Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Regressão beta não linear robusta

por

Eduardo de Sousa Carvalho

Brasília, 3 de setembro de 2025

Regressão beta não linear robusta

por

Eduardo de Sousa Carvalho

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília como requisito parcial para obtenção do título de Mestre em Estatística.

Orientadora: Profa. Dra. Terezinha Késsia de Assis Ribeiro

Brasília, 3 de setembro de 2025

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Profa. Dra. Terezinha Késsia de Assis Ribeiro
Orientadora, EST/UnB

Prof. Dr. Guilherme Souza Rodrigues
EST/UnB

Profa. Dra. Silvia Lopes de Paula Ferrari
IME/USP

Agradecimentos

Agradeço à minha amada esposa Flávia, a minha maior incentivadora e motivadora, que sempre foi compreensiva com as minhas escolhas e com quem sempre pude contar em todos os momentos. Ao meu filho Augusto, que nasceu durante a elaboração desse trabalho, e tem trazido mais alegria às nossas vidas.

A toda minha família, que sempre apoiou minhas decisões e me deu suporte em todos os momentos, em especial, minha mãe Maria Francisca, meu pai José Ribeiro, hoje já falecido, e minha irmã Fernanda.

À minha orientadora, Profa. Dra. Terezinha Késsia de Assis Ribeiro, com quem tenho tido a satisfação de trabalhar desde a graduação e que tem me motivado a trilhar caminhos que não são os mais fáceis, mas com certeza são os mais recompensadores.

Meus sinceros agradecimentos aos professores e servidores do EST/UnB e PPGEST/UnB, pelo profissionalismo e competência. Agradeço também à Profa. Dra. Cira Etheowalda Guevara Otiniano, ao Prof. Dr. Guilherme Souza Rodrigues e à Profa. Dra. Silvia Lopes de Paula Ferrari, por dedicarem seus tempos para participar das bancas avaliadoras do exame de qualificação e/ou da defesa final do mestrado, contribuindo com suas experiências e orientações valiosas.

E, por fim, agradeço à Universidade de Brasília (UnB), que me acolheu ao longo desses 12 anos da minha jornada acadêmica, desde as duas graduações até o mestrado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Regressão beta não linear robusta

A regressão beta é frequentemente utilizada para modelar dados restritos ao intervalo contínuo unitário, a exemplo de taxas, frações e proporções. O método inferencial padrão utilizado para estimação dos parâmetros da regressão beta é o método da máxima verossimilhança. Entretanto, este método é sensível a observações discrepantes nos dados, podendo, em muitos casos, conduzir a resultados errôneos sobre a relação entre a resposta e as covariáveis de interesse. Nesse sentido, Ribeiro e Ferrari (2023) e, mais recentemente, Maluf, Ferrari e Queiroz (2025) propuseram métodos de estimação robustos alternativos ao método da máxima verossimilhança, objetivando reduzir a influência de observações atípicas no processo de estimação. Os referidos métodos de estimação robustos foram desenvolvidos sob modelos de regressão beta que consideram em suas estruturas de regressão preditores que são funções lineares de seus parâmetros. Assim, o presente trabalho se propõe a adaptar os métodos de estimação robustos aqui mencionados, estendendo-os a modelos de regressão beta não lineares.

O processo teórico de obtenção dos estimadores robustos sob estruturas não lineares de regressão beta foi desenvolvido e estudado por meio de suas propriedades teóricas. Também foi mostrada uma adaptação do teste de Wald como alternativa robusta para avaliação da significância dos parâmetros da regressão. Para seleção do valor ótimo da constante de afinação necessária nos procedimentos robustos, propusemos uma adaptação ao método orientado a dados desenvolvido por Ribeiro e Ferrari (2023), com o objetivo de deixar o processo de seleção mais estável e computacionalmente mais eficiente em cenários onde se utiliza o erro padrão estimado por meio de método *bootstrap.*

Foram realizados estudos de simulações de Monte Carlo e aplicação com dados simulados, por meio dos quais foi verificado que os modelos de regressão beta não lineares sob os estimadores robustos proporcionam menores viéses na presença de contaminação quando comparados aos modelos sob os estimadores tradicionais. Por fim, apresentamos outros resultados relacionados ao processo de estimação robusta e efetuamos discussões a partir de uma aplicação com dados reais.

