamente, a proporção de zeros na amostra,  $\hat{\eta}$  e  $\hat{\phi}$ , em que os dois últimos são funções dos EMV de  $W|I_{\{0\}}(W) = 0 \sim LN(\eta, \phi)$ . Para o critério de convergência, é necessário estabelecer um critério de parada para o processo iterativo baseado em alguma medida, por exemplo, quando o valor absoluto das distancias entre as iterações for menor que uma constante fixada  $\epsilon$ ,  $\left| \frac{\hat{\theta}^{(l)} - \hat{\theta}^{(l-1)}}{\hat{\theta}^{(l-1)}} \right| < \epsilon$ .

### 3.1.4 Distribuição assintótica

Seja  $\hat{\theta}$  a estimativa máxima verossimilhança de  $\theta = (\beta^\top, \alpha^\top, \gamma^\top)^\top$ , sob as condições usuais de regularidade pode se mostrar que  $\hat{\theta}$  é um estimador consistente (ver Cysneiros et al., 2010; Cox e Hinkley, 1974, Cap. 9), assim, para  $n$  suficientemente grande temos

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}_{M_1+M_2+M_3}(0, \mathbf{J}_{\theta\theta}^{-1}); \text{ sendo } \mathbf{J}_{\theta\theta} = \lim_{n \rightarrow \infty} \mathbf{K}(\theta)/n$$

em que  $\mathcal{N}_{M_1+M_2+M_3}$  segue uma distribuição normal  $M_1 + M_2 + M_3$ -variada e  $\mathbf{K}(\theta)^{-1}$  é a matriz de variância-covariância assintótico de  $\hat{\theta}$ . Temos que  $\mathbf{K}(\hat{\theta})^{-1}$  é um estimador consistente de  $\mathbf{K}(\theta)^{-1}$ , como  $\mathbf{K}(\theta) = \text{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\alpha\alpha}, \mathbf{K}_{\gamma\gamma}\}$  é uma matriz bloco diagonal, então assintoticamente segue que  $\hat{\beta} \xrightarrow{D} \mathcal{N}_{M_1}(\beta, \hat{\mathbf{K}}^{\beta\beta})$ ,  $\hat{\alpha} \xrightarrow{D} \mathcal{N}_{M_2}(\alpha, \hat{\mathbf{K}}^{\alpha\alpha})$  e  $\hat{\gamma} \xrightarrow{D} \mathcal{N}_{M_3}(\gamma, \hat{\mathbf{K}}^{\gamma\gamma})$ , em que  $\hat{\mathbf{K}}^{\beta\beta}$ ,  $\hat{\mathbf{K}}^{\alpha\alpha}$  e  $\hat{\mathbf{K}}^{\gamma\gamma}$  são as matrizes invertidas de  $\hat{\mathbf{K}}_{\beta\beta}$ ,  $\hat{\mathbf{K}}_{\alpha\alpha}$  e  $\hat{\mathbf{K}}_{\gamma\gamma}$ , respectivamente. Desta forma pela normalidade assintótica do estimador de máxima verossimilhança, podemos construir intervalos de confiança assintóticos para os parâmetros do modelo de regressão. Temos assim,

$$\left[ \hat{\beta}_{j_1} - \Phi^{-1}\left(1 - \frac{\xi}{2}\right)(\hat{\mathbf{K}}_{j_1 j_1}^{\beta\beta})^{1/2}, \quad \hat{\beta}_{j_1} + \Phi^{-1}\left(1 - \frac{\xi}{2}\right)(\hat{\mathbf{K}}_{j_1 j_1}^{\beta\beta})^{1/2} \right],$$

para  $j_1 = 1, \dots, M_1$ ,

$$\left[ \hat{\alpha}_{j_2} - \Phi^{-1}\left(1 - \frac{\xi}{2}\right)(\hat{\mathbf{K}}_{j_2 j_2}^{\alpha\alpha})^{1/2}, \quad \hat{\alpha}_{j_2} + \Phi^{-1}\left(1 - \frac{\xi}{2}\right)(\hat{\mathbf{K}}_{j_2 j_2}^{\alpha\alpha})^{1/2} \right],$$

para  $j_2 = 1, \dots, M_2$  e

$$\left[ \hat{\gamma}_{j_3} - \Phi^{-1}\left(1 - \frac{\xi}{2}\right)(\hat{\mathbf{K}}_{j_3 j_3}^{\gamma\gamma})^{1/2}, \quad \hat{\gamma}_{j_3} + \Phi^{-1}\left(1 - \frac{\xi}{2}\right)(\hat{\mathbf{K}}_{j_3 j_3}^{\gamma\gamma})^{1/2} \right],$$

para  $j_3 = 1, \dots, M_3$ , são intervalos de confiança assintóticos para  $\beta_{j_1}$ ,  $\alpha_{j_2}$  e  $\gamma_{j_3}$  com um coeficiente de confiança de  $100(1 - \xi)\%$ . Para  $0 < \xi < 1/2$ ,  $\Phi^{-1}(1 - \frac{\xi}{2})$  representa o quantil  $1 - \frac{\xi}{2}$  de um normal padrão.

### 3.1.5 Teste de razão de verossimilhança

Suponha que o interesse é testar a significância de alguns parâmetros, no qual, pode ser utilizado como uma forma de avaliar a inclusão ou exclusão de variáveis explanatórias. Em geral podemos considerar a hipótese  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^0$  contra  $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}^0$ , em que  $\boldsymbol{\theta}^0$  é um vetor conhecido de dimensão  $(p \times 1)$  com  $p = M_1 + M_2 + M_3$ . A estatística da razão de verossimilhança  $\xi_{RV}$  é dada por  $2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}^0)\}$ . Segue que sob  $H_0$  e para amostras  $n$  suficientemente grande  $\xi_{RV}$  converge para uma distribuição qui-quadrada com  $p$  graus de liberdade.

### 3.1.6 Seleção de variáveis explanatórias

Uma vez que o conjunto de variáveis explanatórias está definido para um modelo proposto, um ponto importante é de saber qual a melhor maneira de obter um modelo reduzido que estejam só as variáveis explanatórias que melhor possam explicar a variável resposta. Para esse problema utilizaremos a metodologia do método de Akaike, que para nosso caso, consiste em obter as melhores variáveis explanatórias para cada parâmetro a ser modelado minimizando a AIC pela metodologia de abordagem para frente, para melhores detalhes ver Yamashita et al. (2007). O abordagem StepAIC e a metodologia a seguir para o modelo de regressão ZALS consiste em

#### **Algoritmo 4 - Abordagem StepAIC:**

- Passo 1: Modelo nulo: Ajustar o modelo de regressão sem variáveis explanatórias (modelo nulo) e obter o AIC.
- Passo 2: Seleção inicial: Ajustar  $k$  modelos para as  $k$  variáveis explanatórias, obter o AIC para cada modelo e selecionar a variável explanatórias para o modelo que obtenha o menor AIC que em comparação ao modelo nulo. Se não ter seleção, ir ao passo 5
- Passo 3: Seleção de outra variável explanatória: ajustar outros modelos de regressão com a variável selecionada anterior e selecionar a nova variável que minimize o AIC que em comparação ao modelo do passo anterior. Se não ter mais seleções, ir ao passo 5, caso contrário ir ao passo 4.

Passo 4: Deletar variáveis explanatórias insignificantes: comparar o AIC para todos os modelos com as variáveis explanatórias  $x_i$ 's sem a variável  $x_j$  e deletar  $x_j$  se o modelo obtido obtenha menor AIC. Se não ter mais variáveis insignificantes fazer o passo 3, caso contrario fazer passo 4.

Passo 5: Parar o algoritmo.

**Algoritmo 5 - Seleção de variáveis para modelos da classe ZALS:**

1. Para a componente discreta, o submodelo  $\pi$ , construir o modelo em base ao algoritmo do abordagem StepAIC.
2. Para a componente contínua:
  - 2.1. No submodelo  $\eta$ , construir o modelo em base ao algoritmo do abordagem StepAIC.
  - 2.2. Dado o submodelo  $\eta$ , construir o modelo para  $\phi$  em base ao algoritmo do abordagem StepAIC.

Vale a pena dizer que eventualmente algumas variáveis explanatórias ou covariáveis selecionadas podem não ser significativas marginalmente para o nosso modelo estimado. Assim, podemos utilizar o teste de razão de verossimilhança para avaliar a retirada delas.

### 3.2 RESÍDUOS E MÉTODOS DE DIAGNÓSTICO

Os métodos de análise de resíduos são usados após ajustar um modelo de regressão para verificar adequacidade, ou melhor, possíveis afastamentos das suposições feitas ao modelo, como por exemplo a escolha da função de ligação ou da distribuição da componente aleatória.

#### 3.2.1 Resíduos

Vários tipos de resíduos podem ser encontrados na literatura, Cox e Snell (1968) apresentam uma forma bastante geral de resíduos. Pregibon (1981) propõe utilizar o componente do desvio como resíduo para os modelos lineares generalizados. McCullagh (1987) apresenta uma forma padronizada para o componente do desvio. Dunn e Smith (1996) propõem o resíduo quantílico. Cysneiros e Paula (2005) e Cysneiros e Vanegas (2008) apresentam resíduos para a classe

dos modelos de regressão não lineares simétricos, Vanegas e Paula (2015) propõem o resíduo componente do desvio para os modelos log-simétricos. Recentemente Pereira et al. (2020) apresentam uma nova classe de resíduos para os modelos de regressão zero ajustados em que proporciona uma melhor forma de identificar pontos atípicos na componente contínua.

### 3.2.1.1 Resíduo componente do desvio

Para o nosso modelo de regressão, com três parâmetros a ser modelados, definimos as funções desvio para  $\eta$ ,  $\phi$  e  $\pi$ .

#### **Resíduo componente do desvio para $\hat{\eta}$**

A função desvio para  $\hat{\eta}$  dado  $\hat{\phi}$  e  $\hat{\pi}$  é

$$D(\hat{\eta}|\hat{\phi}, \hat{\pi}) = 2 \sum_{i=1}^n [\ell_i(\tilde{\eta}_i, \hat{\phi}_i, \hat{\pi}_i) - \ell_i(\hat{\eta}_i, \hat{\phi}_i, \hat{\pi}_i)], \quad (3.9)$$

pela separabilidade dos parâmetros  $(\eta^\top, \phi^\top)^\top$  e  $\pi$ , a Equação (3.9) pode ser expressada

$$D(\hat{\eta}|\hat{\phi}) = 2 \sum_{i:w_i>0} [\ell_i(\tilde{\eta}_i, \hat{\phi}_i) - \ell_i(\hat{\eta}_i, \hat{\phi}_i)] = \sum_{i:w_i>0} d_i(\hat{\eta}|\hat{\phi}), \quad (3.10)$$

em que  $\hat{\eta}_i$  é a estimativa de máxima verossimilhança de  $\eta_i$  e  $\tilde{\eta}_i$  é o valor de  $\eta_i$  que maximiza  $\ell_i(\eta_i, \hat{\phi}_i)$ . Vale notar que a Equação (3.10) coincide com a função dada por Vanegas e Paula (2015). O resíduo componente de desvio associado a  $\hat{\eta}$  é definido como

$$t_\eta(\hat{w}_i) = \text{sign}(\hat{w}_i)[d_i(\hat{\eta}|\hat{\phi})]^{1/2}, \quad \text{para } w_i > 0. \quad (3.11)$$

Substituindo o valor de  $\tilde{\eta}_i$  por  $w_i$ , obtemos,  $d_i(\hat{\eta}|\hat{\phi}) = 2\log[g(0)/g(\hat{w}_i^2)]$ .

#### **Resíduo componente do desvio para $\hat{\phi}$**

A função de desvio para  $\hat{\phi}$  dado  $\hat{\eta}$  é definido por

$$D(\hat{\phi}|\hat{\eta}) = 2 \sum_{i:w_i>0} [\ell_i(\hat{\eta}_i, \tilde{\phi}_i) - \ell_i(\hat{\eta}_i, \hat{\phi}_i)] = \sum_{i:w_i>0} d_i(\hat{\phi}|\hat{\eta}), \quad (3.12)$$

em que  $\tilde{\phi}_i$  é o valor de  $\phi_i$  que maximiza  $\ell_i(\hat{\eta}_i, \phi_i)$ . Portanto,  $d_i(\hat{\phi}|\hat{\eta}) = 2\log[g(\varrho^2)/g(\hat{w}_i^2)] - \log[\hat{w}_i^2/\varrho^2]$ .  $\ell_i(\hat{\eta}_i, \phi_i)$  tem só um ponto crítico, logo  $\varrho$  é a solução da equação  $v(\varrho)\varrho^2 = 1$ . O resíduo componente de desvio associado a  $\hat{\phi}$  é dado por

$$t_{\phi}(\hat{w}_i) = \text{sinal}(\hat{w}_i)[d_i(\hat{\phi}|\hat{\eta})]^{1/2}, \quad \text{para } w_i > 0, \quad (3.13)$$

as Equações (3.11) e (3.13) coincidem com os resíduos propostos por Vanegas e Paula (2015).

### **Resíduo componente do desvio para $\hat{\pi}$**

A função de desvio para  $\hat{\pi}$  pode ser obtido separadamente através do logaritmo da função de verosimilhança da componente discreta e é dado por

$$D(\hat{\pi}) = 2 \sum_{i=1}^n [\ell_i(\tilde{\pi}_i) - \ell_i(\hat{\pi}_i)] = \sum_{i=1}^n d_i(\hat{\pi}), \quad (3.14)$$

em que  $\tilde{\pi}_i$  é o valor de  $\pi_i$  que maximiza  $\ell_i(\pi_i)$ . Assim  $d_i(\hat{\pi}) = 2\{I_{\{0\}}(w_i)\log[I_{\{0\}}(w_i)/\hat{\pi}_i] + [1 - I_{\{0\}}(w_i)]\log[(1 - I_{\{0\}}(w_i))/(1 - \hat{\pi}_i)]\}$ . Todavia, quando  $I_{\{0\}}(w_i) = 0$  ou  $I_{\{0\}}(w_i) = 1$ , o  $i$ -ésimo termo,  $d_i(\hat{\pi})$ , vale  $-2\log[1 - \hat{\pi}_i]$  ou  $-2\log[\hat{\pi}_i]$ , respectivamente. Assim o resíduo componente de desvio associado a  $\hat{\pi}$  é dado por

$$t_{\pi_i} = \begin{cases} -(-2\log[1 - \hat{\pi}_i])^{1/2}, & \text{se } I_{\{0\}}(w_i) = 0, \\ (-2\log[\hat{\pi}_i])^{1/2}, & \text{se } I_{\{0\}}(w_i) = 1. \end{cases}$$

#### **3.2.1.2 Resíduo quantílico**

Seja  $F_W(w)$  a função de distribuição acumulada definida na Equação 2.1. Assim, considerando  $F$  contínua no intervalo  $w_i > 0$ , segue que  $F_W(w)$  é uniformemente distribuída no intervalo unitário. Portanto, o resíduo quantílico baseado em Dunn e Smyth (1996),  $t_Q(\hat{w}_i)$ , é expressado por

$$t_Q(\hat{w}_i) = \begin{cases} \Phi^{-1}[u_i], & \text{se } w_i = 0, \\ \Phi^{-1}[F_W(\hat{w}_i)], & \text{se } w_i > 0, \end{cases}$$

em que  $\Phi$  é a distribuição acumulada de uma normal padronizada e  $u_i$  é uma variável uniformemente distribuída no intervalo  $(0, \hat{\pi}_i)$ .

### 3.2.1.3 Resíduo ZAR

De acordo com Pereira et al. (2020) o resíduo quantal apresenta limitações quando a  $w_i$  é muito próximo a zero e a probabilidade,  $\hat{\pi}_i$ , é em torno de 0,5, pois o valor de  $t_Q(\hat{w}_i)$  será negativo e o valor absoluto não será grande. Assim, Pereira et al. (2020), propuseram uma classe de resíduos muito mais sensível no sentido de identificar pontos atípicos nos modelos de regressão zero ajustado que em comparação a outros resíduos usualmente utilizados. A classe de resíduos para modelos zero ajustados (ZAR) é uma função de qualquer resíduo definido para a componente contínua e da probabilidade estimada da observação assumir zero.

$$t_*(\hat{w}_i) = \begin{cases} \Phi^{-1}[\Phi(t_i)(1 - \hat{\pi}_i)], & \text{se } t_i < 0, \\ \Phi^{-1}[\hat{\pi}_i + \Phi(t_i)(1 - \hat{\pi}_i)], & \text{se } t_i > 0, \end{cases} \quad (3.15)$$

em que  $t_i$  é qualquer tipo de resíduo para a componente contínua da variável resposta. Assim, a Equação (3.15) representa uma classe de resíduos para os modelos de regressão zero ajustados.

### 3.2.2 Influência Global

Outro aspecto importante é a identificação de observações influentes nas estimativas de máxima verossimilhança dos parâmetros do modelo proposto. Assim, uma forma de identificar as observações influentes, baseia-se na metodologia proposta por Cook (1977) para os modelos lineares de resposta normal, denominada modelo de deleção de casos (CDM). Esta metodologia consiste em avaliar o impacto nas estimativas dos parâmetros do modelo através de uma medida de distância quando a  $i$ -ésima observação é excluída,  $\hat{\boldsymbol{\theta}}_{(i)}$ , através do afastamento da verossimilhança

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

em que  $\ell(\boldsymbol{\theta})$  e  $\ell(\boldsymbol{\theta}_{(i)})$  é o logaritmo da função de verossimilhança para os dados completos e excluindo a observação  $i$ , respectivamente. Pelo alto custo computacional que pode requerer estimar  $\boldsymbol{\theta}_{(i)}$  para  $i = 1, \dots, n$ , podemos obter uma aproximação de  $\hat{\boldsymbol{\theta}}_{(i)}$  maximizando uma aproximação de  $\ell_{(i)}(\boldsymbol{\theta})$ , o logaritmo da função de verossimilhança excluindo a  $i$ -ésima observação. Seja a expansão de Taylor até a segunda ordem de  $\ell_{(i)}(\boldsymbol{\theta})$  em torno de  $\hat{\boldsymbol{\theta}}$  dada por

$$\ell_{(i)}(\boldsymbol{\theta}) \approx \ell_{(i)}(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \mathbf{U}_{(i)}(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top (-\ddot{L}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}}^{(i)})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),$$

em que  $U_{(i)}(\hat{\boldsymbol{\theta}})$  e  $\ddot{L}_{\theta\theta}^{(i)}$  são o vector escore e a matriz de informação de Fisher observada avaliados na estimativa de máxima verossimilhança, respectivamente, excluída a  $i$ -ésima observação. Esta aproximação sugerida por Pregibon (1981), denominada de aproximação a um passo, consiste em tomar a primeira iteração do processo iterativo pelo método de Scoring de Fisher. A expressão acima atinge o máximo quando

$$\hat{\boldsymbol{\theta}}_{(i)}^I = \hat{\boldsymbol{\theta}} - \ddot{L}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}}^{(i)-1} U_{(i)}(\hat{\boldsymbol{\theta}}).$$

Substituindo  $-\ddot{L}_{\theta\theta}^{(i)}$  pelo seu valor esperado  $\mathbf{K}(\boldsymbol{\theta})^{(i)}$ , logo  $\hat{\boldsymbol{\theta}}_{(i)}^I$  pode ser expressado como

$$\hat{\boldsymbol{\theta}}_{(i)}^I = \hat{\boldsymbol{\theta}} + \mathbf{K}(\hat{\boldsymbol{\theta}})^{(i)-1} U_{(i)}(\hat{\boldsymbol{\theta}}).$$

Após algumas manipulações algébricas, temos

$$\hat{\boldsymbol{\theta}}_{(i)}^I = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{(i)}^I \\ \hat{\boldsymbol{\alpha}}_{(i)}^I \\ \hat{\boldsymbol{\gamma}}_{(i)}^I \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}} - c_i^{(\eta)} |_{\hat{\boldsymbol{\theta}}} (\mathbf{X}^\top \hat{\mathbf{R}} \hat{\boldsymbol{\Gamma}} \hat{\mathbf{Q}}^2 \mathbf{X})^{-1} \mathbf{X}_i / (1 - \hat{h}_{ii}^{(\eta)}) \\ \hat{\boldsymbol{\alpha}} - c_i^{(\phi)} |_{\hat{\boldsymbol{\theta}}} (\mathbf{Z}^\top \hat{\mathbf{R}} \hat{\boldsymbol{\Gamma}}^2 \hat{\boldsymbol{\Delta}}_\phi^2 \mathbf{Z})^{-1} \mathbf{Z}_i / (1 - \hat{h}_{ii}^{(\phi)}) \\ \hat{\boldsymbol{\gamma}} - c_i^{(\pi)} |_{\hat{\boldsymbol{\theta}}} (\mathbf{D}^\top \hat{\mathbf{P}} \hat{\boldsymbol{\Delta}}_\pi^2 \mathbf{D})^{-1} \mathbf{D}_i / (1 - \hat{h}_{ii}^{(\pi)}) \end{bmatrix}, \quad (3.16)$$

sendo que  $\mathbf{X}_i^\top$ ,  $\mathbf{Z}_i^\top$  e  $\mathbf{D}_i^\top$  são a  $i$ -ésima linha de  $\mathbf{X}$ ,  $\mathbf{Z}$  e  $\mathbf{D}$ , além, tem-se que  $c_i^{(\eta)} = \frac{1}{d_g} [1 - I_{\{0\}}(w_i)] (\tilde{w}_i / \sqrt{\phi_i}) \mathbf{v}(\tilde{w}_i) (\partial \eta_i / (\eta_i \partial \xi_{1i}))$ ,  $c_i^{(\phi)} = \frac{2}{(f_g - 1)} [1 - I_{\{0\}}(w_i)] (\mathbf{v}(\tilde{w}_i) \tilde{w}_i^2 - 1) (\partial \phi_i / (\phi_i \partial \xi_{2i}))$ ,  $c_i^{(\pi)} = \frac{(I_{\{0\}}(w_i) - \pi_i)}{\pi_i (1 - \pi_i)} (\partial \pi_i / \partial \xi_{3i})$ ,  $h_{ii}^{(\eta)} = \frac{1}{\phi_i} (1 - \pi_i) (\partial \eta_i / (\eta_i \partial \xi_{1i}))^2 \mathbf{X}_i^\top (\mathbf{X}^\top \mathbf{R} \mathbf{T} \mathbf{Q}^2 \mathbf{X})^{-1} \mathbf{X}_i$ ,  $h_{ii}^{(\phi)} = (1 - \pi_i) (\partial \phi_i / (\phi_i \partial \xi_{2i}))^2 \mathbf{Z}_i^\top (\mathbf{Z}^\top \mathbf{R} \mathbf{T}^2 \boldsymbol{\Delta}_\phi^2 \mathbf{Z})^{-1} \mathbf{Z}_i$  e  $h_{ii}^{(\pi)} = \frac{(\partial \pi_i / (\partial \xi_{3i}))^2}{\pi_i (1 - \pi_i)} \mathbf{D}_i^\top (\mathbf{D}^\top \mathbf{P} \boldsymbol{\Delta}_\pi^2 \mathbf{D})^{-1} \mathbf{D}_i$ .