Palavras-chave: Constante de afinação; Distribuição beta; Inferência robusta; L_q-verossimilhança; Regressão beta não linear robusta; Robustez.

Abstract

Robust nonlinear beta regression

Beta regression models are frequently employed for modeling data restricted to the unit interval, such as rates, fractions, and proportions. Parameter estimation in beta regression is typically performed using maximum likelihood estimation. However, this method is known to be sensitive to outliers, which can lead to biased or misleading inferences regarding the relationship between the response variable and the covariates of interest. To address this issue, Ribeiro and Ferrari (2023), and more recently, Maluf, Ferrari and Queiroz (2024), have proposed robust estimation methods as alternatives to the maximum likelihood approach. These methods aim to mitigate the influence of atypical observations on the estimation process. Their techniques were developed under beta regression models in which predictors are incorporated linearly into the regression structures. This study aims to extend these robust estimation methods to nonlinear beta regression models, thereby broadening their applicability.

To this end, we develop a theoretical framework for obtaining robust estimators under nonlinear regression structures and investigate the theoretical properties of the resulting models. In addition, it is shown a robust adaptation of the Wald test for assessing the statistical significance of regression parameters. To select the optimal value of the tuning constant, we propose a modification of the data-driven procedure introduced by Ribeiro and Ferrari (2023), designed to improve stability and computational efficiency, particularly in settings where standard errors are estimated via bootstrap methods.

A comprehensive Monte Carlo simulation study and an application using simulated data are conducted to evaluate the performance of the proposed methods. The results demonstrate that nonlinear beta regression models estimated via robust methods yield reduced bias in the presence of contamination when compared to models fitted with conventional estimators. Finally, we present additional findings regarding the robust estimation process and discuss their implications through an application to real-world data.

Palavras-chave: Beta distribution; L_q-likelihood; Nonlinear robust beta regression model; Robust inference; Robustness; Tuning constant.

Sumário

Lista de Tabelas

Lista de Algoritmos

Lista de Figuras

Lista de Siglas e Abreviaturas

AE Eficiência Assintótica (*Asymptotic Efficiency*)

BGFS Broyden-Fletcher-Goldfarb-Shanno

CDF Função de Distribuição Acumulada (*Cumulative Distribution Function*)

EDF Função de Distribuição Empírica (*Empirical Distribution Function*)

EGB Beta Exponencial Generalizada do Segundo Tipo (*Exponential Generalized Beta of the Second Type*)

ENPEV Norma Euclidiana dos Vetores das Estimativas dos Parâmetros (*Euclidian Norm of the Parameter Estimate Vectors*)

FIM Matriz de Informação de Fisher (*Fisher Information Matrix*)

GLM Modelos Lineares Generalizados (*Generalized Linear Models*)

IF Função de Influência (*Influence Function*)

IID Independentes e Identicamente Distribuídas (*Independent and Identically Distributed*)

ISS Sensibilidade Padronizada pela Informação (*Information-standardized Sensitivity*)

LMDPDE Estimador Logit de Mínima Divergência Potência entre Densidades (*Logit Minimum Density Power Divergence Estimator*)

LSMLE Estimador Logit de Máxima Verossimilhança Substituto (*Logit Surrogate Maximum Likelihood Estimator*)

MDPDE Estimador de Mínima Divergência Potência entre Densidades (*Minimum Density Power Divergence Estimator*)

MLE Estimador de Máxima Verossimilhança (*Maximum Likelihood Estimator*)

ML$_q$E Estimador de Máxima L$_q$-verossimilhança (*Maximum L_q-likelihood Estimator*)

PDF Função Densidade de Probabilidade (*Probability Density Function*)

SE Erro Padrão (*Standard Error*)

SMLE Estimador de Máxima Verossimilhança Substituto (*Surrogate Maximum Likelihood Estimator*)

SQV Variações Quadráticas Padronizadas (*Standardized Quadratric Variations*)

SSS Sensibilidade Auto-padronizada (*Self-standardized Sensitivity*)

SST Temperatura da Superfície do Mar (*Sea Surface Temperature*)

TMSE Erros Quadráticos Médios Totais (*Total Mean Squared Errors*)

TTP Porcentagem de Atum Tropical (*Tropical Tuna Percentage*)