Podemos obter uma aproximação para  $LD(\hat{\boldsymbol{\theta}}_{(i)})$  de segunda ordem de Taylor, obtendo

$$LD(\hat{\boldsymbol{\theta}}_{(i)}) \approx (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top [-\ddot{L}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}}] (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}). \quad (3.17)$$

Neste trabalho no lugar de  $-\ddot{L}_{\theta\theta}$  consideramos o valor esperado  $\mathbf{K}(\boldsymbol{\theta})$ . Substituindo (3.16) em (3.17) obtemos uma aproximação para  $LD(\hat{\boldsymbol{\theta}}_{(i)})$  que pode ser expressa como

$$LD(\hat{\boldsymbol{\theta}}_{(i)}^I) \approx (\hat{\boldsymbol{\theta}}_{(i)}^I - \hat{\boldsymbol{\theta}})^\top \mathbf{K}(\hat{\boldsymbol{\theta}}) (\hat{\boldsymbol{\theta}}_{(i)}^I - \hat{\boldsymbol{\theta}}),$$

dado que os estimadores de máxima verossimilhança de  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$  e  $\boldsymbol{\gamma}$  são assintoticamente independentes, podemos usar a expressão acima para mostrar que o efeito da  $i$ -ésima observação na estimativa de  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$  e  $\boldsymbol{\gamma}$  pode ser expressada como

$$\begin{aligned}
LD(\hat{\beta}_{(i)}^I) &= \frac{1}{d_g(1 - \hat{\pi}_i)} [1 - I_{\{0\}}(w_i)] \hat{w}_i^2 v^2(\hat{w}_i) \frac{\hat{h}_{ii}^{(\eta)}}{(1 - \hat{h}_{ii}^{(\eta)})^2}, \\
LD(\hat{\alpha}_{(i)}^I) &= \frac{1}{(f_g - 1)(1 - \hat{\pi}_i)} [1 - I_{\{0\}}(w_i)] (v(\hat{w}_i) \hat{w}_i^2 - 1)^2 \frac{\hat{h}_{ii}^{(\phi)}}{(1 - \hat{h}_{ii}^{(\phi)})^2}, \\
LD(\hat{\gamma}_{(i)}^I) &= \frac{(I_{\{0\}}(w_i) - \hat{\pi}_i)^2}{\hat{\pi}_i(1 - \hat{\pi}_i)} \frac{\hat{h}_{ii}^{(\pi)}}{(1 - \hat{h}_{ii}^{(\pi)})^2},
\end{aligned}$$

pode ser observado que quando  $\{w_i = 0\}$ ,  $LD(\hat{\beta}_{(i)}^I) = 0$  e  $LD(\hat{\alpha}_{(i)}^I) = 0$ .

Gráficos de  $LD(\hat{\beta}_{(i)}^I)$ ,  $LD(\hat{\alpha}_{(i)}^I)$  e  $LD(\hat{\gamma}_{(i)}^I)$  contra os índices das observações são recomendados para detectar pontos aberrantes.

### 3.2.3 Influência Local

Influência local tem como objetivo verificar os efeitos de pequenas perturbações no modelo e/ou nos dados por meio de algum tipo de perturbação. Caso tais perturbações causem variações desproporcionais, isto indica que pode haver indícios de que o modelo esteja mal ajustado ou com problemas de afastamento das suposições feitas para o mesmo.

Cook (1986) propôs avaliar a influência das observações sob pequenas perturbações nos dados ou no modelo através do estudo da curvatura normal do afastamento da verossimilhança. Conhecido como influência local, nesta mesma direção, Lawrence (1988) investigou a aplicação de influência local nos modelos lineares com parâmetros na transformação da resposta, Paula (1993) aplica influência local em modelos lineares com restrições, Galea, Paula e Cysneiros (2005) desenvolveram influência local em modelos de regressão simétricos, Vanegas e Paula (2015) aplicaram influência local para os modelos log-simétricos, Tomazella et al. (2019) desenvolveram influência local utilizando o esquema de perturbação de casos ponderados para os modelos de regressão zero ajustado reparametrizado Birnbaum-Saunders, recentemente, Queiroz e Lemonte (2020) propuseram influência local para os esquemas de perturbação de casos ponderados e nas variáveis explanatórias para uma classe de modelos de regressão para taxas e proporções.

Seja o vetor de perturbações  $\omega = (\omega_1, \dots, \omega_n)^\top$  e  $l(\theta)$  o logaritmo da função de verossimilhança. A medida de influência mais conhecida é o afastamento da verossimilhança dada por

$$LD(\omega) = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\omega)\} \geq 0,$$

em que  $\hat{\boldsymbol{\theta}}$  e  $\hat{\boldsymbol{\theta}}_\omega$  denota as estimativas de máxima verossimilhança sob o modelo não perturbado e perturbado, respectivamente.

A sugestão de Cook (1986) é estudar o comportamento do afastamento da verossimilhança,  $LD(\omega) = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_\omega)\}$ , em torno ao vetor de não perturbação  $\omega_0$  através da curvatura normal da linha projetada  $LD(\omega_0 + a\boldsymbol{\ell})$ , em que  $a \in \mathbb{R}$  e  $\boldsymbol{\ell}$  é um vetor unitário, ou seja,  $\|\boldsymbol{\ell}\| = 1$ . Assim, Cook (1986) mostra que a curvatura normal pode ser expressada numa forma geral dada por

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

em que  $\Delta$  é uma matriz de dimensão  $(M_1 + M_2 + M_3) \times n$  dada por  $\Delta = \partial^2 \ell(\boldsymbol{\theta}|\omega) / \partial \theta \partial \omega$  avaliada em  $\hat{\boldsymbol{\theta}}$  e  $\omega_0$ .

Cook (1986) sugere utilizar o autovetor,  $\boldsymbol{\ell}_{max}$ , correspondente ao maior autovalor  $C_{\boldsymbol{\ell}_{max}}$  da matriz  $\ddot{F} = -\Delta^\top \ddot{L}_{\theta\theta}^{-1} \Delta$ . O gráfico de índices de  $\boldsymbol{\ell}_{max}$  contra a ordem das observações pode mostrar os pontos que exercem maior influência nas estimativas de  $\boldsymbol{\theta}$  mediante pequenas perturbações. Contudo, é possível avaliar a influência local para um subvetor do vetor de parâmetro  $\boldsymbol{\theta}$ , por exemplo, se o interesse é somente no vetor da componente discreta  $\boldsymbol{\gamma}$ , nesse caso a curvatura normal fica dada por

$$C_\ell(\boldsymbol{\gamma}) = 2|\boldsymbol{\ell}^\top \Delta^\top (\ddot{L}_{\theta\theta}^{-1} - L_1) \Delta \boldsymbol{\ell}|,$$

em que

$$L_1 = \begin{bmatrix} 0 & 0 \\ 0 & \ddot{L}_{\theta_1\theta_1}^{-1} \end{bmatrix},$$

em que  $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$ .

Outra alternativa, proposta por Escobar e Meeker (1992), consiste em tomar como medida de influência os elementos da diagonal principal  $\ddot{F}$ . Lesaffre e Verbeke (1998) sugerem em avaliar a curvatura normal na direção da  $i$ -ésima observação. Essa curvatura é denominada e dada por  $C_i = 2|\Delta_i^\top \ddot{L}_{\theta\theta}^{-1} \Delta_i|$ , em que  $\Delta_i$  é a  $i$ -ésima coluna da matriz  $\Delta$ . Assim, é sugerido um gráfico de  $C_i$  versus o índice, e que as observações tais que  $C_i > 2\bar{C}$  tenham uma atenção especial.

Neste estudo, apresentamos dois esquemas de perturbação nos modelos lineares zero ajustado log-simétricas: perturbação de casos ponderados e perturbação nas variáveis explanatórias. No caso de perturbação na variável resposta, não é possível obter uma forma de perturbar por causa da variável resposta ser semi-contínua.

### 3.2.3.1 Perturbação de casos ponderados

Seja o vetor de perturbações  $\omega = (\omega_1, \dots, \omega_n)^\top$ . A função de log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\theta}|\omega) = \ell_{(1)}(\boldsymbol{\gamma}|\omega) + \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega) \quad (3.18)$$

é a função de log-verossimilhança perturbada em que,

$$\ell_{(1)}(\boldsymbol{\gamma}|\omega) = \sum_{i=1}^n \omega_i \ell_i(\pi_i), \quad \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega) = \sum_{i=1}^n \omega_i \ell_i(\eta_i, \phi_i),$$

em que  $\ell_i(\pi_i)$  e  $\ell_i(\eta_i, \phi_i)$  são as funções de log-verossimilhança definidas em (3.4). Para o esquema de perturbação de casos ponderados, o vetor de não perturbação é  $\omega_0 = (1, 1, \dots, 1)^\top$  e a matriz  $\Delta$  é dada por

$$\Delta = \begin{bmatrix} \Delta_\beta \\ \Delta_\alpha \\ \Delta_\gamma \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \hat{\mathbf{T}}^{1/2} \hat{\mathbf{D}}_{(v)} \hat{\mathbf{Q}} (\mathbf{I}_n - \mathbf{W}^c) \hat{\mathbf{D}}_{(\tilde{w})} \\ \frac{1}{2} \mathbf{Z}^\top \hat{\mathbf{T}} \hat{\mathbf{A}}_\phi (\mathbf{I}_n - \mathbf{W}^c) (\hat{\mathbf{D}}_{(v)} \hat{\mathbf{D}}_{(\tilde{w})}^2 - \mathbf{I}_n) \\ \mathbf{D}^\top \hat{\mathbf{P}} \hat{\mathbf{A}}_\pi \hat{\mathcal{E}} \end{bmatrix}, \quad (3.19)$$

em que  $\hat{\mathbf{D}}_{(\tilde{w})}$  corresponde a  $\mathbf{D}_{(\tilde{w})} = \text{diag}\left\{\frac{w_1^+ - \eta_1^*}{\sqrt{\phi_1}}, \dots, \frac{w_n^+ - \eta_n^*}{\sqrt{\phi_n}}\right\}$ ,  $\hat{\mathcal{E}}$  corresponde a  $\mathcal{E} = \text{diag}\{(I_{\{0\}}(w_1) - \pi_1), \dots, (I_{\{0\}}(w_n) - \pi_n)\}$  avaliados no estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  e  $\omega_0$ . Para maiores detalhes ver Apêndice B.

### 3.2.3.2 Perturbação nas variáveis explanatórias

Nesta seção apresentamos a perturbação nas variáveis explanatórias de forma aditiva, o esquema será dividido em: perturbação nas variáveis explanatórias individualmente e simultaneamente.

#### *Perturbação nas variáveis explanatórias individualmente*

#### ***Perturbação individual nas variáveis explanatórias da componente discreta***

Considere o esquema em que é adicionado um vetor de perturbação a uma variável explanatória particular da componente discreta,  $d_j$ , ponderado por um fator de escala  $s_{d_j}$

podendo ser o desvio padrão, ou seja,

$$d_{ij\omega} = d_{ij} + \omega_i s_{d_j}, \quad i = 1, \dots, n;$$

a perturbação leva ao esquema em que  $m_3(\pi_{i\omega}) = \mathbf{D}_{i\omega}^\top \boldsymbol{\gamma}$  e  $\mathbf{D}_{i\omega} = (d_{i1}, \dots, d_{ij} + \omega_i s_{d_j}, \dots, d_{iM_3})^\top$ .

Logo, o logaritmo da função de verossimilhança perturbada assume a forma

$$\ell(\boldsymbol{\theta}) = \ell_{(1)}(\boldsymbol{\gamma}|\omega) + \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}),$$

em que  $\ell_{(1)}(\boldsymbol{\gamma}|\omega) = \sum_{i=1}^n \ell_i(\pi_{i\omega})$ , aqui  $\ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha})$  é desprezível. Logo, depois de algumas manipulações algébricas, avaliando em  $\hat{\gamma}_j$  e no vetor de não perturbação  $\omega_0 = (0, \dots, 0)^\top$  obtemos

$$\Delta_\gamma = s_{d_j} \{ \mathbf{F}_1 \hat{\mathbf{P}} \hat{\boldsymbol{\Delta}}_\pi \hat{\boldsymbol{\epsilon}} + \hat{\gamma}_j \mathbf{D}^\top \hat{\boldsymbol{\psi}}_1 \},$$

em que  $\mathbf{F}_1$  é uma matriz  $M_3 \times n$  de zeros com uns na  $j$ -ésima linha,  $\boldsymbol{\psi}_1$  é uma matriz diagonal definida na Equação (3.6).

### ***Perturbação individual nas variáveis explanatórias da componente contínua do submodelo associado a $\eta$***

Agora, considere o esquema em que é adicionado um vetor de perturbação a uma variável explanatória particular da componente contínua do submodelo associado a  $\eta$  ponderado por um fator de escala  $s_{x_p}$  podendo ser o desvio padrão, i.e.,

$$x_{ip\omega} = x_{ip} + \omega_i s_{x_p}, \quad i = 1, \dots, n;$$

a perturbação leva ao esquema em que  $m_1(\eta_{i\omega}) = \mathbf{X}_{i\omega}^\top \boldsymbol{\beta}$  e  $\mathbf{X}_{i\omega} = (x_{i1}, \dots, x_{ip} + \omega_i s_{x_p}, \dots, x_{iM_1})^\top$ .

Logo, o logaritmo da função de verossimilhança perturbada assume a forma

$$\ell(\boldsymbol{\theta}) = \ell_{(1)}(\boldsymbol{\gamma}) + \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega),$$

em que  $\ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega) = \sum_{i=1}^n \ell_i(\eta_{i\omega}, \phi_i)$ , aqui  $\ell_{(1)}(\boldsymbol{\gamma})$  é desprezível. Logo, depois de algumas manipulações algébricas, avaliando em  $\hat{\boldsymbol{\theta}}_1$  e no vetor de não perturbação  $\omega_0 = (0, \dots, 0)^\top$  obtemos

$$\Delta_{\boldsymbol{\theta}_1} = \begin{bmatrix} \Delta_\beta \\ \Delta_\alpha \end{bmatrix} = \begin{bmatrix} s_{x_p} \{ \mathbf{F}_2 \hat{\mathbf{T}}^{1/2} \hat{\mathbf{D}}_{(v)} \hat{\mathbf{Q}} (\mathbf{I}_n - \mathbf{W}^c) \hat{\mathbf{D}}_{(\tilde{w})} + \hat{\beta}_p \mathbf{X}^\top \hat{\boldsymbol{\psi}}_2 \} \\ s_{x_p} \hat{\beta}_p \mathbf{Z}^\top \hat{\boldsymbol{\psi}}_3 \end{bmatrix},$$

em que  $\mathbf{F}_2$  é uma matriz  $M_1 \times n$  de zeros com uns na  $p$ -ésima linha,  $\boldsymbol{\psi}_2$  e  $\boldsymbol{\psi}_3$  são matrizes diagonais definidas na Equação (3.6).

### ***Perturbação individual nas variáveis explanatórias da componente contínua do submodelo associado a $\phi$***

De forma análoga, considere o esquema em que é adicionado um vetor de perturbação a uma variável explanatória particular da componente contínua do submodelo associado a  $\eta$  ponderado por um fator de escala  $s_{z_q}$  podendo ser o desvio padrão, i.e.,

$$z_{iq\omega} = z_{iq} + \omega_i s_{z_q}, \quad i = 1, \dots, n;$$

a perturbação leva ao esquema em que  $m_2(\phi_{i\omega}) = \mathbf{Z}_{i\omega}^\top \boldsymbol{\alpha}$  e  $\mathbf{Z}_{i\omega} = (z_{i1}, \dots, z_{iq} + \omega_i s_{z_q}, \dots, z_{iM_2})^\top$ . Logo, o logaritmo da função de verossimilhança perturbada assume a forma

$$\ell(\boldsymbol{\theta}) = \ell_{(1)}(\boldsymbol{\gamma}) + \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega),$$

em que  $\ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega) = \sum_{i=1}^n \ell_i(\eta_i, \phi_{i\omega})$ , aqui também  $\ell_{(1)}(\boldsymbol{\gamma})$  é desprezível. Logo, depois de algumas manipulações algébricas, avaliando em  $\hat{\boldsymbol{\theta}}_1$  e no vetor de não perturbação  $\omega_0 = (0, \dots, 0)^\top$  obtemos

$$\Delta_{\boldsymbol{\theta}_1} = \begin{bmatrix} \Delta_{\boldsymbol{\beta}} \\ \Delta_{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} s_{z_q} \hat{\alpha}_q \mathbf{X}^\top \hat{\boldsymbol{\psi}}_3 \\ s_{z_q} \left\{ \frac{1}{2} \mathbf{F}_3 \hat{\mathbf{T}} \hat{\mathbf{D}}_\phi (\mathbf{I}_n - \mathbf{W}^c) (\hat{\mathbf{D}}_{(v)} \hat{\mathbf{D}}_{(\bar{w})}^2 - \mathbf{I}_n) + \hat{\alpha}_q \mathbf{Z}^\top \hat{\boldsymbol{\psi}}_4 \right\} \end{bmatrix},$$

em que  $\mathbf{F}_3$  é uma matriz  $M_2 \times n$  de zeros com uns na  $q$ -ésima linha,  $\boldsymbol{\psi}_4$  é uma matriz diagonal definida na Equação (3.6).