UGES Sensibilidade a Erro Grosseiro não Padronizada (*Unstandardized Gross-error Sensitivity*)

1 Introdução

A modelagem adequada de dados contínuos limitados ao intervalo unitário surge naturalmente como um obstáculo a ser superado em diversas áreas do conhecimento. Tais tipos de dados são geralmente utilizados para representar fenômenos e situações que envolvem, por exemplo, taxas, proporções, porcentagens e frações. Dentre os diversos casos práticos, pode-se citar a fração da renda familiar gasta com alimentação, escores de qualidade de vida e taxas específicas de mortalidade. Para lidar com dados que possuem tal característica dentro do contexto de regressão, pode-se modelar a média μ_t de uma variável y_t denominada de resposta que assume valores no intervalo (0,1) em função de outras variáveis que são conhecidas e fixadas. Estas últimas são comumente chamadas de covariáveis ou variáveis explicativas.

Para dados desta natureza, se torna apropriado supor uma distribuição de probabilidades para y_t que tenha suporte no intervalo contínuo (0,1) e acomode diversas formas. Considerando um contexto de regressão, na literatura existem algumas propostas baseadas em distribuições de probabilidades com suporte no intervalo (0,1), a exemplo dos trabalhos de Kieschnick e McCullough (2003), Ferrari e Cribari-Neto (2004), Gómez-Déniz, Sordo e Calderín-Ojeda (2014), Lemonte e Bazán (2016), Smithson e Shou (2017) e Queiroz e Ferrari (2024). Para o desenvolvimento deste trabalho, focaremos em abordagens nas quais a distribuição de probabilidades da variável resposta, condicionada aos valores das covariáveis, segue uma distribuição beta.

O modelo probabilístico beta é uma distribuição de probabilidades associada a uma variável aleatória contínua que assume valores no intervalo (0,1). A função densidade de probabilidade (*probability density function*; PDF) da distribuição beta possui dois parâmetros e, a depender da combinação entre estes, pode assumir diversas formas, incluindo formas assimétricas. Considerando uma reparametrização desta distribuição, Ferrari e Cribari-Neto (2004) propuseram uma classe de modelos de regressão em que y_t segue uma distribuição beta indexada pela média μ_t e por um parâmetro de precisão ϕ. Nesta abordagem, a média μ_t é modelada através de uma estrutura de regressão linear $g_\mu(\mu_t) = \boldsymbol{X}_t^\top \boldsymbol{\beta}$ com $g_{\boldsymbol{\mu}}(\cdot) : (0,1) \to \mathbb{R}$ denominada de função de ligação. Sendo assim, obtém-se que $\mu_t = g_{\boldsymbol{\mu}}^{-1}(\boldsymbol{X}_t^\top \boldsymbol{\beta}) \in (0,1)$. Assim, ao estimar μ_t por $\widehat{\mu}_t$, sempre será obtido um valor ajustado dentro do intervalo (0,1). Também, este modelo é heteroscedástico pois a variância de y_t varia com as covariáveis através de sua média μ_t. Uma extensão natural desta abordagem é supor que a precisão dos dados também varie de acordo com as covariáveis. Tal proposta foi introduzida por Smithson e Verkuilen (2006) onde supõe-se que y_t segue uma distribuição beta indexada pela média μ_t e precisão ϕ_t. Nesse sentido, atribui-se uma estrutura de regressão linear também para a precisão ϕ_t.

O método mais utilizado para estimar os parâmetros do modelo de regressão beta é o método da máxima verossimilhança, por meio do qual é obtido um estimador de máxima verossimilhança (*maximum likelihood estimator*; MLE) para cada parâmetro. Entretanto, os MLEs não são considerados robustos, ou seja, podem ser fortemente influenciados pela presença de observações discrepantes nos dados de interesse. Ribeiro e Ferrari (2023) ilustraram esse comportamento do MLE ao analisar um conjunto de dados sobre práticas de gestão de risco de algumas firmas a partir de um conjunto de dados disponibilizados por Schmit e Roth (1990). No trabalho, verificou-se que a presença de observações atípicas conduziu a efeitos desproporcionais na curva de regressão beta ajustada, quando comparada às curvas obtidas com os dados sem as observações tidas como atípicas. Ribeiro e Ferrari (2023) demonstraram matematicamente que o procedimento de estimação de máxima verossimilhança não é robusto para os parâmetros dos modelos de regressão beta e, portanto, podem ser desproporcionalmente influenciados pela presença de observações atípicas.