### ***Perturbação nas variáveis explanatórias simultaneamente***

#### ***Perturbação simultânea nas variáveis explanatórias da componente contínua***

Considere o esquema em que é adicionado um vetor de perturbação a uma variável explanatória particular da componente contínua do submodelo associado a  $\eta$ , e considere também que é feito o mesmo para o submodelo associado a  $\phi$ , assim, o esquema de ponderações é definido como

$$x_{ip\omega} = x_{ip} + \omega_i s_{x_p},$$

$$z_{iq\omega} = z_{iq} + \omega_i s_{z_q},$$

para qualquer  $p = 1, \dots, M_1$  e  $q = 1, \dots, M_2$ . Contudo, neste caso se considera que  $x_p \neq z_q$ . Essa perturbação leva ao esquema em que  $m_1(\eta_{i\omega}) = \mathbf{X}_{i\omega}^\top \boldsymbol{\beta}$ ,  $\mathbf{X}_{i\omega} = (x_{i1}, \dots, x_{ip} + \omega_i s_{x_p}, \dots, x_{iM_1})^\top$ ,

$m_2(\phi_{i\omega}) = \mathbf{Z}_{i\omega}^\top \boldsymbol{\alpha}$  e  $\mathbf{Z}_{i\omega} = (z_{i1}, \dots, z_{iq} + \omega_i s_{z_q}, \dots, z_{iM_2})^\top$ . Logo, o logaritmo da função de verossimilhança perturbada assume a forma

$$\ell(\boldsymbol{\theta}) = \ell_{(1)}(\boldsymbol{\gamma}) + \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega),$$

em que  $\ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega) = \sum_{i=1}^n \ell_i(\eta_{i\omega}, \phi_{i\omega})$ , aqui também  $\ell_{(1)}(\boldsymbol{\gamma})$  é desprezível. Logo, depois de algumas manipulações algébricas, avaliando em  $\hat{\boldsymbol{\theta}}_1$  e no vetor de não perturbação  $\omega_0 = (0, \dots, 0)^\top$  obtemos

$$\Delta_{\theta_1} = \begin{bmatrix} \Delta_\beta \\ \Delta_\alpha \end{bmatrix} = \begin{bmatrix} s_{x_p} \{ \mathbf{F}_2 \hat{\mathbf{T}}^{1/2} \hat{\mathbf{D}}_{(v)} \hat{\mathbf{Q}} (\mathbf{I}_n - \mathbf{W}^c) \hat{\mathbf{D}}_{(\tilde{w})} + \hat{\beta}_p \mathbf{X}^\top \hat{\boldsymbol{\psi}}_2 \} + s_{z_q} \hat{\alpha}_q \mathbf{X}^\top \hat{\boldsymbol{\psi}}_3 \\ s_{z_q} \{ \frac{1}{2} \mathbf{F}_3 \hat{\mathbf{T}} \hat{\mathbf{A}}_\phi (\mathbf{I}_n - \mathbf{W}^c) (\hat{\mathbf{D}}_{(v)} \hat{\mathbf{D}}_{(\tilde{w})}^2 - \mathbf{I}_n) + \hat{\alpha}_q \mathbf{Z}^\top \hat{\boldsymbol{\psi}}_4 \} + s_{x_p} \hat{\beta}_p \mathbf{Z}^\top \hat{\boldsymbol{\psi}}_3 \end{bmatrix},$$

em que  $\mathbf{F}_2$  é uma matriz  $M_1 \times n$  de zeros com uns na  $p$ -ésima linha,  $\mathbf{F}_3$  é uma matriz  $M_2 \times n$  de zeros com uns na  $q$ -ésima linha.

### ***Perturbação simultânea nas variáveis explanatórias da componente discreta e contínua***

Agora, considere o esquema em que é adicionado um vetor de perturbação a uma variável explanatória particular da componente discreta, considere-se também que é perturbado uma variável explanatória da componente contínua associado a componente  $\eta$  assim também, a componente  $\phi$ , i.e,

$$x_{ip\omega} = x_{ip} + \omega_i s_{x_p},$$

$$z_{iq\omega} = z_{iq} + \omega_i s_{z_q},$$

$$d_{ij\omega} = d_{ij} + \omega_i s_{d_j},$$

para qualquer  $p = 1, \dots, M_1$ ,  $q = 1, \dots, M_2$  e  $j = 1, \dots, M_3$ . Contudo, neste caso se considera que  $x_p \neq z_q \neq d_j$ . O logaritmo da função de verossimilhança assume a forma

$$\ell(\boldsymbol{\theta}) = \ell_{(1)}(\boldsymbol{\gamma}|\omega) + \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\omega),$$

Depois de algumas manipulações, avaliando no vetor de não perturbação  $\omega_0 = (0, \dots, 0)^\top$  e em  $\hat{\boldsymbol{\theta}}$  obtemos

$$\Delta_\theta = \begin{bmatrix} \Delta_\beta \\ \Delta_\alpha \\ \Delta_\gamma \end{bmatrix} = \begin{bmatrix} s_{x_p} \{ \mathbf{F}_2 \hat{\mathbf{T}}^{1/2} \hat{\mathbf{D}}_{(v)} \hat{\mathbf{Q}} (\mathbf{I}_n - \mathbf{W}^c) \hat{\mathbf{D}}_{(\tilde{w})} + \hat{\beta}_p \mathbf{X}^\top \hat{\boldsymbol{\psi}}_2 \} + s_{z_q} \hat{\alpha}_q \mathbf{X}^\top \hat{\boldsymbol{\psi}}_3 \\ s_{z_q} \{ \frac{1}{2} \mathbf{F}_3 \hat{\mathbf{T}} \hat{\mathbf{A}}_\phi (\mathbf{I}_n - \mathbf{W}^c) (\hat{\mathbf{D}}_{(v)} \hat{\mathbf{D}}_{(\tilde{w})}^2 - \mathbf{I}_n) + \hat{\alpha}_q \mathbf{Z}^\top \hat{\boldsymbol{\psi}}_4 \} + s_{x_p} \hat{\beta}_p \mathbf{Z}^\top \hat{\boldsymbol{\psi}}_3 \\ s_{d_j} \{ \mathbf{F}_1 \hat{\mathbf{P}} \hat{\mathbf{A}}_\pi \hat{\boldsymbol{\epsilon}} + \hat{\gamma}_j \mathbf{D}^\top \hat{\boldsymbol{\psi}}_1 \} \end{bmatrix},$$

### 3.2.4 Alavancagem Generalizada

A ideia principal de pontos de alavanca é avaliar a influência de  $w_i$  sobre o próprio valor predito (veja, por exemplo, Welsh, 1978; Cook e Weisberg, 1982; Laurent e Cook, 1992; Wei, Hu e Fung, 1998; Galea, Paula e Cysneiros, 2005). Uma forma de medir essa influência é por meio da derivada  $\partial\hat{w}_i/\partial w_i$ , que coincide com  $h_{ii}$  no caso normal linear, em que  $h_{ii}$  representa a  $i$ -ésimo elemento da diagonal de projeção, conhecida como matriz hat. Wei, Hu e Fung (1998) propuseram uma simples mas poderoso expressão para  $\partial\hat{w}_i/\partial w_i$ , desenvolvendo uma forma explícita para alavancagem.

Seja o vetor  $\mathbf{w} = (w_1, \dots, w_n)^\top$  de respostas observadas com densidade de probabilidade  $f_W(\mathbf{w}, \boldsymbol{\theta})$ , definida em (2.2). Denotamos por  $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\mathbf{w})$  a estimativa de máxima verossimilhança de  $\boldsymbol{\theta}$  e por  $\boldsymbol{\mu}$  o vetor de valores esperados, então  $\hat{\mathbf{w}} = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$  é o vetor de respostas preditas. Para o nosso caso, pode haver interesse em três medidas de pontos de alavanca, a influência de  $w_i$  em  $\hat{\pi}_i$  representada por  $\frac{\partial\hat{\pi}_i}{\partial w_i}$ , a influência de  $w_i$  em  $\hat{\eta}_i$  representada por  $\frac{\partial\hat{\eta}_i}{\partial w_i}$  e a influência de  $w_i$  em  $\hat{\phi}_i$  representada por  $\frac{\partial\hat{\phi}_i}{\partial w_i}$ . Pela propriedade da separabilidade dos parâmetros  $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$  e  $\boldsymbol{\gamma}$ , podemos obter a matriz de medidas de alavancagem generalizada de pontos de alavanca  $\frac{\partial\hat{\pi}_i}{\partial w_i}$  dada por

$$GL^\pi(\hat{\boldsymbol{\gamma}}) = \{\mathbf{D}_\gamma^\pi(-\ddot{L}_{\boldsymbol{\gamma}\boldsymbol{\gamma}})^{-1}\ddot{L}_{\boldsymbol{\gamma}\mathbf{w}^*}\}_{|\hat{\boldsymbol{\gamma}}} \quad (3.20)$$

em que  $\mathbf{D}_\gamma^\pi = \partial\boldsymbol{\pi}/\partial\boldsymbol{\gamma}^\top$ ,  $\ddot{L}_{\boldsymbol{\gamma}\mathbf{w}^*} = \partial^2\ell(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}\partial\mathbf{w}^{*\top}$  e  $\mathbf{w}^* = (I_{\{0\}}(w_1), \dots, I_{\{0\}}(w_n))^\top$ . Observe que em (3.20) envolve só a componente discreta.

A matriz de medidas de alavancagem generalizada de pontos de alavanca  $\frac{\partial\hat{\eta}_i}{\partial w_i}$  e  $\frac{\partial\hat{\phi}_i}{\partial w_i}$  pode ser expressa como

$$GL^\eta(\hat{\boldsymbol{\theta}}_1) = \{\mathbf{D}_{\boldsymbol{\theta}_1}^\eta(-\ddot{L}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1})^{-1}\ddot{L}_{\boldsymbol{\theta}_1\mathbf{w}}\}_{|\hat{\boldsymbol{\theta}}_1}, \quad GL^\phi(\hat{\boldsymbol{\theta}}_1) = \{\mathbf{D}_{\boldsymbol{\theta}_1}^\phi(-\ddot{L}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1})^{-1}\ddot{L}_{\boldsymbol{\theta}_1\mathbf{w}}\}_{|\hat{\boldsymbol{\theta}}_1},$$

respectivamente, em que  $\mathbf{D}_{\boldsymbol{\theta}_1}^\eta = (\mathbf{D}_\beta^\eta, \mathbf{D}_\alpha^\eta) = (\partial\boldsymbol{\eta}/\partial\boldsymbol{\beta}^\top, \partial\boldsymbol{\eta}/\partial\boldsymbol{\alpha}^\top)$ ,  $\mathbf{D}_{\boldsymbol{\theta}_1}^\phi = (\mathbf{D}_\beta^\phi, \mathbf{D}_\alpha^\phi) = (\partial\boldsymbol{\phi}/\partial\boldsymbol{\beta}^\top, \partial\boldsymbol{\phi}/\partial\boldsymbol{\alpha}^\top)$ ,  $\ddot{L}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} = \partial^2\ell(\boldsymbol{\theta}_1)/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^\top$  definida em (3.5) e  $\ddot{L}_{\boldsymbol{\theta}_1\mathbf{w}} = \partial^2(\boldsymbol{\theta}_1)/\partial\boldsymbol{\theta}_1\partial\mathbf{w}^\top$ .

Segundo o Apêndice C, matricialmente temos  $\mathbf{D}_\gamma^\pi = \boldsymbol{\Delta}_\pi\mathbf{D}$ ,  $\mathbf{D}_{\boldsymbol{\theta}_1}^\eta = (\mathbf{Q}(\mathbf{I}_n - \mathbf{W}^c)\mathbf{X}, \mathbf{0})$ ,  $\mathbf{D}_{\boldsymbol{\theta}_1}^\phi = (\mathbf{0}, \boldsymbol{\Delta}_\phi(\mathbf{I}_n - \mathbf{W}^c)\mathbf{Z})$

$$\ddot{L}_{\boldsymbol{\gamma}\mathbf{w}} = \mathbf{D}^\top\boldsymbol{\Phi}_1, \quad \ddot{L}_{\boldsymbol{\theta}_1\mathbf{w}} = \begin{bmatrix} \ddot{L}_{\beta\mathbf{w}} \\ \ddot{L}_{\alpha\mathbf{w}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\boldsymbol{\Phi}_2 \\ \mathbf{Z}^\top\boldsymbol{\Phi}_3 \end{bmatrix},$$

sendo  $\Delta_\pi$ ,  $\mathbf{Q}$  e  $\Delta_\phi$  definidas em (3.1.1),  $\Phi_1$ ,  $\Phi_2$  e  $\Phi_3$  são matrizes diagonais em que cada elemento  $i$  da respectiva diagonal para  $i = 1, \dots, n$ , é dado por

$$\begin{aligned}\Phi_{1i} &= \frac{1}{\pi_i(1-\pi_i)} \left( \frac{\partial \pi_i}{\partial \xi_{3i}} \right), \\ \Phi_{2i} &= (1 - I_{\{0\}}(w_i)) \left[ \frac{1}{w_i \phi_i} \left( \frac{\partial \eta_i}{\partial \xi_{1i}} \right) (\mathbf{v}_i + \mathbf{v}'_i \tilde{w}_i) \right], \\ \Phi_{3i} &= (1 - I_{\{0\}}(w_i)) \left[ \frac{1}{w_i \sqrt{\phi_i}} \left( \frac{\partial \phi_i}{\partial \xi_{2i}} \right) [\mathbf{v}_i \tilde{w}_i + \mathbf{v}'_i \tilde{w}_i^2 / 2] \right].\end{aligned}$$

### 3.3 APLICAÇÃO

O objetivo deste capítulo é apresentar e ilustrar como os modelos lineares zero ajustado log-simétricas se aplicam a um conjunto de dados reais. Para obter as estimativas dos parâmetros foi usado as extensões do pacote GAMLSS implementado pelo autor no software R.

#### Descrição do banco de dados

O banco de dados foi obtido através de um estudo longitudinal denominado Young Lives, que trata sobre aspectos sócio econômicos tais como gastos em saúde, educação, qualidade de vida, entre outros para identificar as causas ou consequências da pobreza infantil realizado nos países de Etiópia, Índia, Peru e Vietnã. Foram realizadas 5 rodadas até o momento (2002, 2006, 2009, 2013 e 2016) nos quais estão disponíveis no repositório UK Data Service. Cada rodada teve como finalidade estudar o mesmo grupo de crianças através de uma série de entrevistas baseadas em questionários.

A finalidade da presente aplicação é de identificar as variáveis que influenciam com os gastos em educação dos adolescentes que residem no Peru. A idade frequente para um aluno terminar o ensino secundário no Peru é de 16 anos, assim, foi considerado a rodada 3 (Boyden, 2018) já que a idade dos adolescentes foi de 14 a 15 anos, sendo uma idade frequente e obrigatória para estudar. As rodadas 4 e 5 consideram jovens maiores de 18 anos em que não necessariamente tem a obrigação de estudar. Além disso, pela natureza dos dados, não foi considerado os estudos nos anos 2002 e 2006, rodada 1 e 2, já que requiere-se modelagem para dados longitudinais sendo uma diferente metodologia dos modelos propostos. Inicialmente, consideramos uma amostra de 678 adolescentes e realizamos um pré-processamento de dados

tal como remover as observações faltantes, resultando em uma amostra final de 661 indivíduos. A variável resposta de interesse é definida como gastos totais em educação sendo a soma dos gastos diretos e indiretos destinado à educação de adolescentes. Vale a pena dizer que os valores zeros indicam que o adolescente não apresentou despesas em educação no momento do estudo.

As variáveis que foram consideradas para o estudo são

1. Resposta ( $y$ ): gasto total (diretos e indiretos) individual da pessoa em educação no ano 2009 em milhares de soles.;
2. IM: índice sobre a qualidade da moradia do adolescente;
3. IC: índice sobre o consumo de bens duráveis da moradia do adolescente;
4. ME: número de pessoas menores de 18 anos que moram na mesma moradia e que estudam (sem considerar o indivíduo da amostra);
5. Sexo: 0 = masculino, 1 = feminino;
6. CA: se um adolescente estuda na capital 0 = sim, 1 = não;
7. AE: número de anos de estudo formal do adolescente até o ano 2009.

### **Análise exploratória de dados**

Temos 661 adolescentes com 53% sendo do sexo masculino, 75% estudam na capital e tem entre 14 e 15 anos de idade. 49 adolescente não tiveram gastos em educação no ano 2009, representando 7,41%. Na Tabela 9 encontra-se algumas medidas descritivas das variáveis contínuas utilizadas na análise. Note-se que a variável resposta apresenta um alto grau de curtose, uma forte assimetria e alguns valores atípicos (veja a Figura 10).

Tabela 9 – Medidas de resumo das variáveis contínuas.(IIQ = Intervalo interquartil)

Medida	$y > 0$	IM	IC	ME	AE
Mínimo	0,050	0,004	0,000	0,000	2,000
Quartil 1	1,172	0,270	0,333	0,000	8,000
Mediana	2,050	0,510	0,500	1,000	9,000
Média	3,699	0,460	0,464	1,207	8,649
Quartil 3	3,250	0,763	0,667	2,000	10,000
Máximo	49,000	0,884	1,000	6,000	11,000
Desvio padrão	5,466	0,237	0,217	1,078	1,396
IIQ	2,078	0,493	0,333	2,000	2,000
Assimetria	3,728	0,058	-0,097	0,868	-1,391
Curtose	20,018	1,876	2,155	3,660	5,634

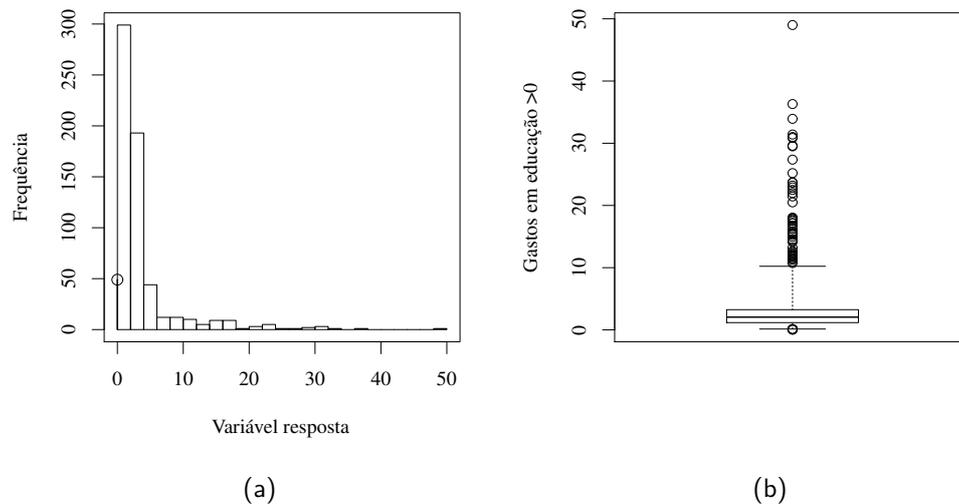


Figura 10 – Gráfico de frequência (a) e boxplot ajustado (b) aos gastos totais em educação em adolescentes do Peru no ano 2009.

Na Figura 11 apresentam os box-plots dos gastos totais em educação por categoria das demais variáveis explicativas não contínuas. Na Figura 11(a) observa-se que o valor máximo do gasto em educação está presente no sexo masculino, além disso o valor mediano é similar para ambos sexos. Na Figura 11(b) é possível notar que para as pessoas que moram na capital, apresentam maiores gastos em educação e maior variabilidade em comparação com uma pessoa que não mora na capital. Destaca-se a Figura 11(c), correspondente à variável ME, quanto maior o número de menores que estudam na moradia, menor é o gasto em educação no

adolescente; observamos também, na Figura 11(d), que quanto maior os anos de escolaridade maior os gastos em educação e maior variabilidade.