Ao longo do tempo foram propostas diversas abordagens para lidar com observações discrepantes em problemas que envolvem a modelagem de dados onde a variável resposta assume valores no intervalo $(0, 1)$ ou ao menos em um intervalo contínuo limitado. Nos trabalhos de Bayes, Bazán e Garcıa (2012), Migliorati, Brisco e Ongaro (2018) e Brisco, Migliorati e Ongaro (2020), por exemplo, foram introduzidas abordagens que não envolvem procedimentos robustos de estimação dos parâmetros, mas a utilização de diferentes tipos de misturas de distribuições beta. Contudo, conforme pontuado por Ribeiro e Ferrari (2023), tais abordagens garantiram uma maior flexibilidade para acomodar dados com observações atípicas, porém, ao custo de trazer maior quantidade de parâmetros e complexidades adicionais aos respectivos modelos.

Recentemente foram publicados trabalhos que, de fato, consideram métodos robustos para estimação de parâmetros nos quais é mantida a distribuição beta reparametrizada por Ferrari e Cribari-Neto (2004) como base para construção dos modelos de regressão. Ghosh (2019) propôs o estimador de mínima divergência potência entre densidades (*minimum density power divergence estimator*; MDPDE), um estimador robusto baseado na minimização da divergência potência entre densidades. Tal método envolve uma constante de afinação, denotada por α ($\alpha \geq 0$), que, conforme demonstrado por Basu et al. (1998), controla o balanceamento entre eficiência assintótica e robustez do referido estimador. A escolha de um valor ideal para α constitui um problema adicional no processo de estimação, uma vez que valores mais altos para α privilegiam a robustez do estimador em detrimento da eficiência.

Ribeiro e Ferrari (2023) propuseram um estimador robusto, denominado estimador de máxima verossimilhança substituto (*surrogate maximum likelihood estimator*; SMLE), baseado na maximização da L_q-verossimilhança reparametrizada introduzida por Ferrari

e La Vecchia (2012). Este procedimento de estimação também depende de uma constante de afinação similar à do MDPDE. Além disso, o processo desenvolvido contempla um método para seleção de um valor ótimo para a constante de afinação, que se baseia nos dados de interesse e assegura eficiência assintótica máxima na ausência de observações atípicas. Ribeiro e Ferrari (2023) demonstraram, ainda, que tanto o MDPDE quanto o SMLE são estimadores bem definidos, robustos e possuem boas propriedades assintóticas para distribuições beta que sejam limitadas. Portanto, não há garantias desses resultados caso o método de estimação seja aplicado sob distribuições beta que sejam ilimitadas. Mais recentemente, Maluf, Ferrari e Queiroz (2025) efetuaram simulações com cenários envolvendo distribuições beta ilimitadas nas quais o MDPDE e o SMLE apresentaram consideráveis índices de falhas no processo de estimação, seja por não alcançar convergência para as estimativas dos parâmetros ou seja por não ser possível calcular seus erros padrão assintóticos.

Maluf, Ferrari e Queiroz (2025) propõem duas novas abordagens para obtenção de estimadores robustos sob modelos de regressão beta, usando a PDF da variável resposta transformada pela função logito. O primeiro deles, chamado de estimador logit de mínima divergência potência entre densidades (*logit minimum density power divergence estimator*; LMDPDE), é baseado no método de Ghosh (2019), enquanto o segundo, denominado estimador logit de máxima verossimilhança substituto (*logit surrogate maximum likelihood estimator*; LSMLE), é uma adaptação do método introduzido por Ribeiro e Ferrari (2023). Para ambos os estimadores propostos foi implementado o método orientado aos dados de interesse desenvolvido por Ribeiro e Ferrari (2023) para seleção da constante de afinação α. Maluf, Ferrari e Queiroz (2025) demostraram matematicamente e ilustraram por meio de simulações que ambos estimadores propostos são bem definidos, robustos e mantém boas propriedades assintóticas sem a necessidade de exigir restrições em relação à distribuição beta para a qual os parâmetros de regressão estão sendo estimados. Portanto, o LMDPDE e o LSMLE representaram uma evolução dos estimadores anteriores, uma vez que funcionam bem mesmo sob distribuições beta não limitadas.

Considerando a capacidade da distribuição beta em assumir uma grande quantidade de formas, o modelo heteroscedástico introduzido por Ferrari e Cribari-Neto (2004) se mostrou bastante flexível e adequado para modelar dados oriundos de uma grande quantidade de fenômenos. A modelagem simultânea do parâmetro de precisão introduzida por Smithson e Verkuilen (2006), trouxe mais flexibilidade ao ajuste de modelos de regressão beta. No entanto, em ambos os casos foram consideradas estruturas de regressão lineares nos parâmetros. Posteriormente, Simas, Barreto-Souza e Rocha (2010) apresentaram uma forma mais geral para o modelo de regressão beta com precisão variável, na qual as estruturas de regressão são descritas por relações não necessariamente lineares nos parâmetros, e da qual os modelos anteriores representam casos particulares. Essa última

abordagem tem o potencial de flexibilizar ainda mais a classe de modelos de regressão beta com precisão variável.

Todos os métodos de estimação robustos aqui mencionados foram desenvolvidos e aplicados a modelos de regressão beta que consideram em suas estruturas de regressão preditores que representam funções lineares de seus parâmetros. Segundo a definição de robustez de estimadores que utilizaremos ao longo desse trabalho, e de acordo com nosso conhecimento, não existe trabalho publicado recentemente que tenha se proposto a desenvolver métodos de estimação robustos para modelos de regressão beta nos quais a estrutura de regressão é flexibilizada para contemplar formas não lineares. É importante ressaltar que o conceito de linearidade que será considerado neste trabalho diz respeito à estrutura de regressão vinculada aos parâmetros dos modelos e não à relação entre a média da variável resposta e as variáveis preditoras. Nesse sentido, o presente trabalho se propõe a replicar os métodos de estimação robustos referentes ao SMLE e LSMLE, estendendo-os a modelos de regressão beta não lineares, e possibilitando a utilização dos referidos métodos sob modelos nos quais uma estrutura linear de regressão não seja adequada. Além disso, pretende-se estudar o comportamento de tais modelos por meio de suas propriedades teóricas e aplicações a dados reais e simulados.

O presente texto está organizado em 6 capítulos. Nesse primeiro capítulo foi feita uma introdução ao tema e apresentado um breve histórico com os trabalhos mais recentes relacionados à regressão beta robusta, incluindo a delimitação do escopo do estudo desenvolvido na dissertação. No Capítulo 2 é descrito o modelo probabilístico beta e as suas principais características. Adicionalmente, no mesmo capítulo são introduzidos os modelos de regressão lineares e não lineares baseados na distribuição beta, e o processo de estimação dos parâmetros do modelo não linear por meio do método da máxima verossimilhança. No Capítulo 3 são revisados alguns conceitos e medidas referentes à inferência robusta que serão usadas na sequência do trabalho. No quarto capítulo é desenvolvido o processo de estimação por meio do SMLE e LSMLE sob modelos de regressão beta não lineares, e é proposta uma adaptação ao método orientado a dados para seleção da constante de afinação. Neste Capítulo também é apresentado um teste de hipóteses robusto para avaliação da significância dos coeficientes de regressão sob o modelo de regressão beta robusto. Ainda, no Capítulo 4 é feita uma breve introdução sobre *bootstrap* e o detalhamento dos processos de reamostragem utilizados no trabalho e, na sequência, uma explanação referente à implementação computacional efetuada para viabilizar a realização dos estudos e aplicações práticas. No Capítulo 5 são apresentados os resultados obtidos nos estudos de simulação e nas aplicações a dados reais e simulados, objetivando ilustrar características da regressão beta não linear robusta, suas vantagens e situações nas quais é recomendada a sua utilização. Por fim, no Capítulo 6 é feita uma recapitulação dos resultados obtidos no trabalho, incluindo limitações observadas na pesquisa, além de su-

gestões de tópicos decorrentes desta dissertação que podem ser melhor aprofundados em trabalhos futuros.