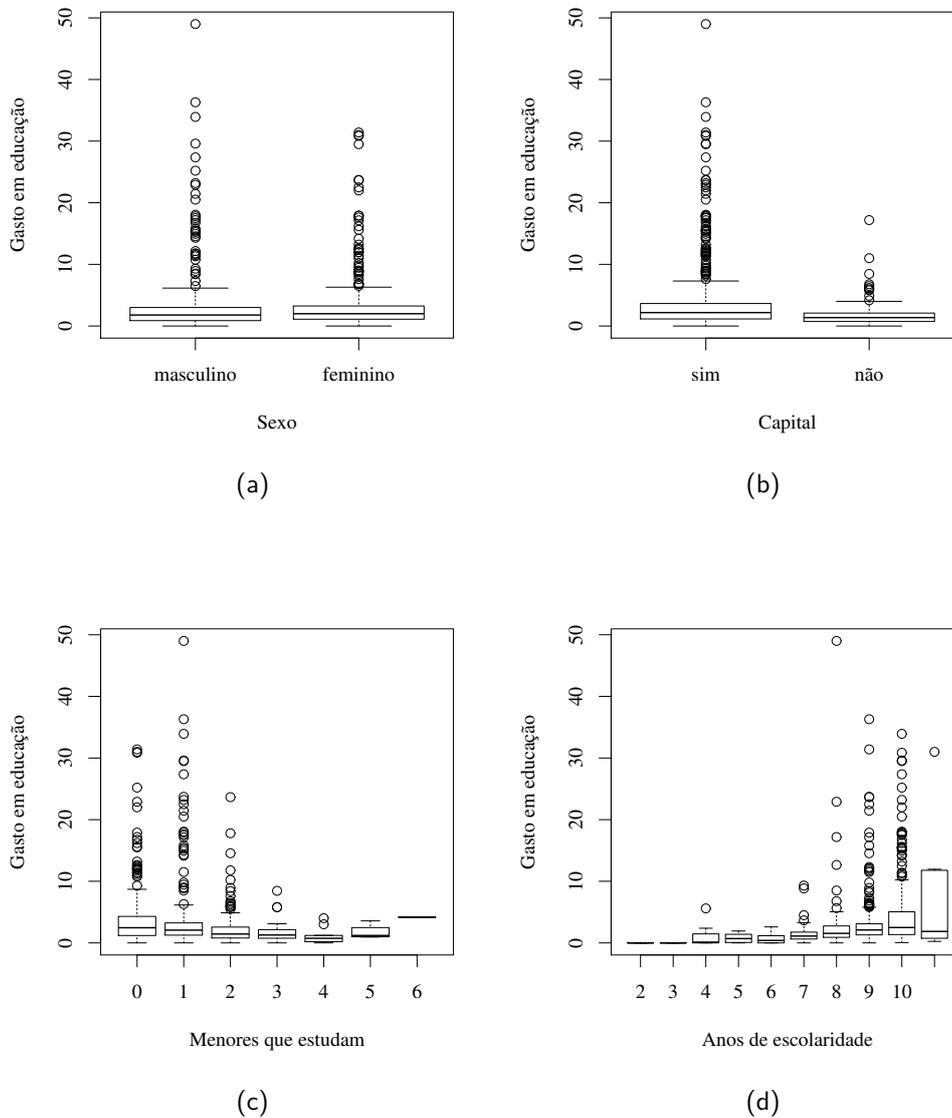


Figura 11 – Boxplots para gastos em educação de acordo ao sexo (a), capital (b), menores que estudam (c) e anos de escolaridade (d).

A Figura 12 apresenta os gráficos de dispersão dos gastos com educação por índice de moradia e de consumo. Podemos notar um aumento nos gastos quando o índice de moradia e o índice de consumo aumentam, além disso também há um aumento da variabilidade por faixa observada.

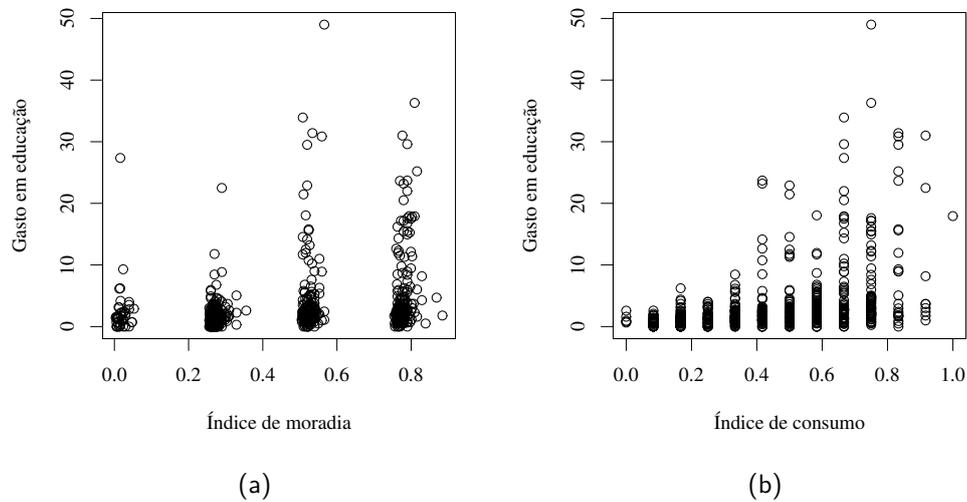


Figura 12 – Diagramas de dispersão em relação aos gastos em educação de acordo ao índice de moradia (a) e índice de consumo (b).

### Ajuste do modelo

Dado a natureza da variável resposta com um comportamento assimétrico e a presença de zeros consideramos cinco modelos de regressão, zero ajustado gama (ZAGA), zero ajustado gaussiana inversa (ZAIG), zero ajustado log-normal (ZALN), zero ajustado log-t-Student (ZALSt) e zero ajustado log-exponencial-potência (ZALPE) para explicar a relação com as variáveis explicativas. A metodologia a seguir para o modelagem consistiu em duas etapas uma que envolve a parte discreta ( $\pi$ ) e outra parte contínua ( $\eta$  ou  $\mu$ ,  $\sigma$  ou  $\phi$ ). Estas partes foram feitas independentes devido a propriedade da separabilidade. Em ambas partes descrevemos a estrutura sistemática da componente mediante o critério AIC de seleção de variáveis.

Foi usada a parametrização de casela de referência para as variáveis do tipo categóricas, tomando como referência ao sexo feminino e a uma pessoa que mora na capital. Em uma primeira etapa, modelamos a componente discreta e seguinte modelamos a parte contínua.

#### Modelo para a componente discreta

O submodelo para a componente discreta tem como objetivo explicar a probabilidade de um aluno ter efetuado gastos em educação em função de algumas variáveis explanatórias. Como o submodelo para  $\pi$  é estimado separadamente dos parâmetros correspondentes a componente contínua, as estimativas de máxima verossimilhança para o submodelo  $\pi$  será o mesmo para todos os modelos ajustados. O submodelo modelo para a parte discreta é dada por

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \xi_{3i},$$

em que

$$\xi_{3i} = \gamma_0 + \gamma_1 \text{IM}_i + \gamma_2 \text{IC}_i + \gamma_3 \text{ME}_i + \gamma_4 \text{Sexo}_i + \gamma_5 \text{CA}_i + \gamma_6 \text{AE}_i,$$

$\text{Sexo}_i$  assume zero para um indivíduo  $i$  com sexo masculino ou um para o sexo feminino e  $\text{CA}_i$  assume zero para uma indivíduo  $i$  que mora na capital ou um caso contrário para  $i = 1, \dots, 661$ . A Tabela 10 apresenta as estimativas de máxima verossimilhança para a componente discreta.

Tabela 10 – Estimativas de máxima verossimilhança para a componente discreta ajustado aos dados de gastos em educação.

	Parâmetro						
	$\gamma_0$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\gamma_5$	$\gamma_6$
Estimativa	6,649	-1,409	-1,170	-0,250	-0,394	-0,485	-0,967
Erro padrão	1,086	0,958	1,013	0,154	0,371	0,425	0,128

O critério AIC para seleção de variáveis sugeriu que a variável AE estivesse presente no modelo. Sendo assim, consideramos o teste da razão de verossimilhança para as hipótese  $H_0 : \gamma_* = 0$  contra  $H_1 : \text{pelo menos } \gamma_j \neq 0, j = 1, 2, 3, 4, 5$  em que  $\gamma_* = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)^\top$ , o valor do teste deu  $\xi_{RV} = 7,385$  com um  $p$ -valor de 0,194, assim no modelo final só a variável AE foi incluído para a componente discreta. A Tabela 11 apresenta as estimativas do modelo final para a componente discreta.

Tabela 11 – Estimativas de máxima verossimilhança para a componente discreta com as variáveis selecionadas ajustadas aos dados de gastos em educação.

	Parâmetro	
	$\gamma_0$	$\gamma_6$
Estimativa	5,283	-1,001
Erro padrão	0,841	0,115

### Modelo para a componente contínua

Para a componente contínua, iniciamos a modelagem do parâmetro de localização  $\mu$  para os modelos ZAGA e ZAIG e  $\eta$  para os modelos da classe log-simétrica através da função de ligação log em que  $\log(\cdot) = \xi_{1i}$ . Os preditores lineares para o correspondente parâmetro é dado por:

$$\xi_{1i} = \beta_0 + \beta_1 IM_i + \beta_2 IC_i + \beta_3 ME_i + \beta_4 \text{Sexo}_i + \beta_5 CA_i + \beta_6 AE_i,$$

para  $i = 1, \dots, 661$ . Para o caso de distribuições ZALPE e ZALSt, os parâmetros extras foram considerado fixo mas desconhecido. Para a escolha desse parâmetro consideramos uma faixa de valores e selecionamos o valor do parâmetro associado ao modelo com menor valor de AIC, sendo que para o modelo ZALSt foi considerado um intervalo de valores  $\nu = [3; 15]$  e para o modelo ZALPE foi considerado  $\nu = [-0,9; 0,9]$ . O valor do parâmetro extra selecionado foi de  $\nu = 4$  e  $\nu = 0,8$  para o modelo de regressão ZALSt e para a ZALPE, respectivamente. A Tabela 12 apresenta as estimativas de máxima verossimilhança para a componente contínua dos modelos considerados.

Tabela 12 – Estimativas de máxima verossimilhança (erro padrão) para o submodelo da componente contínua  $\eta$  e  $\mu$  dos modelos ajustados aos dados de gastos em educação.

Parâmetro	Modelo				
	ZAGA	ZAIG	ZALN	ZALSt	ZALPE
$\beta_0$	-0,829 (0,275)	-0,907 (0,291)	-0,781 (0,301)	-0,606 (0,265)	-0,612 (0,240)
$\beta_1$	0,454 (0,167)	0,402 (0,219)	0,349 (0,181)	0,291 (0,159)	0,359 (0,138)
$\beta_2$	2,029 (0,195)	1,917 (0,254)	1,686 (0,211)	1,497 (0,185)	1,371 (0,161)
$\beta_3$	-0,102 (0,031)	-0,060 (0,032)	-0,106 (0,035)	-0,125 (0,031)	-0,113 (0,030)
$\beta_4$	-0,044 (0,067)	-0,039 (0,086)	0,079 (0,071)	0,060 (0,063)	0,037 (0,054)
$\beta_5$	0,001 (0,092)	-0,038 (0,105)	0,074 (0,097)	0,067 (0,085)	0,040 (0,074)
$\beta_6$	0,104 (0,030)	0,116 (0,033)	0,072 (0,033)	0,068 (0,029)	0,072 (0,025)

Pode-se observar a presença de parâmetros não significativas em todos os modelos apresentados, do qual, se observou que em comum os parâmetros  $\beta_4$  e  $\beta_5$  correspondentes as variáveis Sexo e CA não foram significativas a um nível de 10% de significância. O critério AIC sugeriu que só as variáveis Sexo e CA não devem estar presente nos cinco modelos de regressão ajustados. A Tabela 13 apresenta as estimativas dos parâmetros para o submodelo final da componente contínua  $\eta$  e  $\mu$ .

Tabela 13 – Estimativas de máxima verossimilhança (erro padrão) para o submodelo final da componente contínua  $\eta$  e  $\mu$  dos modelos ajustados aos dados de gastos em educação.

Parâmetro	Modelo				
	ZAGA	ZAIG	ZALN	ZALSt	ZALPE
$\beta_0$	-0,944 (0,240)	-1,058 (0,295)	-0,825 (0,258)	-0,570 (0,232)	-0,585 (0,203)
$\beta_1$	0,433 (0,160)	0,435 (0,215)	0,312 (0,173)	0,300 (0,151)	0,304 (0,138)
$\beta_2$	1,991 (0,173)	1,996 (0,227)	1,584 (0,188)	1,398 (0,163)	1,359 (0,151)
$\beta_3$	-0,077 (0,028)	-0,056 (0,033)	-0,098 (0,032)	-0,118 (0,028)	-0,107 (0,025)
$\beta_6$	0,114 (0,028)	0,124 (0,034)	0,090 (0,030)	0,074 (0,027)	0,076 (0,026)

Em seguida ajustamos o submodelo para o parâmetro  $\psi$ , em que  $\psi = \phi$  para os modelos ZALN, ZALSt e ZALPE e  $\psi = \sigma$  para ZAGA e ZAIG. Vale a pena dizer que a modelagem de  $\psi$  envolve superdispersão,  $\psi$  é encarregado de modelar a dispersão relativa para os modelos da classe log-simétrica e dispersão para ZAGA e ZAIG. Temos que submodelo completo é dado por

$$\log(\psi_i) = \xi_{2i} = \alpha_0 + \alpha_1 IM_i + \alpha_2 IC_i + \alpha_3 ME_i + \alpha_4 \text{Sexo}_i + \alpha_5 CA_i + \alpha_6 AE_i,$$

Inicialmente, verificamos a hipótese de dispersão variável, as hipóteses consideradas são  $H_0 : \alpha_* = 0$  contra  $H_1 : \text{pelo menos } \alpha_j \neq 0, j = 1, \dots, 6$  em que  $\alpha_* = (\alpha_1, \dots, \alpha_6)^T$ . A Tabela 14 apresenta os valores das estatísticas de razão de verossimilhança com seus respectivos  $p$ -valores do qual podemos observar que há indício, ao nível de significância de 1%, de dispersão variável nos cinco modelos de regressão.

Tabela 14 – Valor da estatística de razão de verossimilhança e  $p$ -valor para o respectivo teste de hipótese nos modelos de regressão ajustados ao dados de gastos em educação.

	ZAGA	ZAIG	ZALN	ZALSt	ZALPE
$\xi_{RV}$	41,19	36,00	37,57	35,79	26,12
$p$ -valor	< 0,01	< 0,01	< 0,01	< 0,01	< 0,01

A Tabela 15 apresenta as estimativas dos parâmetros com as variáveis selecionadas em cada modelo de regressão correspondente ao segundo parâmetro da componente contínua.

Tabela 15 – Estimativas de máxima verossimilhança (erro padrão) para o submodelo da componente contínua  $\phi$  e  $\sigma$  dos modelos ajustados aos dados de gastos em educação.

Parâmetro	Modelo				
	ZAGA	ZAIG	ZALN	ZALSt	ZALPE
$\alpha_0$	-0,376 (0,088)	0,054 (0,101)	-0,582 (0,184)	-2,652 (0,569)	-2,502 (0,202)
$\alpha_2$	0,514 (0,144)	-0,762 (0,160)	0,988 (0,306)	1,444 (0,377)	1,527 (0,359)
$\alpha_4$	-0,116 (0,053)	-0,225 (0,059)	-0,274 (0,115)	-0,316 (0,152)	-0,272 (0,154)
$\alpha_5$	-0,160 (0,075)	-0,179 (0,082)	-0,321 (0,153)	– –	– –
$\alpha_6$	– –	– –	– –	0,138 (0,068)	– –

### **Análise de resíduo**

As Figuras 13-17 apresentam diferentes gráficos para os resíduos do tipo quantílico para os modelos ZAGA, ZAIG, ZALN, ZALSt e ZALPE, respectivamente. Podemos ver que para os diferentes modelos ajustados não há evidencia de afastamento das suposições, mas podemos observar que para os valores ajustados contra o resíduo para os modelos ZAGA e ZAIG há uma ligeira tendência, uma variabilidade menor do resíduo a medida que os valores ajustados aumentam. Os resíduos para os modelos ZAGA, ZAIG e ZALN apresentam um comportamento leptocúrtica, já para os modelos ZALSt e ZALPE apresentam um comportamento mesocúrtico. Podemos observar nos gráficos normal de probabilidade que só no modelo ZAGA existe um ligeiro afastamento na cauda direita, já para os outros modelos, podemos ver que não apresentam nenhum comportamento inusual.

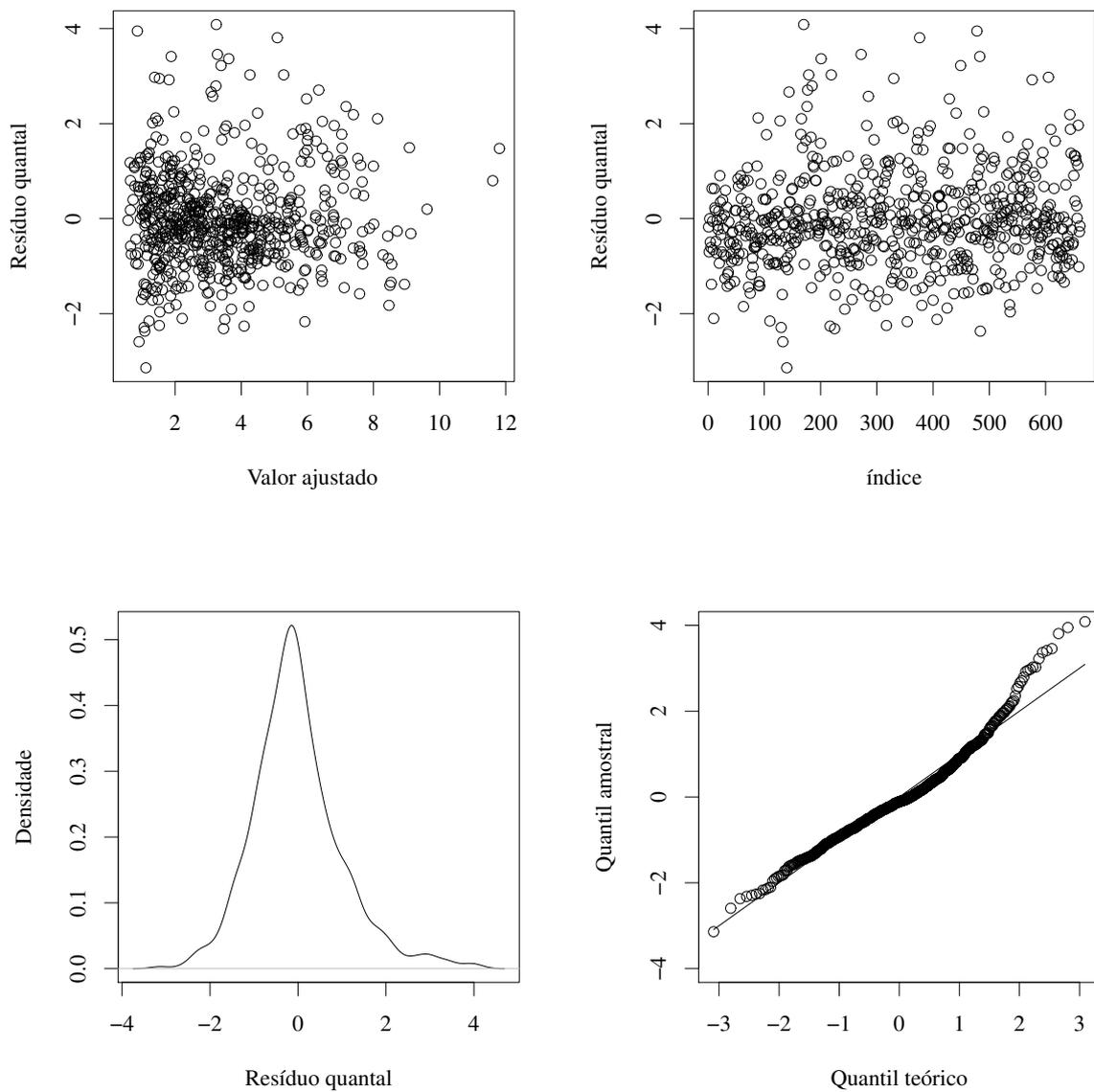


Figura 13 – Resíduo do tipo quantílico do modelo ZAGA ajustado aos dados de gastos em educação.

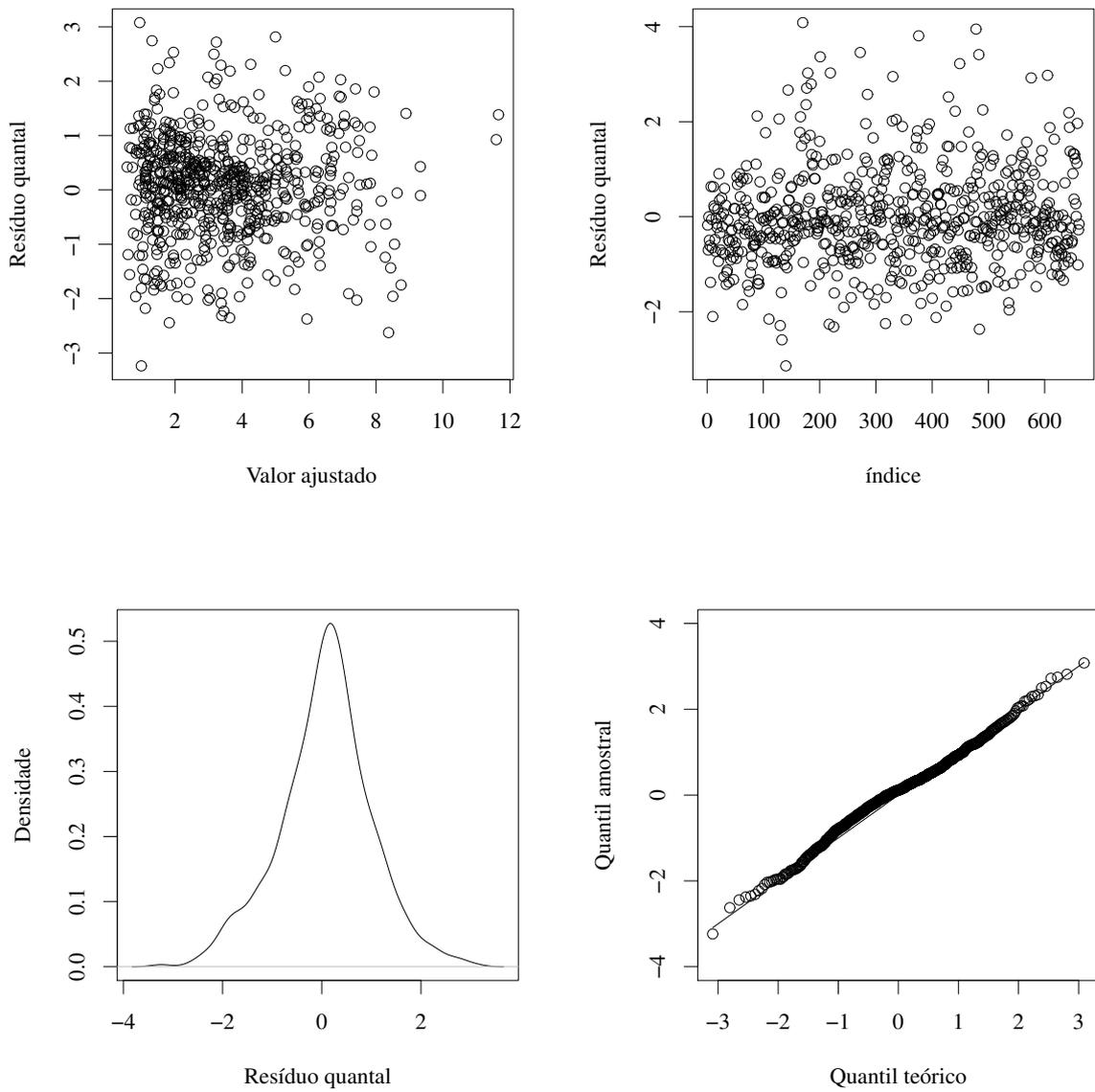


Figura 14 – Resíduo do tipo quantílico do modelo ZAIG ajustado aos dados de gastos em educação.

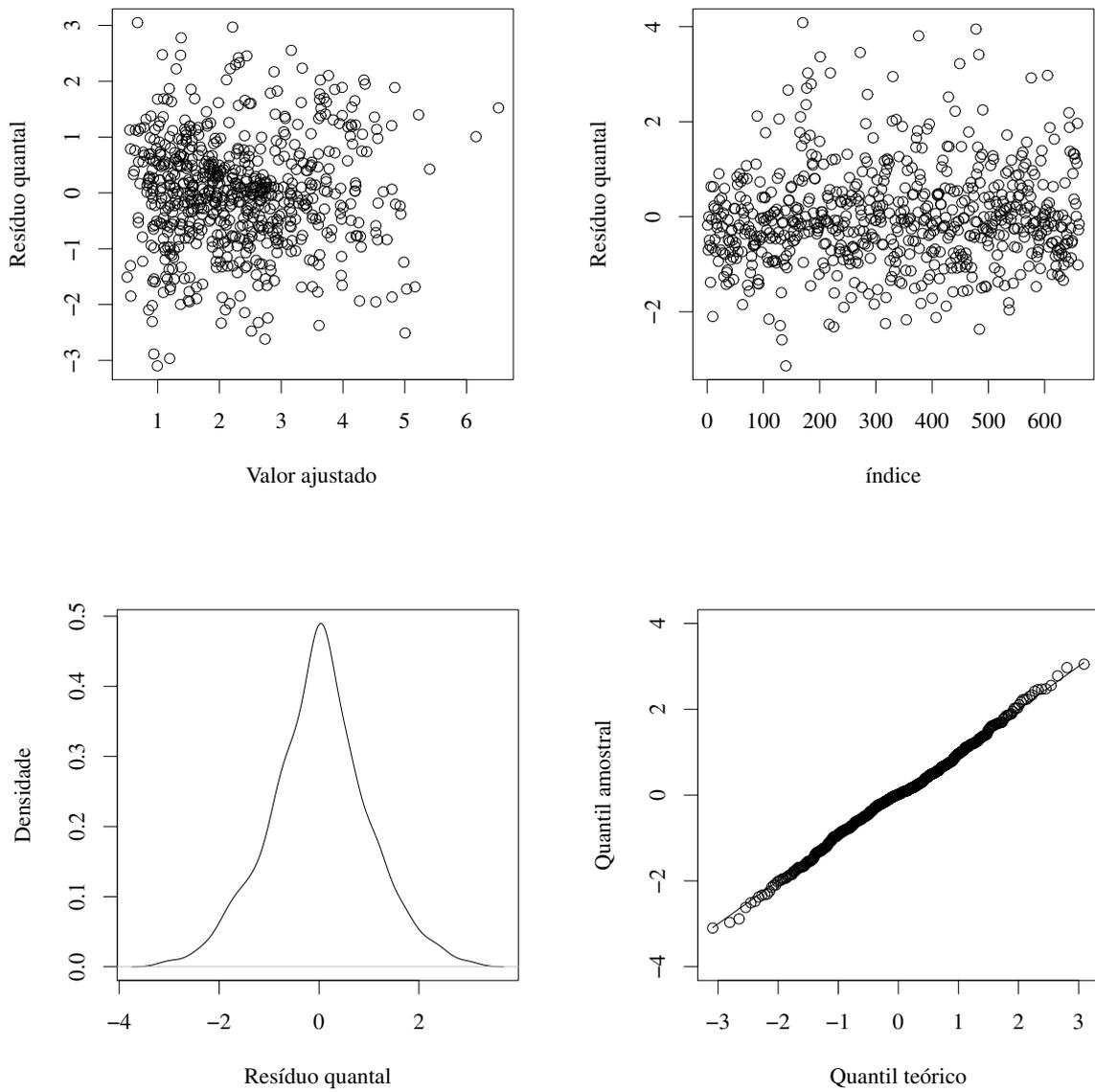


Figura 15 – Resíduo do tipo quantílico do modelo ZALN ajustado aos dados de gastos em educação.

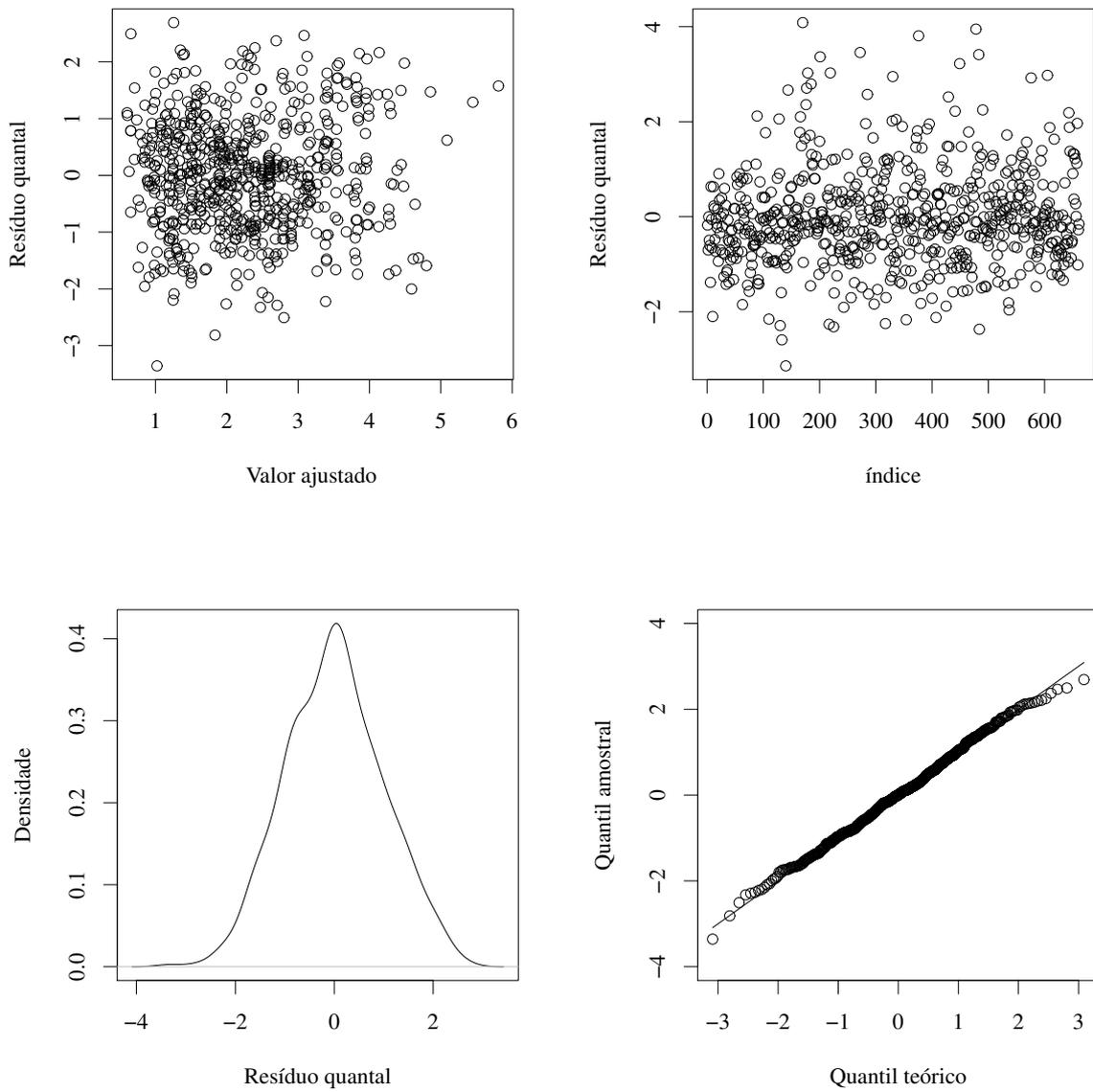


Figura 16 – Resíduo do tipo quantílico do modelo ZALSt ajustado aos dados de gastos em educação.

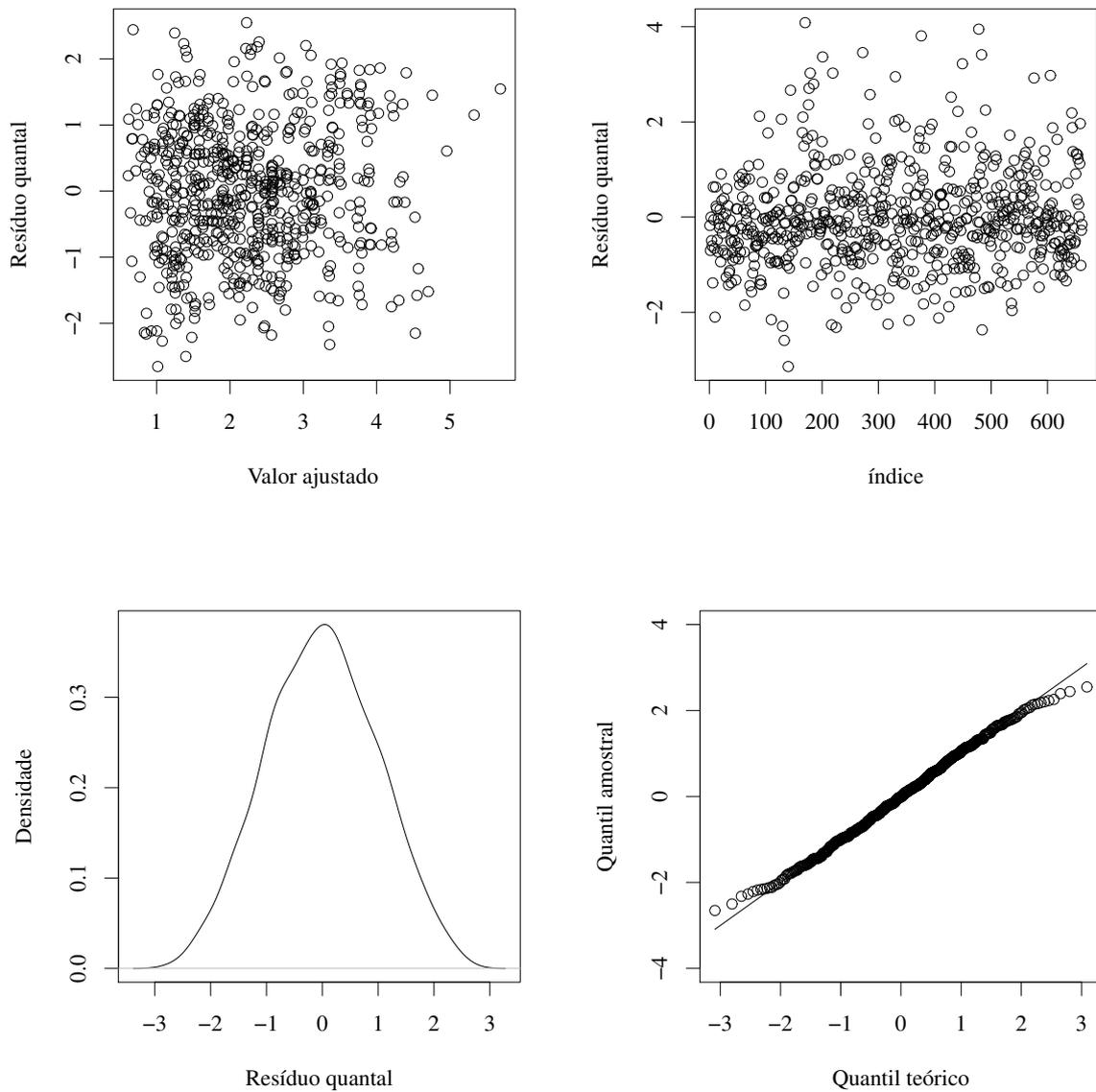


Figura 17 – Resíduo do tipo quantílico do modelo ZALPE ajustado aos dados de gastos em educação.

A tabela 16 apresenta algumas medidas de adequabilidade tais como tais como  $-2\log(\hat{L})$ , AIC, BIC e pseudo  $R^2$  para cada modelo ajustado podendo observar que o modelo ZALPE tem os menores valores para as medidas mencionadas e o segundo maior pseudo  $R^2$ , logo o modelo ZALPE foi o que se ajustou melhor aos dados de gastos em educação.

Tabela 16 – Valores do  $-2\log(\hat{L})$ , AIC e BIC para os modelos considerados ajustados aos dados de gastos em educação.

	Modelo de regressão				
	ZAGA	ZAIG	ZALN	ZALSt	ZALPE
$-2\log(\hat{L})$	2753,002	2792,975	2699,653	2661,195	<b>2659,105</b>
AIC	2775,002	2814,975	2721,653	2683,195	<b>2679,105</b>
BIC	2824,433	2864,407	2771,084	2732,626	<b>2724,043</b>
pseudo $R^2$	<b>0,471</b>	0,353	0,368	0,369	0,370

## Análise de resíduos e diagnóstico para o modelo selecionado

### Resíduos

O análise de resíduos proposta neste trabalho consistiu em avaliar os resíduos da componente contínua e discreta do modelo selecionado ZALPE sob o resíduo componente do desvio e o resíduo ZAR mediante diferentes gráficos. O resíduo componente do desvio teve como objetivo avaliar o ajuste individualmente em cada submodelo  $\pi$ ,  $\eta$  e  $\phi$ , já o resíduo ZAR permitiu avaliar o ajuste e identificar pontos atípicos na componente contínua associado ao ajuste do parâmetro  $\theta_1$ . Na Figura 18 é apresentado os valores ajustados contra os resíduos componente do desvio e também os gráficos normais de probabilidade com envelope para cada componente ajustado no modelo de regressão ZALPE. As Figuras 18(a), 18(c) e 18(e) correspondentes aos componentes  $\eta$ ,  $\phi$  e  $\pi$ , respectivamente, mostram que os resíduos não apresentaram nenhuma tendência sistemática e apresentaram uma aleatoriedade em torno do zero. Já nos gráficos normais de probabilidade com envelope indicam que não há evidência de afastamento das suposições para as três componentes do modelo ZALPE mas pode-se observar um ligeiro afastamento na cauda esquerda correspondente ao resíduo componente do desvio para o submodelo  $\eta$ . Por último, valores ajustados de  $\eta$  contra os resíduos e gráfico normal de probabilidade com envelope para o resíduo ZAR são apresentados na Figura 19, da mesma forma, observamos que não há evidência de afastamento das suposições e de pontos atípicos.

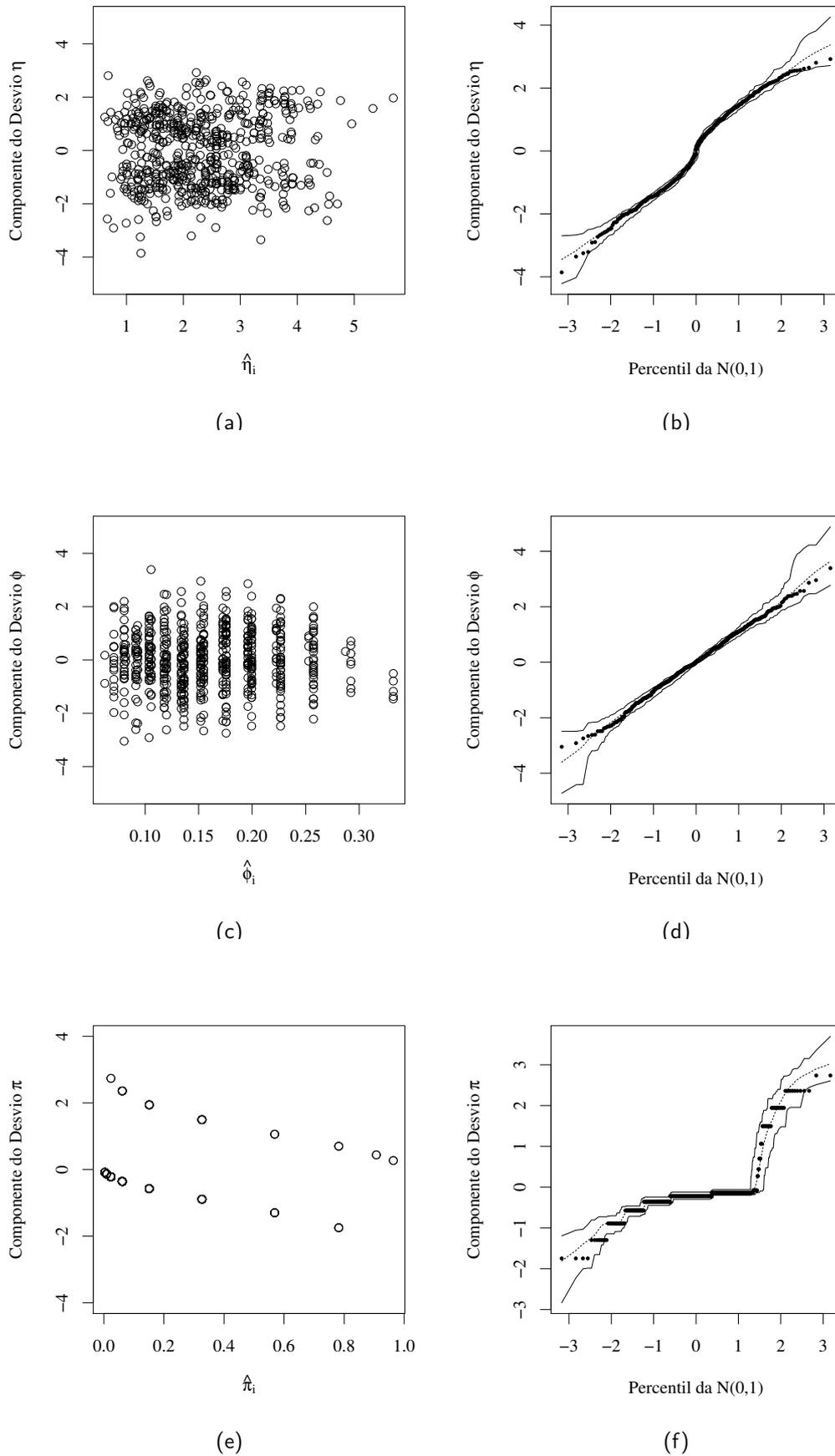


Figura 18 – Gráfico de valores ajustados contra resíduo componente do desvio (a), (c), (e) e gráfico normal de probabilidade com envelope (b), (d), (f) para a componente contínua e discreto ajustados aos dados de gastos em educação.