2 Modelos de regressão baseados na distribuição beta

2.1 A distribuição beta

A distribuição beta é uma família de distribuições de probabilidade contínuas definida com suporte em um intervalo limitado e parametrizada por dois elementos, ambos positivos, aqui denotados por a e b, que aparecem como expoentes na função densidade da variável aleatória e controlam a forma da distribuição. Conforme apresentada por Johnson, Kotz e Balakrishnan (1995), a distribuição beta constitui um caso particular da distribuição de Pearson Tipo I, quando variável resposta é transformada para restringir o suporte ao intervalo (0,1).

A PDF de uma variável aleatória y que segue uma distribuição beta de parâmetros $a > 0$ e $b > 0$ é definida por

$$f(y; a,b) = \frac{y^{a-1}(1-y)^{b-1}}{B(a,b)}, \quad 0 < y < 1, \tag{2.1.1}$$

em que $B(a,b)$ é a função beta dada por

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

e $\Gamma(\cdot)$ é a função gama definida por

$$\Gamma(z) = \int_0^\infty u^{z-1}e^{-u}du,$$

sendo z um número complexo cuja parte real é estritamente positiva. Assim, a expressão (2.1.1) pode ser reescrita como

$$f(y; a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1}, \tag{2.1.2}$$

A função de distribuição acumulada (*cumulative distribution function*; CDF) da distribuição beta é definida por

$$\begin{aligned} F(y; a,b) &= \int_{-\infty}^{y} f(t; a,b)dt = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^y t^{a-1}(1-t)^{b-1}dt \\ &= \frac{1}{B(a,b)}\int_0^y t^{a-1}(1-t)^{b-1}dt = \frac{B_y(a,b)}{B(a,b)}, \end{aligned}$$

em que $0 < y < 1$ e $B_y(a,b) = \int_0^y t^{a-1}(1-t)^{b-1}dt$ é, segundo Johnson, Kotz e Balakrishnan

(1995), conhecida por função beta incompleta.

A PDF em (2.1.2) pode ser reescrita como

$$\begin{aligned} f(y;a,b) &= \exp\{\log[f(y;a,b)]\}\mathcal{I}_{[y\in(0,1)]} \\ &= \exp\left\{\log\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1}\right]\right\}\mathcal{I}_{[y\in(0,1)]} \\ &= \exp\left\{\log\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]+\log\left[y^{a-1}\right]+\log\left((1-y)^{b-1}\right)\right\}\mathcal{I}_{[y\in(0,1)]} \\ &= \exp\{(a-1)\log(y)+(b-1)\log(1-y)-\log(B(a,b))\}\mathcal{I}_{[y\in(0,1)]}, \end{aligned} \tag{2.1.3}$$

em que $\mathcal{I}_{[y\in(0,1)]}$ é uma função indicadora que assume valor 1 quando $y \in (0,1)$ e 0, caso contrário. Considerando $\boldsymbol{\theta} = (a\ \ b)^\top$ e fazendo $\omega_1(\boldsymbol{\theta}) = a-1$, $\omega_2(\boldsymbol{\theta}) = b-1$, $T_1(y) = \log(y)$, $T_2(y) = \log(1-y)$, $c(\boldsymbol{\theta}) = B(a,b)$ e $h(y) = \mathcal{I}_{[y\in(0,1)]}$, obtemos que a expressão em (2.1.3) se reduz à

$$\begin{aligned} f(y;a,b) &= \exp\{\omega_1(\boldsymbol{\theta})T_1(y)+\omega_2(\boldsymbol{\theta})T_2(y)-c(\boldsymbol{\theta})\}h(y) \\ &= \exp\left\{\sum_{i=1}^{s}\omega_i(\boldsymbol{\theta})T_i(y)-c(\boldsymbol{\theta})\right\}h(y), \end{aligned} \tag{2.1.4}$$

em que $s = 2$ é a quantidade de parâmetros envolvidos. Observa-se que $\omega_i(\boldsymbol{\theta})$, i =1, 2, e $c(\boldsymbol{\theta}) \geq 0$ são funções que dependem somente dos parâmetros a e b, e $T_i(y)$, i =1, 2, e $h(y)$, funções que dependem somente de y. Famílias de distribuições cujas PDFs possam ser reescritas na forma da expressão em (2.1.4) são ditas pertencer à família exponencial s-dimensional. Portanto, a distribuição beta faz parte da família exponencial bidimensional. Famílias exponenciais são de particular interesse na Estatística pois apresentam propriedades matemáticas úteis, além de estarem ligadas a conceitos importantes tais como suficiência e redução de dados (CASELLA; BERGER, 2011).