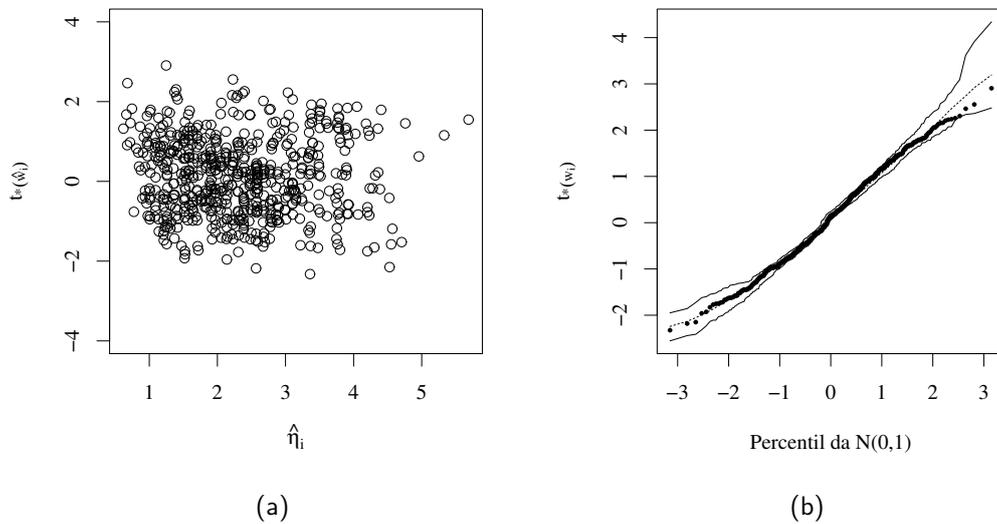


Figura 19 – Gráfico de valores ajustados contra resíduos do tipo ZAR (a) e gráfico normal de probabilidade com envelope (b) para a componente contínua ajustado aos dados de gastos em educação.

### ***Influência global e local***

Avaliamos a influência global mediante gráficos de índice contra  $LD(\hat{\theta}_{1(i)}^I)$ ,  $LD(\hat{\gamma}_{(i)}^I)$  e  $LD(\hat{\theta}_{(i)}^I)$  para os parâmetros da componente contínua, componente discreta e no geral, respectivamente, apresentados na Figura 20. Assim, o ponto #303 aparece como possivelmente influente nos parâmetros da componente contínua, os pontos #225, #400 e #407 aparecem como observações possivelmente influentes nos parâmetros da componente discreta, já no geral, notamos os pontos #225, #303 e #400. Por último, avaliamos a influência dos pontos nas estimativas de máxima verossimilhança dos parâmetros mediante o esquema de perturbação de casos ponderados para os parâmetros da componente discreta (Figura 21(a)) e componente contínua (Figura 21(b)). Considere-se como potencialmente pontos influentes aquelas que estejam com uma distância desproporcional dos demais. Verificamos que os pontos destacados nesse gráficos são os mesmos destacados na análise de influência global.

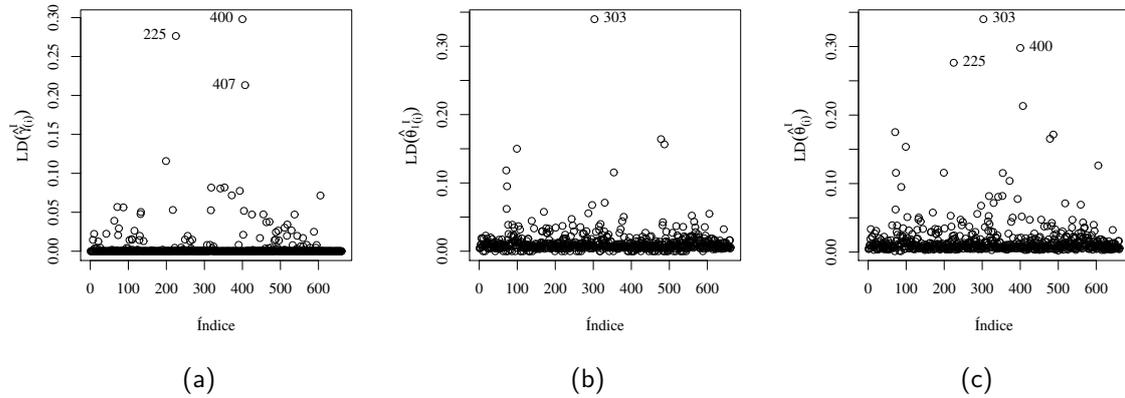


Figura 20 – Gráfico de influência global para a componente discreta (a), componente contínua (b) e geral (c) para o modelo ajustado aos dados dos gastos totais em educação em adolescentes do Peru no ano 2009.

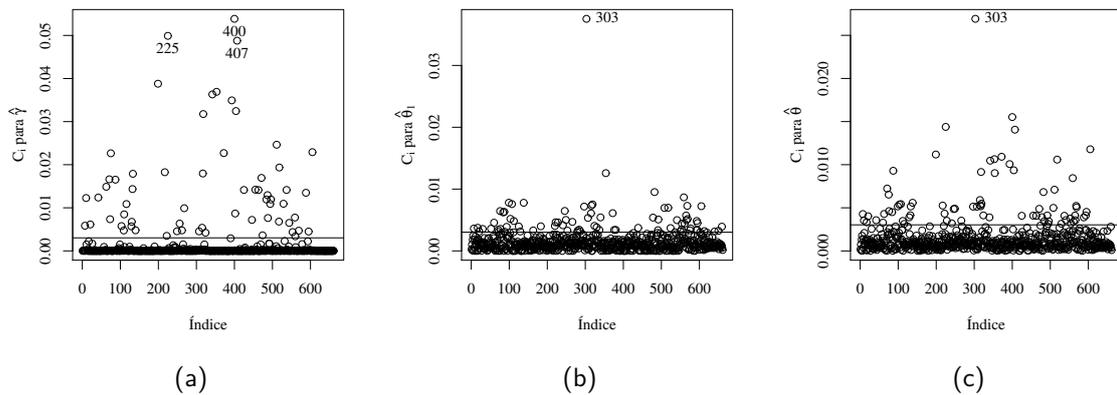


Figura 21 – Gráfico de influência local ponderação de casos para a componente discreta (a), componente contínua (b) e geral (c) para o modelo ajustado aos dados dos gastos totais em educação em adolescentes do Peru no ano 2009.

### ***Alavancagem generalizado***

Nas Figura 22(a), 22(b) e 22(c) é apresentado os pontos de alavancagem para o ajuste de  $\pi$ ,  $\eta$  e  $\phi$ , respectivamente. A partir dessas figura, nota-se que as observações #315, #538 e #303 aparecem com alta alavancagem. Observamos de novo a presença do ponto #303 para a componente contínua especificamente para o ajuste de  $\phi$ .

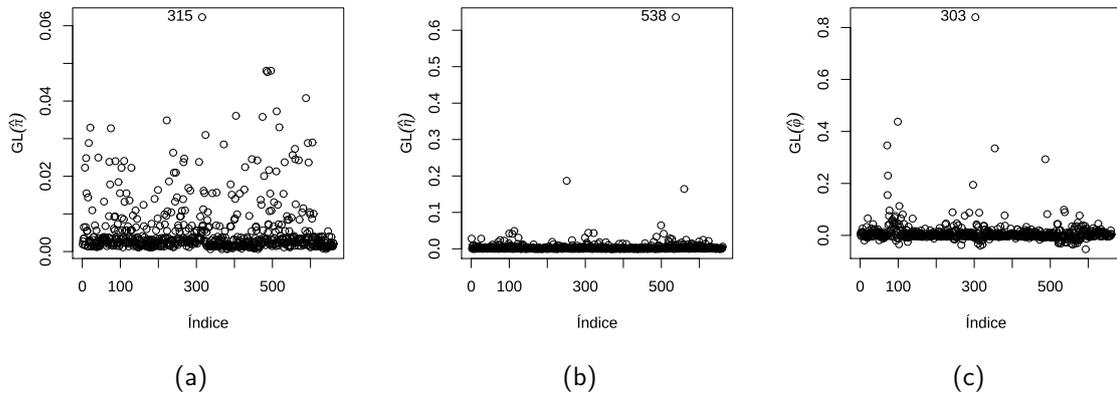


Figura 22 – Gráfico de pontos de alavanca generalizados para  $\pi$  (a),  $\eta$  (b) e  $\phi$  (c) para o modelo ajustado aos dados dos gastos totais em educação em adolescentes do Peru no ano 2009.

A seguir, A fim de verificar o impacto da retirada de cada ponto, ajustamos modelos sem a presença do ponto e verificamos a variação percentual dada por  $VR = \left( \frac{\hat{\theta} - \hat{\theta}^{(i)}}{\hat{\theta}} \right)$  em que  $\hat{\theta}$  e  $\hat{\theta}^{(i)}$  são as estimativas de máxima verossimilhança no modelo completo e no modelo sem a  $i$ -ésima observação. A Tabela 17 apresenta as estimativas e a variação percentual para os modelos ajustados sem as observações #225, #303 e #400 individual e coletivamente. Vale a pena dizer que os pontos #225 e #400 são observações possivelmente influentes para os parâmetros da componente discreta, assim, podemos observar que ao momento de retirar os pontos #225 e #400 os parâmetros correspondentes ao componente contínua não apresenta muita mudança sendo próximo a 0% (\* significa que o impacto porcentual em absoluta é  $< 0,001\%$ ), já para os parâmetros da componente discreta, as novas estimativas apresentam uma maior mudança porcentual, ainda mais ao retirar o ponto #225 sendo que para o parâmetro  $\gamma_6$  correspondente a variável AE tem uma mudança de 2,36% em relação a estimativa sem retirar a observação. As estimativas dos parâmetros  $\alpha_2$  e  $\alpha_4$  apresentam maior mudanças ao momento de retirar o ponto #303 com 8,384% e 14,07% de variação, respectivamente.

Tendo em vista que os pontos #225, #303 e #400 foram destacados em todas análises, apresentamos as características de cada um, a saber:

**#225:** É uma observação que se destaca para a componente discreta, o indivíduo tem um ano de escolaridade de 9 anos, representando o valor máximo quando o estudante não apresenta gastos educativos.

**#303:** É uma observação que se destaca para a componente contínua, assim observando a variável IM, o indivíduo tem um valor menor do que o primeiro quartil com um valor de 0,269;

acontece o mesmo para a variável IC, com um valor de 0,167. Para a variável ME o valor foi de 1, próximo a zero e para a variável AE apresenta um valor de 8, sendo o limite dos 25% menores.

**#400:** Também é uma observação que se destaca para a componente discreta, o indivíduo tem um ano de escolaridade de 8 anos, um valor próximo ao máximo quando o estudante não apresenta gastos educativos.

Tabela 17 – Estimativas dos parâmetros do modelo ZALPE final ajustado aos dados de gastos em educação e impacto porcentual (entre parênteses) após retirar as observações influentes consideradas.

Parâmetro	Retirando observação						
	#225	#303	#400	#225 e #303	#225 e #400	#303 e #400	#225, #303 e #400
$\beta_0$	-0,585 (*)	-0,584 (-0,159%)	-0,585 (*)	-0,584 (-0,159%)	-0,585 (*)	-0,584 (-0,159%)	-0,584 (-0,160%)
$\beta_1$	0,304 (*)	0,298 (-1,978%)	0,305 (*)	0,298 (-1,978%)	0,304 (*)	0,298 (-1,978%)	0,298 (-1,978%)
$\beta_2$	1,359 (*)	1,339	1,359 (*)	1,339	1,359 (*)	1,339	1,339 (-1,506%)
$\beta_3$	-0,107 (*)	-0,111 (3,949%)	-0,107 (*)	-0,111 (3,949%)	-0,107 (*)	-0,111 (3,949%)	-0,111 (3,949%)
$\beta_6$	0,076 (*)	0,079 (2,737%)	0,076 (*)	0,079 (2,737%) <sup>3</sup>	0,076 (*)	0,079 (2,737%)	0,079 (2,737%)
$\alpha_0$	-2,502 (*)	-2,570 (2,626%)	-2,502 (*)	-2,570 (2,626%)	-2,502 (*)	-2,570 (2,626%)	-2,570 (2,626%)
$\alpha_2$	1,527 (*)	1,666 (8,384%)	1,527 (*)	1,666 (8,384%)	1,527 (*)	1,666 (8,384%)	1,666 (8,384%)
$\alpha_4$	-0,272 (*)	-0,317 (14,07%)	-0,272 (*)	-0,317 (14,07%)	-0,272 (*)	-0,317 (14,07%)	-0,317 (14,07%)
$\gamma_0$	5,426 (2,635%)	5,281 (-0,055%)	5,329 (0,86%)	5,423 (2,579%)	5,478 (3,559%)	5,326 (0,804%)	5,475 (3,502%)
$\gamma_6$	-1,026 (2,368%)	-1,001 (-0,067%)	-1,012 (1,039%)	-1,025 (2,301%)	-1,037 (3,459%)	-1,011 (0,972%)	-1,036 (3,393%)

## 4 CONCLUSÕES

Na presente dissertação, inicialmente foi proposta uma nova classe de distribuições zero ajustada log-simétrica, sendo a estrutura da nova classe composta por uma componente discreta distribuída por uma bernoulli e uma componente contínua sendo esta última a classe log-simétrica. Foi apresentado a estimação pontual e intervalar dos parâmetros e um estudo de simulação para três distribuições pertencentes a classe, a saber, ZALN, ZALSt e ZALPE, sob diferentes cenários concluíram que os estimadores de máxima verossimilhança para os parâmetros da componente contínua chegam a ser afetados pela proporção de zeros, mas apresentam um bom desempenho para amostras grandes.

Na segunda etapa deste trabalho foi proposto uma classe de modelos de regressão linear bastante flexível. Foi obtido a função escore assim como a a matriz de informação de Fisher, as estimativas de máxima verossimilhança para os parâmetros foi desenvolvido mediante um processo iterativo de estimação. Foi apresentado 3 tipos de resíduos: componente de desvio (Pregibon, 1981), quantílico (Dunn e Smyth, 1996) e ZAR (Pereira et al., 2020). O resíduo componente de desvio, bastante utilizado para os modelos lineares generalizados e adaptado para este trabalho, avalia as suposições em cada componente a ser modelado. O resíduo quantílico avalia as suposições globalmente e o resíduo zero ajustado (ZAR) foi desenvolvido pelas limitações do resíduo quantílico para a detecção de pontos atípicos no componente contínuo. Outro ponto que representa uma importante contribuição foi propor medidas de diagnóstico sob o enfoque de influência global, local e alavancagem generalizada.

Por último, no contexto prático, foi discutido um exemplo com um conjunto de dados reais sobre despesas educativas em adolescentes entre 14 a 15 anos de idade no Peru, 2009. Foi ajustado cinco modelos de regressão a saber: ZAGA, ZAIG, ZALN, ZALSt e ZALPE; os modelos ZAGA e ZAIG. Selecionamos o modelo que ajustou-se melhor aos dados com base as medidas  $-2\log(\hat{L})$ , AIC e BIC. A análise de resíduos mostrou evidências que o modelo está adequado e o análise de diagnóstico evidenciou alguns pontos influentes em que os pontos #225 e #400 destacam para a componente discreta e o ponto #303 destaca para a componente contínua. O ponto #225 apresenta um valor máximo de anos de escolaridade, já o ponto #400 apresenta um valor de 8 anos de escolaridade, sendo bem próximo ao máximo. O ponto #303 em particular apresenta um índice de moradia menor do que o primeiro quartil, um valor de 8 anos de escolaridade pertencendo ao limite dos 25% menores.

## REFERÊNCIAS

- AITCHISON, J.; BROWN, J. A. The lognormal distribution with special reference to its uses in economics. Cambridge Univ. Press, 1957.
- ANDERSON, T.; FANG, K.-T. Theory and applications of elliptically contoured. 1990.
- BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. *Journal of applied probability*, JSTOR, p. 319–327, 1969.
- BOYDEN, J. *Young Lives: an International Study of Childhood Poverty: Round 3, 2009.[data collection]*. UK Data Service. SN: 6853. 2018.
- CASTELLACCI, G. A formula for the quantiles of mixtures of distributions with disjoint supports. *Available at SSRN 2055022*, 2012.
- COOK, R. D. Detection of influential observation in linear regression. *Technometrics*, Taylor & Francis Group, v. 19, n. 1, p. 15–18, 1977.
- COOK, R. D. Assessment of local influence. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 48, n. 2, p. 133–155, 1986.
- COOK, R. D.; WEISBERG, S. *Residuals and influence in regression*. New York: Chapman and Hall, 1982.
- CORDEIRO, G. M.; FERRARI, S. L.; URIBE-OPAZO, M. A.; VASCONCELLOS, K. L. Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics & probability letters*, Elsevier, v. 46, n. 4, p. 317–328, 2000.
- COX, D. R.; HINKLEY, D. V. *Theoretical statistics*. CRC Press, 1979.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 30, n. 2, p. 248–265, 1968.
- CYSNEIROS, F.; CORDEIRO, G.; CYSNEIROS, A. Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 80, n. 4, p. 451–461, 2010.
- CYSNEIROS, F. J. A.; PAULA, G. A.; GALEA, M. Heteroscedastic symmetrical linear models. *Statistics & probability letters*, Elsevier, v. 77, n. 11, p. 1084–1090, 2007.
- CYSNEIROS, F. J. A.; VANEGAS, L. H. Residuals and their statistical properties in symmetrical nonlinear models. *Statistics & Probability Letters*, Elsevier, v. 78, n. 18, p. 3269–3273, 2008.
- DUAN, N.; MANNING, W. G.; MORRIS, C. N.; NEWHOUSE, J. P. A comparison of alternative models for the demand for medical care. *Journal of business & economic statistics*, Taylor & Francis, v. 1, n. 2, p. 115–126, 1983.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996.
- ESCOBAR, L. A.; JR, W. Q. M. Assessing influence in regression analysis with censored data. *Biometrics*, JSTOR, p. 507–528, 1992.

- FANG, K. W. *Symmetric multivariate and related distributions*. Chapman & Hall, 1990.
- GALEA, M.; PAULA, G. A.; CYSNEIROS, F. J. A. On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters*, Elsevier, v. 73, n. 4, p. 459–467, 2005.
- HASHIMOTO, E. M.; ORTEGA, E. M.; CORDEIRO, G. M.; CANCHO, V. G.; KLAUBERG, C. Zero-spiked regression models generated by gamma random variables with application in the resin oil production. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 89, n. 1, p. 52–70, 2019.
- HELLER, G.; STASINOPOULOS, M.; RIGBY, B. et al. The zero-adjusted inverse gaussian distribution as a model for insurance claims. In: GALWAY. *Proceedings of the 21th International Workshop on Statistical Modelling*. , 2006. v. 226233.
- HOAGLIN, D. C.; WELSCH, R. E. The hat matrix in regression and anova. *The American Statistician*, Taylor & Francis Group, v. 32, n. 1, p. 17–22, 1978.
- JONES, M. On reciprocal symmetry. *Journal of Statistical Planning and Inference*, Elsevier, v. 138, n. 10, p. 3039–3043, 2008.
- KAI-TAI; YAO-TING, Z. *Generalized multivariate analysis*. Science Press Beijing and Springer-Verlag, Berlin, 1990. v. 19.
- LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, Taylor & Francis, v. 34, n. 1, p. 1–14, 1992.
- LAURENT, R. T. S.; COOK, R. D. Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 87, n. 420, p. 985–990, 1992.
- LAWLESS, J. F. *Statistical models and methods for lifetime data*. John Wiley & Sons, 2003. v. 362.
- LAWRANCE, A. J. Regression transformation diagnostics using local influence. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 83, n. 404, p. 1067–1072, 1988.
- LEHMANN, E. L.; CASELLA, G. *Theory of point estimation*. Springer Science & Business Media, 1998.
- LEIVA, V.; SANTOS-NETO, M.; CYSNEIROS, F. J. A.; BARROS, M. A methodology for stochastic inventory models based on a zero-adjusted birnbaum-saunders distribution. *Applied Stochastic Models in Business and Industry*, Wiley Online Library, v. 32, n. 1, p. 74–89, 2016.
- LESAFFRE, E.; VERBEKE, G. Local influence in linear mixed models. *Biometrics*, JSTOR, p. 570–582, 1998.
- MAIOR, V. Q.; CYSNEIROS, F. J. A. Symarma: a new dynamic model for temporal data on conditional symmetric distribution. *Statistical Papers*, Springer, v. 59, n. 1, p. 75–97, 2018.
- MARTIN, T. G.; WINTLE, B. A.; RHODES, J. R.; KUHNERT, P. M.; FIELD, S. A.; LOW-CHOY, S. J.; TYRE, A. J.; POSSINGHAM, H. P. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology letters*, Wiley Online Library, v. 8, n. 11, p. 1235–1246, 2005.

- MCCULLAGH, P. Tensor methods in statistics, 1987. *Chapman & Hall*.
- MEDEIROS, F. M.; FERRARI, S. L. Small-sample testing inference in symmetric and log-symmetric linear regression models. *Statistica Neerlandica*, Wiley Online Library, v. 71, n. 3, p. 200–224, 2017.
- MIN, Y.; AGRETI, A. Modeling nonnegative data with clumping at zero: a survey. *JOURNAL OF THE IRANIAN STATISTICAL SOCIETY (JIRSS)*, 2002.
- OSPINA, R.; FERRARI, S. L. Inflated beta distributions. *Statistical Papers*, Springer, v. 51, n. 1, p. 111, 2010.
- PACE, L.; SALVAN, A. *Principles of statistical inference: from a Neo-Fisherian perspective*. World scientific, 1997. v. 4.
- PAULA, G. A. Assessing local influence in restricted regression models. *Computational statistics & data analysis*, Elsevier, v. 16, n. 1, p. 63–79, 1993.
- PEREIRA, G. H.; SCUDILIO, J.; SANTOS-NETO, M.; BOTTER, D. A.; SANDOVAL, M. C. A class of residuals for outlier identification in zero adjusted regression models. *Journal of Applied Statistics*, Taylor & Francis, v. 47, n. 10, p. 1833–1847, 2020.
- PREGIBON, D. et al. Logistic regression diagnostics. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 9, n. 4, p. 705–724, 1981.
- PUIG, P. A note on the harmonic law: a two-parameter family of distributions for ratios. *Statistics & probability letters*, Elsevier, v. 78, n. 3, p. 320–326, 2008.
- QUEIROZ, F. F.; LEMONTE, A. J. A broad class of zero-or-one inflated regression models for rates and proportions. *Canadian Journal of Statistics*, Wiley Online Library, 2020.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 54, n. 3, p. 507–554, 2005.
- RODRIGUES-MOTTA, M.; SOTO, D. M. G.; LACHOS, V. H.; VILCA, F.; BALTAR, V. T.; JUNIOR, E. V.; FISBERG, R. M.; MARCHIONI, D. M. L. A mixed-effect model for positive responses augmented by zeros. *Statistics in medicine*, Wiley Online Library, v. 34, n. 10, p. 1761–1778, 2015.
- SU, L.; TOM, B. D.; FAREWELL, V. T. Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics*, Oxford University Press, v. 10, n. 2, p. 374–389, 2009.
- TEAM, R. C. et al. *R: A language and environment for statistical computing*. Vienna, Austria, 2019.
- TOMAZELLA, V.; PEREIRA, G. H.; NOBRE, J. S.; SANTOS-NETO, M. Zero-adjusted reparameterized birnbaum–saunders regression model. *Statistics & Probability Letters*, Elsevier, v. 149, p. 142–145, 2019.
- TRASANDE, L.; CHATTERJEE, S. The impact of obesity on health service utilization and costs in childhood. *Obesity*, Wiley Online Library, v. 17, n. 9, p. 1749–1754, 2009.

- 
- VANEGAS, L. H.; CYSNEIROS, F. J. A. Assessment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics & Data Analysis*, Elsevier, v. 54, n. 4, p. 1002–1016, 2010.
- VANEGAS, L. H.; PAULA, G. A. A semiparametric approach for joint modeling of median and skewness. *Test*, Springer, v. 24, n. 1, p. 110–135, 2015.
- VANEGAS, L. H.; PAULA, G. A. et al. Log-symmetric distributions: statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 30, n. 2, p. 196–220, 2016.
- VENTURA, M.; SAULO, H.; LEIVA, V.; MONSUETO, S. Log-symmetric regression models: information criteria and application to movie business and industry data with economic implications. *Applied Stochastic Models in Business and Industry*, Wiley Online Library, v. 35, n. 4, p. 963–977, 2019.
- WEI, B.-C.; HU, Y.-Q.; FUNG, W.-K. Generalized leverage and its applications. *Scandinavian Journal of statistics*, Wiley Online Library, v. 25, n. 1, p. 25–37, 1998.
- WEI, G. C.; TANNER, M. A. A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association*, Taylor & Francis Group, v. 85, n. 411, p. 699–704, 1990.
- YAMASHITA, T.; YAMASHITA, K.; KAMIMURA, R. A stepwise aic method for variable selection in linear regression. *Communications in Statistics—Theory and Methods*, Taylor & Francis, v. 36, n. 13, p. 2395–2403, 2007.

## APÊNDICE A – PRIMEIRAS E SEGUNDAS DERIVADAS DOS MODELOS LINEARES ZERO AJUSTADO LOG-SIMÉTRICAS

### A.1 OBTENÇÃO DA FUNÇÃO ESCORE

O vetor escore é obtido derivando a função de log-verossimilhança em relação a cada parâmetro assim para  $j_1 = 1, \dots, M_1$ ,  $j_2 = 1, \dots, M_2$  e  $j_3 = 1, \dots, M_3$  temos que

$$\begin{aligned}
 U_{\beta_{j_1}} &= \frac{\partial \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \beta_{j_1}} = \sum_{i:w_i>0} \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \beta_{j_1}} \\
 &= \sum_{i:w_i>0} \left\{ \frac{g'(\tilde{w}_i^2)(2\tilde{w}_i)}{g(\tilde{w}_i^2)} \left( \frac{-1}{\sqrt{\phi_i \eta_i}} \right) \right\} \frac{\partial \eta_i}{\partial \xi_{1i}} x_{ij_1} \\
 &= \sum_{i:w_i>0} v(\tilde{w}_i) \frac{\log(w_i/\eta_i)}{\eta_i \phi_i} \frac{\partial \eta_i}{\partial \xi_{1i}} x_{ij_1} \\
 &= \sum_{i:w_i>0} v(\tilde{w}_i) \frac{\log(w_i) - \log(\eta_i)}{\eta_i \phi_i} \frac{\partial \eta_i}{\partial \xi_{1i}} x_{ij_1}
 \end{aligned}$$

$$\begin{aligned}
 U_{\alpha_{j_2}} &= \frac{\partial \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \alpha_{j_2}} = \sum_{i:w_i>0} \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \alpha_{j_2}} \\
 &= \sum_{i:w_i>0} \left\{ -\frac{1}{2\phi_i} + \frac{g'(\tilde{w}_i^2)(2\tilde{w}_i)}{g(\tilde{w}_i^2)} \left( \frac{-\tilde{w}_i^2}{2\phi_i} \right) \right\} \frac{\partial \phi_i}{\partial \xi_{2i}} z_{ij_2} \\
 &= \sum_{i:w_i>0} \left[ \frac{1}{2\phi_i} \right] (v(\tilde{w}_i)\tilde{w}_i^2 - 1) \frac{\partial \phi_i}{\partial \xi_{2i}} z_{ij_2}
 \end{aligned}$$

$$\begin{aligned}
 U_{\gamma_{j_3}} &= \frac{\partial \ell_{(1)}(\boldsymbol{\gamma})}{\partial \gamma_{j_3}} = \sum_{i=1}^n \frac{\partial \ell_i(\pi_i)}{\partial \pi_i} \frac{\partial \pi_i}{\partial \xi_{3i}} \frac{\partial \xi_{3i}}{\partial \gamma_{j_3}} \\
 &= \sum_{i=1}^n \left[ \frac{I_{\{0\}}(w_i)}{\pi_i} - \frac{1 - I_{\{0\}}(w_i)}{1 - \pi_i} \right] \frac{\partial \pi_i}{\partial \xi_{3i}} d_{ij_3} \\
 &= \sum_{i=1}^n \left[ \frac{I_{\{0\}}(w_i) - \pi_i}{\pi_i(1 - \pi_i)} \right] \frac{\partial \pi_i}{\partial \xi_{3i}} d_{ij_3}
 \end{aligned}$$

## A.2 SEGUNDAS DERIVADAS

A matriz de segundas derivadas da função de log-verossimilhança para  $p_1 = 1, \dots, M_1$ ,  $p_2 = 1, \dots, M_2$  e  $p_3 = 1, \dots, M_3$ . Pela separabilidade dos parâmetros dos componentes contínuo e discretos segue que  $U_{\gamma_{j_3}\beta_{j_1}} = 0$  e  $U_{\gamma_{j_3}\alpha_{j_2}} = 0$

Para o componente discreto temos que

$$\begin{aligned} U_{\gamma_{j_3}\gamma_{p_3}} &= \frac{\partial^2 \ell_{(1)}(\boldsymbol{\gamma})}{\partial \gamma_{j_3} \partial \gamma_{p_3}} = \sum_{i=1}^n \frac{\partial}{\partial \pi_i} \left( \frac{\partial \ell_i(\pi_i)}{\partial \pi_i} \frac{\partial \pi_i}{\partial \xi_{3i}} \frac{\partial \xi_{3i}}{\partial \gamma_{j_3}} \right) \frac{\partial \pi_i}{\partial \xi_{3i}} \frac{\partial \xi_{3i}}{\partial \gamma_{p_3}} \\ &= \sum_{i=1}^n \left\{ \frac{\partial^2 \ell_i(\pi_i)}{\partial \pi_i^2} \left( \frac{\partial \pi_i}{\partial \xi_{3i}} \right)^2 + \frac{\partial \ell_i(\pi_i)}{\partial \pi_i} \left( \frac{\partial}{\partial \pi_i} \frac{\partial \pi_i}{\partial \xi_{3i}} \right) \frac{\partial \pi_i}{\partial \xi_{3i}} \right\} \frac{\partial \xi_{3i}}{\partial \gamma_{j_3}} \frac{\partial \xi_{3i}}{\partial \gamma_{p_3}} \\ &= \sum_{i=1}^n \left\{ \frac{\partial^2 \ell_i(\pi_i)}{\partial \pi_i^2} \left( \frac{\partial \pi_i}{\partial \xi_{3i}} \right)^2 + \frac{\partial \ell_i(\pi_i)}{\partial \pi_i} \frac{\partial^2 \pi_i}{\partial \xi_{3i}^2} \right\} \frac{\partial \xi_{3i}}{\partial \gamma_{j_3}} \frac{\partial \xi_{3i}}{\partial \gamma_{p_3}} \\ &= \sum_{i=1}^n \psi_{1i} d_{ij_3} d_{ip_3}. \end{aligned}$$

Para o componente contínuo temos que

$$\begin{aligned} U_{\beta_{j_1}\beta_{p_1}} &= \frac{\partial^2 \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \beta_{j_1} \partial \beta_{p_1}} = \sum_{i:w_i>0} \frac{\partial}{\partial \eta_i} \left( \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \beta_{j_1}} \right) \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \beta_{p_1}} \\ &= \sum_{i:w_i>0} \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \eta_i^2} \left( \frac{\partial \eta_i}{\partial \xi_{1i}} \right)^2 + \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \eta_i} \left( \frac{\partial}{\partial \eta_i} \frac{\partial \eta_i}{\partial \xi_{1i}} \right) \frac{\partial \eta_i}{\partial \xi_{1i}} \right\} \frac{\partial \xi_{1i}}{\partial \beta_{j_1}} \frac{\partial \xi_{1i}}{\partial \beta_{p_1}} \\ &= \sum_{i=1}^n [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \eta_i^2} \left( \frac{\partial \eta_i}{\partial \xi_{1i}} \right)^2 + \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \eta_i} \frac{\partial^2 \eta_i}{\partial \xi_{1i}^2} \right\} x_{ij_1} x_{ip_1} \\ &= \sum_{i=1}^n \psi_{2i} x_{ij_1} x_{ip_1}, \end{aligned}$$

$$\begin{aligned} U_{\beta_{j_1}\alpha_{p_2}} &= \frac{\partial^2 \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \beta_{j_1} \partial \alpha_{p_2}} = \sum_{i:w_i>0} \frac{\partial}{\partial \phi_i} \left( \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \beta_{j_1}} \right) \frac{\partial \phi_i}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \alpha_{p_2}} \\ &= \sum_{i:w_i>0} \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \eta_i \partial \phi_i} \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \phi_i}{\partial \xi_{2i}} \right\} \frac{\partial \xi_{1i}}{\partial \beta_{j_1}} \frac{\partial \xi_{2i}}{\partial \alpha_{p_2}} \\ &= \sum_{i=1}^n [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \eta_i \partial \phi_i} \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \phi_i}{\partial \xi_{2i}} \right\} x_{ij_1} z_{ip_2} \\ &= \sum_{i=1}^n \psi_{3i} x_{ij_1} x_{ip_1}, \end{aligned}$$

$$\begin{aligned}
U_{\alpha_{j_2} \alpha_{p_2}} &= \frac{\partial^2 \ell_{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \alpha_{j_2} \partial \alpha_{p_2}} = \sum_{i:w_i > 0} \frac{\partial}{\partial \phi_i} \left( \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \alpha_{j_2}} \right) \frac{\partial \phi_i}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \alpha_{p_2}} \\
&= \sum_{i:w_i > 0} \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \phi_i^2} \left( \frac{\partial \phi_i}{\partial \xi_{2i}} \right)^2 + \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \phi_i} \left( \frac{\partial}{\partial \phi_i} \frac{\partial \phi_i}{\partial \xi_{2i}} \right) \frac{\partial \phi_i}{\partial \xi_{2i}} \right\} \frac{\partial \xi_{2i}}{\partial \alpha_{j_2}} \frac{\partial \xi_{2i}}{\partial \alpha_{p_2}} \\
&= \sum_{i=1}^n [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \phi_i^2} \left( \frac{\partial \phi_i}{\partial \xi_{2i}} \right)^2 + \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \phi_i} \frac{\partial^2 \phi_i}{\partial \xi_{2i}^2} \right\} z_{ij_2} z_{ip_2} \\
&= \sum_{i=1}^n \psi_{4i} z_{ij_2} z_{ip_2};
\end{aligned}$$

em que

$$\begin{aligned}
\psi_{1i} &= \left\{ \frac{\partial^2 \ell_i(\pi_i)}{\partial \pi_i^2} \left( \frac{\partial \pi_i}{\partial \xi_{3i}} \right)^2 + \frac{\partial \ell_i(\pi_i)}{\partial \pi_i} \frac{\partial^2 \pi_i}{\partial \xi_{3i}^2} \right\}, \\
\psi_{2i} &= [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \eta_i^2} \left( \frac{\partial \eta_i}{\partial \xi_{1i}} \right)^2 + \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \eta_i} \frac{\partial^2 \eta_i}{\partial \xi_{1i}^2} \right\}, \\
\psi_{3i} &= [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \eta_i \partial \phi_i} \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \phi_i}{\partial \xi_{2i}} \right\}, \\
\psi_{4i} &= [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_i)}{\partial \phi_i^2} \left( \frac{\partial \phi_i}{\partial \xi_{2i}} \right)^2 + \frac{\partial \ell_i(\eta_i, \phi_i)}{\partial \phi_i} \frac{\partial^2 \phi_i}{\partial \xi_{2i}^2} \right\}.
\end{aligned}$$

## APÊNDICE B – INFLUÊNCIA LOCAL

### B.1 PERTURBAÇÃO DE CASOS PONDERADOS

No esquema de perturbação de casos ponderados temos a função de log-verossimilhança perturbada  $l(\theta|\omega) = l_{(1)}(\gamma|\omega) + l_{(2)}(\beta, \alpha|\omega)$ , em que:

$$l_{(1)}(\gamma|\omega) = \sum_{i=1}^n \omega_i l_i(\pi_i)$$

$$l_{(2)}(\beta, \alpha|\omega) = \sum_{i=1}^n \omega_i l_i(\eta_i, \phi_i).$$

Em base ao Apêndice A, temos a  $i$ -ésima linha de  $\Delta$  para  $j_1 = 1, \dots, M_1$ ,  $j_2 = 1, \dots, M_2$  e  $j_3 = 1, \dots, M_3$  é dada por:

$$\frac{\partial^2 l_{(2)}(\beta, \alpha)}{\partial \beta_{j_1} \partial \omega_i} = [1 - I_{\{0\}}(w_i)] v(w_i^+) \frac{\log(w_i^+) - \log(\eta_i)}{\eta_i \phi_i} \frac{\partial \eta_i}{\partial \xi_{1i}} x_{ij_1},$$

$$\frac{\partial^2 l_{(2)}(\beta, \alpha)}{\partial \alpha_{j_2} \partial \omega_i} = [1 - I_{\{0\}}(w_i)] \left[ \frac{1}{2\phi_i} \right] (v(w_i^+) w_i^{+2} - 1) \frac{\partial \phi_i}{\partial \xi_{2i}} z_{ij_2},$$

$$\frac{\partial^2 ell_{(1)}(\gamma)}{\partial \gamma_{j_3} \omega_i} = \left[ \frac{I_{\{0\}}(w_i) - \pi_i}{\pi_i(1 - \pi_i)} \right] \frac{\partial \pi_i}{\partial \xi_{3i}} d_{ij_3}.$$

### B.2 PERTURBAÇÃO NAS VARIÁVEIS EXPLANATÓRIOS

#### B.2.1 Perturbação nas variáveis explanatórias individualmente

##### *Perturbação nas variáveis explanatórias da componente discreta*

Seja  $l(\theta|\omega)$  a função de log-verossimilhança perturbada sobre o esquema de perturbação:

$$d_{ij\omega} = d_{ij} + \omega_i s_{dj},$$

logo,

$$\frac{\partial l(\theta|\omega)}{\partial \omega_i} = \frac{\partial l_{(1)}(\gamma|\omega)}{\partial \omega_i} = \frac{\partial l_i(\pi_{i\omega})}{\partial \pi_{i\omega}} \frac{\partial \pi_{i\omega}}{\partial \xi_{3i}} \frac{\partial \xi_{3i}}{\partial \omega_i}.$$

Derivando em relação a  $\gamma_{j_3}$  para o caso em que  $j_3 \neq j$ :

$$\begin{aligned} \frac{\partial^2 \ell(\theta|\omega)}{\partial \gamma_{j_3} \partial \omega_i} &= \frac{\partial^2 \ell_i(\pi_{i\omega})}{\partial \pi_{i\omega}^2} \left( \frac{\partial \pi_{i\omega}}{\partial \xi_{3i}} \right)^2 \frac{\partial \xi_{3i}}{\gamma_{j_3}} \frac{\partial \xi_{3i}}{\partial \omega_i} \\ &\quad + \frac{\partial \ell_i(\pi_{i\omega})}{\partial \pi_{i\omega}} \frac{\partial^2 \pi_{i\omega}}{\partial \xi_{3i}^2} \frac{\partial \xi_{3i}}{\gamma_{j_3}} \frac{\partial \xi_{3i}}{\partial \omega_i} \\ &= \psi_{1i}(\omega) d_{ij_3} \gamma_j s_{d_j}, \end{aligned}$$

em que  $\psi_{1i}(\omega) = \left[ \frac{\partial^2 \ell_i(\pi_{i\omega})}{\partial \pi_{i\omega}^2} \left( \frac{\partial \pi_{i\omega}}{\partial \xi_{3i}} \right)^2 + \frac{\partial \ell_i(\pi_{i\omega})}{\partial \pi_{i\omega}} \frac{\partial^2 \pi_{i\omega}}{\partial \xi_{3i}^2} \right]$ . Para o caso em que  $j_3 = j$  segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \gamma_j \partial \omega_i} = \psi_{1i}(\omega) d_{ij} \gamma_j s_{d_j} + \left[ \frac{I_{\{0\}}(w_i) - \pi_{i\omega}}{\pi_{i\omega}(1 - \pi_{i\omega})} \right] \frac{\partial \pi_{i\omega}}{\partial \xi_{3i}} s_{d_j}.$$

Assim, em geral segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \gamma_{j_3} \partial \omega_i} = \begin{cases} \psi_{1i}(\omega) d_{ij_3} \gamma_j s_{d_j}, & \text{se } j_3 \neq j, \\ \psi_{1i}(\omega) d_{ij} \gamma_j s_{d_j} + \left[ \frac{I_{\{0\}}(w_i) - \pi_{i\omega}}{\pi_{i\omega}(1 - \pi_{i\omega})} \right] \frac{\partial \pi_{i\omega}}{\partial \xi_{3i}} s_{d_j}, & \text{se } j_3 = j, \end{cases}$$

### ***Perturbação individual nas variáveis explanatórias da componente contínua do sub-modelo associado a $\eta$***

Seja  $\ell(\theta|\omega)$  a função de log-verossimilhança perturbada sobre o esquema de perturbação:

$$x_{ip\omega} = x_{ip} + \omega_i s_{x_p},$$

logo,

$$\frac{\partial \ell(\theta|\omega)}{\partial \omega_i} = \frac{\partial \ell_{(2)}(\beta, \alpha|\omega)}{\partial \omega_i} = \frac{\partial \ell_i(\eta_{i\omega}, \phi_i)}{\partial \eta_{i\omega}} \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \omega_i} \quad (\text{B.1})$$

Derivando (B.1) em relação a  $\beta_{j_1}$  para o caso em que  $j_1 \neq p$ :

$$\begin{aligned} \frac{\partial^2 \ell(\theta|\omega)}{\partial \beta_{j_1} \partial \omega_i} &= \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_i)}{\partial \eta_{i\omega}^2} \left( \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \right)^2 \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \xi_{1i}}{\partial \omega_i} \\ &\quad + \frac{\partial \ell_i(\eta_{i\omega}, \phi_i)}{\partial \eta_{i\omega}} \frac{\partial^2 \eta_{i\omega}}{\partial \xi_{1i}^2} \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \xi_{1i}}{\partial \omega_i} \\ &= \psi_{2i}(\omega) \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \xi_{1i}}{\omega_i} \\ &= \psi_{2i}(\omega) x_{ij_1} \beta_p s_{x_p}, \end{aligned}$$

em que  $\psi_{2i}(\omega) = [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_i)}{\partial \eta_{i\omega}^2} \left( \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \right)^2 + \frac{\partial \ell_i(\eta_{i\omega}, \phi_i)}{\partial \eta_{i\omega}} \frac{\partial^2 \eta_{i\omega}}{\partial \xi_{1i}^2} \right\}$ . Para o caso em que  $j_1 = p$ , segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \beta_p \partial \omega_i} = \psi_{2i} x_{ip} \beta_p s_{x_p} + [1 - I_{\{0\}}(w_i)] \left( \frac{\log(w_i) - \log(\eta_{i\omega})}{\eta_{i\omega} \phi_i} \right) v(\tilde{w}_i) \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} s_{x_p}.$$

Assim, em geral segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \beta_{j_1} \partial \omega_i} = \begin{cases} \psi_{2i}(\omega) x_{ir_1} \beta_p s_{x_p}, & \text{se } j_1 \neq p, \\ \psi_{2i}(\omega) x_{ip} \beta_p s_{x_p} + [1 - I_{\{0\}}(w_i)] \left( \frac{\log(w_i) - \log(\eta_{i\omega})}{\eta_{i\omega} \phi_i} \right) v(\tilde{w}_i) \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} s_{x_p}, & \text{se } j_1 = p, \end{cases}$$

Derivando (B.1) em relação a  $\alpha_{j_2}$  para todo  $j_2 = 1, \dots, M_2$  segue que:

$$\begin{aligned} \frac{\partial^2 \ell(\theta|\omega)}{\partial \alpha_{j_2} \partial \omega_i} &= \left[ \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_i)}{\partial \phi_i \partial \eta_{i\omega}} \frac{\partial \phi_i}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \alpha_{j_2}} \right] \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \omega_i} \\ &= \psi_{3i}(\omega) z_{j_2} s_{x_p} \beta_p, \end{aligned}$$

em que  $\psi_{3i}(\omega) = [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_i)}{\partial \phi_i \partial \eta_{i\omega}} \frac{\partial \phi_i}{\partial \xi_{2i}} \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \right\}$ .

### ***Perturbação individual nas variáveis explanatórias da componente contínua do sub-modelo associado a $\phi$***

Seja  $\ell(\theta|\omega)$  a função de log-verossimilhança perturbada sobre o esquema de perturbação:

$$z_{iq\omega} = z_{iq} + \omega_i s_{z_q},$$

logo,

$$\frac{\partial \ell(\theta|\omega)}{\partial \omega_i} = \frac{\partial \ell_{(2)}(\beta, \alpha|\omega)}{\partial \omega_i} = \frac{\partial \ell_i(\eta_i, \phi_{i\omega})}{\partial \phi_{i\omega}} \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \omega_i} \quad (\text{B.2})$$

Derivando (B.2) em relação a  $\alpha_{j_2}$  para o caso em que  $j_2 \neq q$ :

$$\begin{aligned} \frac{\partial^2 \ell(\theta|\omega)}{\partial \alpha_{j_2} \partial \omega_i} &= \frac{\partial^2 \ell_i(\eta_i, \phi_{i\omega})}{\partial \phi_{i\omega}^2} \left( \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \right)^2 \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \xi_{2i}}{\partial \omega_i} \\ &+ \frac{\partial \ell_i(\eta_i, \phi_{i\omega})}{\partial \phi_{i\omega}} \frac{\partial^2 \phi_{i\omega}}{\partial \xi_{2i}^2} \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \xi_{2i}}{\partial \omega_i} \\ &= \psi_{4i}(\omega) \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \xi_{2i}}{\omega_i} \\ &= \psi_{4i}(\omega) z_{ij_2} \alpha_q s_{z_q} \end{aligned}$$

em que  $\psi_{4i}(\omega) = [1 - I_{\{0\}}(w_i)] \left\{ \frac{\partial^2 \ell_i(\eta_i, \phi_{i\omega})}{\partial \phi_{i\omega}^2} \left( \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \right)^2 + \frac{\partial \ell_i(\eta_i, \phi_{i\omega})}{\partial \phi_{i\omega}} \frac{\partial^2 \phi_{i\omega}}{\partial \xi_{2i}^2} \right\}$ . Para o caso em que  $j_2 = q$ , segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \alpha_q \partial \omega_i} = \psi_{4i}(\omega) z_{iq} \alpha_q s_{z_q} + [1 - I_{\{0\}}(w_i)] \left[ \frac{1}{2\phi_{i\omega}} \right] (v(\tilde{w}_i) \tilde{w}_i^2 - 1) \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} s_{z_q}.$$

Assim, em geral segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \alpha_{j_2} \partial \omega_i} = \begin{cases} \psi_{4i}(\omega) z_{ij_2} \alpha_q s_{z_q}, \\ \psi_{4i}(\omega) z_{iq} \alpha_q s_{z_q} + [1 - I_{\{0\}}(w_i)] \left[ \frac{1}{2\phi_{i\omega}} \right] (v(\tilde{w}_i) \tilde{w}_i^2 - 1) \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} s_{z_q}, & \text{se } j_2 = q, \end{cases}$$

Derivando (B.2) em relação a  $\beta_{j_1}$  para todo  $j_1 = 1, \dots, M_1$  segue que:

$$\begin{aligned} \frac{\partial^2 \ell(\theta|\omega)}{\partial \beta_{j_1} \partial \omega_i} &= \left[ \frac{\partial^2 \ell_i(\eta_i, \phi_{i\omega})}{\partial \eta_i \partial \phi_{i\omega}} \frac{\partial \eta_i}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\beta_{j_1}} \right] \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \omega_i} \\ &= \psi_{3i}(\omega) x_{j_1} s_{z_q} \alpha_q. \end{aligned}$$

## B.2.2 Perturbação nas variáveis explanatórias simultaneamente

### *Perturbação nas variáveis explanatórias da componente contínua*

Seja  $\ell(\theta|\omega)$  a função de log-verossimilhança perturbada sobre o esquema de perturbação:

$$x_{ip\omega} = x_{ip} + \omega_i s_{x_p},$$

$$z_{iq\omega} = z_{iq} + \omega_i s_{z_q},$$

logo,

$$\frac{\partial \ell(\theta|\omega)}{\partial \omega_i} = \frac{\partial \ell_{(2)}(\beta, \alpha|\omega)}{\partial \omega_i} = \frac{\partial \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \eta_{i\omega}} \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \omega_i} + \frac{\partial \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \phi_{i\omega}} \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \omega_i}$$

Derivando em relação a  $\beta_{j_1}$  para o caso em que  $j_1 \neq p$ :

$$\begin{aligned}
\frac{\partial^2 \ell(\theta|\omega)}{\partial \beta_{j_1} \partial \omega_i} &= \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \eta_{i\omega}^2} \left( \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \right)^2 \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \xi_{1i}}{\partial \omega_i} \\
&+ \frac{\partial \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \eta_{i\omega}} \frac{\partial^2 \eta_{i\omega}}{\partial \xi_{1i}^2} \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \xi_{1i}}{\partial \omega_i} \\
&+ \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \eta_{i\omega} \partial \phi_{i\omega}} \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\partial \omega_i} \\
&= \psi_{2i}(\omega) \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \xi_{1i}}{\omega_i} + \psi_{3i}(\omega) \frac{\partial \xi_{1i}}{\beta_{j_1}} \frac{\partial \xi_{2i}}{\partial \omega_i} \\
&= \psi_{2i}(\omega) x_{ij_1} \beta_p s_{x_p} + \psi_{3i}(\omega) x_{ij_1} \alpha_q s_{z_q},
\end{aligned}$$

Para o caso em que  $j_1 = p$ , segue que:

$$\begin{aligned}
\frac{\partial^2 \ell(\theta|\omega)}{\partial \beta_p \partial \omega_i} &= \psi_{2i}(\omega) x_{ip} \beta_p s_{x_p} + \psi_{3i}(\omega) x_{ip} \alpha_q s_{z_q} \\
&+ [1 - I_{\{0\}}(w_i)] \left( \frac{\log(w_i) - \log(\eta_{i\omega})}{\eta_{i\omega} \phi_{i\omega}} \right) \mathbf{v}(\tilde{w}_i) \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} s_{x_p}.
\end{aligned}$$

Assim, em geral segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \beta_{j_1} \partial \omega_i} = \begin{cases} \psi_{2i}(\omega) x_{ij_1} \beta_p s_{x_p} + \psi_{3i}(\omega) x_{ij_1} \alpha_q s_{z_q}, & \text{se } j_1 \neq p, \\ \psi_{2i}(\omega) x_{ip} \beta_p s_{x_p} + \psi_{3i}(\omega) x_{ip} \alpha_q s_{z_q} + [1 - I_{\{0\}}(w_i)] \left( \frac{\log(w_i) - \log(\eta_{i\omega})}{\eta_{i\omega} \phi_{i\omega}} \right) \mathbf{v}(\tilde{w}_i) \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} s_{x_p}, & \text{se } j_1 = p, \end{cases}$$

Agora derivando em relação a  $\alpha_{j_2}$  para o caso em que  $j_2 \neq q$ :

$$\begin{aligned}
\frac{\partial^2 \ell(\theta|\omega)}{\partial \alpha_{j_2} \partial \omega_i} &= \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \phi_{i\omega}^2} \left( \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \right)^2 \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \xi_{2i}}{\partial \omega_i} \\
&+ \frac{\partial \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \phi_{i\omega}} \frac{\partial^2 \phi_{i\omega}}{\partial \xi_{2i}^2} \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \xi_{2i}}{\partial \omega_i} \\
&+ \frac{\partial^2 \ell_i(\eta_{i\omega}, \phi_{i\omega})}{\partial \phi_{i\omega} \partial \eta_{i\omega}} \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \eta_{i\omega}}{\partial \xi_{1i}} \frac{\partial \xi_{1i}}{\partial \omega_i} \\
&= \psi_{4i}(\omega) \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \xi_{2i}}{\omega_i} + \psi_{3i}(\omega) \frac{\partial \xi_{2i}}{\alpha_{j_2}} \frac{\partial \xi_{1i}}{\partial \omega_i} \\
&= \psi_{4i}(\omega) z_{ir_2} \alpha_q s_{z_q} + \psi_{3i}(\omega) z_{ij_2} \beta_p s_{x_p}
\end{aligned}$$

Para o caso em que  $j_2 = q$ , segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \alpha_q \partial \omega_i} = \psi_{4i}(\omega) z_{ij_2} \alpha_q s_{z_q} + \psi_{3i}(\omega) z_{ij_2} \beta_p s_{x_p} + [1 - I_{\{0\}}(w_i)] \left[ \frac{1}{2\phi_{i\omega}} \right] (\mathbf{v}(\tilde{w}_i) \tilde{w}_i^2 - 1) \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} s_{z_q}.$$

Assim, em geral segue que:

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial \alpha_{j_2} \partial \omega_i} = \begin{cases} \psi_{4i}(\omega) z_{ij_2} \alpha_q s_{z_q} + \psi_{3i}(\omega) z_{ij_2} \beta_p s_{x_p}, & \text{se } j_2 \neq q, \\ \psi_{4i}(\omega) z_{iq} \alpha_q s_{z_q} + \psi_{3i}(\omega) z_{iq} \beta_p s_{x_p} + [1 - I_{\{0\}}(w_i)] \left[ \frac{1}{2\phi_{i\omega}} \right] (\mathbf{v}(\tilde{w}_i) \tilde{w}_i^2 - 1) \frac{\partial \phi_{i\omega}}{\partial \xi_{2i}} s_{z_q}, & \text{se } j_2 = q, \end{cases}$$

## APÊNDICE C – ALAVANCAGEM GENERALIZADA

Pela propriedade de separabilidade dos parâmetros, podemos obter duas expressões para alavancagem generalizada, uma que depende dos parâmetros da componente discreta e outra da componente contínua, além disso, para a componente contínua, é possível observar a influência para  $\eta$  e  $\phi$ .

### C.1 CÁLCULOS PARA $GL^\pi(\hat{\gamma})$

Para a componente discreta, a alavancagem generalizada de  $\hat{\gamma}$  é definida como

$$GL^\pi(\hat{\gamma}) = \{ \mathbf{D}_\gamma^\pi (-\ddot{L}_{\gamma\gamma})^{-1} \ddot{L}_{\gamma w^*} \} |_{\hat{\gamma}}$$

em que  $\mathbf{D}_\gamma^\pi = \partial \boldsymbol{\pi} / \partial \boldsymbol{\gamma}^\top$ ,  $\ddot{L}_{\gamma\gamma}$  é a matriz jacobiana e  $\ddot{L}_{\gamma w^*} = \partial^2 \ell_{(1)}(\boldsymbol{\gamma}) / \partial \boldsymbol{\gamma} \partial \mathbf{w}^{*\top}$ . Para  $j_3 = 1, \dots, M_3$ , temos que

$$\begin{aligned} \frac{\partial \pi_i}{\partial \gamma_{j_3}} &= \frac{\partial \pi_i}{\partial \xi_{3i}} \frac{\partial \xi_{3i}}{\partial \gamma_{j_3}} \\ &= \frac{\partial \pi_i}{\partial \xi_{3i}} d_{ij_3} \end{aligned}$$

e

$$\begin{aligned} \frac{\partial^2 \ell_{(1)}(\boldsymbol{\gamma})}{\partial \gamma_{j_3} \partial w_i^*} &= \frac{\partial}{\partial w_i^*} \left( \sum_{i=1}^n \left[ \frac{w_i^* - \pi_i}{\pi_i(1 - \pi_i)} \right] \frac{\partial \pi_i}{\partial \xi_{3i}} d_{ij_3} \right) \\ &= \sum_{i=1}^n \frac{1}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \xi_{3i}} d_{ij_3} \\ &= \sum_{i=1}^n \Phi_{1i} d_{ij_3} \end{aligned}$$

em que  $\Phi_{1i} = \frac{1}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \xi_{3i}}$ .

### C.2 CÁLCULOS PARA $GL^\eta(\hat{\boldsymbol{\theta}}_1)$

Para a componente contínua, especificamente para  $\eta$ , a alavancagem generalizada de  $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$  é definida como

$$GL^\eta(\hat{\boldsymbol{\theta}}_1) = \{ \mathbf{D}_{\boldsymbol{\theta}_1}^\eta (-\ddot{L}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1})^{-1} \ddot{L}_{\boldsymbol{\theta}_1 w} \} |_{\hat{\boldsymbol{\theta}}_1}$$

em que  $\mathbf{D}_{\theta_1}^\eta = (\mathbf{D}_\beta^\eta, \mathbf{D}_\alpha^\eta) = (\partial\eta/\partial\beta^\top, \partial\eta/\partial\alpha^\top)$ ,  $\ddot{L}_{\theta_1\theta_1} = \partial^2\ell(\theta_1)/\partial\theta_1\partial\theta_1^\top$  é a matriz jacobiana de  $\theta_1$  e  $\ddot{L}_{\theta_1 w} = \partial^2\ell(\theta_1)/\partial\theta_1\partial w^\top$ . Para  $j_1 = 1, \dots, M_1$  e  $j_2 = 1, \dots, M_2$ , temos que

$$\begin{aligned} \frac{\partial\eta_i}{\partial\beta_{j_1}} &= \frac{\partial\eta_i}{\partial\xi_{1i}} \frac{\partial\xi_{1i}}{\partial\beta_{j_1}} & \frac{\partial\eta_i}{\partial\alpha_{j_2}} &= \frac{\partial\eta_i}{\partial\xi_{1i}} \frac{\partial\xi_{1i}}{\partial\alpha_{j_2}} = 0 \\ &= \frac{\partial\eta_i}{\partial\eta_{1i}} x_{ij_1}, \end{aligned}$$

e

$$\begin{aligned} \frac{\partial^2\ell_{(2)}(\beta, \alpha)}{\partial\beta_{j_1}\partial w_i} &= \frac{\partial}{\partial w_i} \left( \sum_{i:w_i>0} v(\tilde{w}_i) \frac{\log(w_i) - \log(\eta_i)}{\eta_i\phi_i} \frac{\partial\eta_i}{\partial\xi_{1i}} x_{ij_1} \right) \\ &= \sum_{i=1}^n (1 - I_{\{0\}}(w_i)) \frac{1}{w_i\eta_i\phi_i} [v'(\tilde{w}_i)\tilde{w}_i + v(\tilde{w}_i)] \frac{\partial\eta_i}{\partial\xi_{1i}} x_{ij_1} \\ &= \sum_{i=1}^n \Phi_{2i} x_{ij_1}, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2\ell_{(2)}(\beta, \alpha)}{\partial\alpha_{j_2}\partial w_i} &= \frac{\partial}{\partial w_i} \left( \sum_{i:w_i>0} \left[ \frac{1}{2\phi_i} \right] (v(\tilde{w}_i)\tilde{w}_i^2 - 1) \frac{\partial\phi_i}{\partial\xi_{2i}} z_{ij_2} \right) \\ &= \sum_{i=1}^n (1 - I_{\{0\}}(w_i)) \frac{1}{w_i\phi_i^{(3/2)}} [v'(\tilde{w}_i)\tilde{w}_i^2/2 + \tilde{w}_i v(\tilde{w}_i)] \frac{\partial\phi_i}{\partial\xi_{2i}} z_{ij_2} \\ &= \sum_{i=1}^n \Phi_{3i} z_{ij_2}, \end{aligned}$$

em que  $\Phi_{2i} = \frac{(1-I_{\{0\}}(w_i))}{w_i\eta_i\phi_i} [v'(\tilde{w}_i)(\tilde{w}_i) + v(\tilde{w}_i)] \frac{\partial\eta_i}{\partial\xi_{1i}}$  e  $\Phi_{3i} = \frac{(1-I_{\{0\}}(w_i))}{w_i\phi_i^{(3/2)}} [v'(\tilde{w}_i)\tilde{w}_i^2/2 + \tilde{w}_i v(\tilde{w}_i)] \frac{\partial\phi_i}{\partial\xi_{2i}}$ .

### C.3 CÁLCULOS PARA $GL^\phi(\hat{\theta}_1)$

Para a componente contínua, especificamente para  $\phi$ , a alavancagem generalizada de  $\theta_1 = (\beta^\top, \alpha^\top)^\top$  é definida como

$$GL^\phi(\hat{\theta}_1) = \{\mathbf{D}_{\theta_1}^\phi (-\ddot{L}_{\theta_1\theta_1})^{-1} \ddot{L}_{\theta_1 w}\} |_{\hat{\theta}_1}$$

em que  $\mathbf{D}_{\theta_1}^\phi = (\mathbf{D}_\beta^\phi, \mathbf{D}_\alpha^\phi) = (\partial\phi/\partial\beta^\top, \partial\phi/\partial\alpha^\top)$ . Para  $j_1 = 1, \dots, M_1$  e  $j_2 = 1, \dots, M_2$ , temos que

$$\begin{aligned} \frac{\partial\phi_i}{\partial\beta_{j_1}} &= \frac{\partial\eta_i}{\partial\xi_{2i}} \frac{\partial\xi_{2i}}{\partial\beta_{j_1}} = 0 & \frac{\partial\phi_i}{\partial\alpha_{j_2}} &= \frac{\partial\eta_i}{\partial\xi_{2i}} \frac{\partial\xi_{2i}}{\partial\alpha_{j_2}} \\ & & &= \frac{\partial\phi_i}{\partial\eta_{2i}} z_{ij_2}. \end{aligned}$$