A distribuição beta é bastante flexível a depender dos valores assumidos pelos parâmetros a e b. Esta pode exibir uma infinidade de formas, sendo amplamente utilizada pra modelar o comportamento de diversos tipos de fenômenos aleatórios, desde que a variável de interesse assuma valores limitados ao intervalo contínuo (0,1). Não obstante, a distribuição beta é também aplicável a fenômenos que produzem valores no intervalo contínuo (c,d), com c e d constantes reais. Para tanto, aplica-se a transformação $(y-c)/(d-c)$ para representar esse intervalo contínuo dentro do suporte exigido para a distribuição beta.

Adotando a PDF em (2.1.2), os momentos de ordem n centrados em zero, com $n = 1,2,3,\ldots$, podem ser obtidos diretamente pela definição

$$\mathrm{E}(y^n) = \int_0^1 y^n f(y;a,b)dy.$$

Assim, temos que

$$\begin{aligned}
\mathrm{E}(y^n) &= \frac{1}{B(a,b)}\int_0^1 y^{(a+n)-1}(1-y)^{b-1}dy \\
&= \frac{B(a+n,b)}{B(a,b)}\int_0^1 \frac{1}{B(a+n,b)}y^{(a+n)-1}(1-y)^{b-1}dy \\
&= \frac{B(a+n,b)}{B(a,b)} \\
&= \frac{\Gamma(a+n)\Gamma(b)}{\Gamma(a+n+b)}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{\Gamma(a+n)}{\Gamma(a+b+n)}\frac{\Gamma(a+b)}{\Gamma(a)}.
\end{aligned}$$

Utilizando as propriedades da função gama, podemos reescrever a expressão anterior como

$$\begin{aligned}
\mathrm{E}(y^n) &= \frac{(a+n-1)(a+n-2)\cdots(a+1)a\Gamma(a)}{(a+b+n-1)(a+b+n-2)\cdots(a+b+1)(a+b)\Gamma(a+b)}\frac{\Gamma(a+b)}{\Gamma(a)} \\
&= \frac{(a+n-1)(a+n-2)\cdots(a+1)a}{(a+b+n-1)(a+b+n-2)\cdots(a+b+1)(a+b)} \\
&= \prod_{r=0}^{n-1}\frac{a+r}{a+b+r}. \qquad (2.1.5)
\end{aligned}$$

Tomando $n = 1{,}2$, obtemos os dois primeiros momentos de y, por meio dos quais chegamos a expressões fechadas para, respectivamente, a média e a variância da variável aleatória y. Portanto, obtemos que

$$\mathrm{E}(y) = \prod_{r=0}^{0}\frac{a+r}{a+b+r} = \frac{a}{a+b},$$

$$\begin{aligned}
\mathrm{Var}(y) &= \mathrm{E}(y^2) - [\mathrm{E}(y)]^2 \\
&= \left(\frac{a+0}{a+b+0}\right)\left(\frac{a+1}{a+b+1}\right) - \left(\frac{a}{a+b}\right)^2 \\
&= \left[\frac{a(a+1)}{(a+b)(a+b+1)}\right] - \left(\frac{a}{a+b}\right)^2 \\
&= \frac{ab}{(a+b)^2(a+b+1)}.
\end{aligned}$$

Também é possível obter a média e a variância por meio da função geradora de momentos $M(t)$, que é dada por $\mathrm{E}(\mathrm{e}^{ty})$, $t = 1{,}2{,}...$ (ROSS, 2009). Com isso, expandindo

$\mathrm{E}(\mathrm{e}^{ty})$ como série de Taylor, temos que

$$\begin{aligned}\mathrm{E}(\mathrm{e}^{ty}) &= \mathrm{E}\left(\sum_{n=0}^{\infty}\frac{t^n y^n}{n!}\right)\\ &= \sum_{n=0}^{\infty}\mathrm{E}\left(y^n\right)\frac{t^n}{n!}\\ &= \mathrm{E}\left(y^0\right)\frac{t^0}{0!}+\sum_{n=1}^{\infty}\mathrm{E}\left(y^n\right)\frac{t^n}{n!}\\ &= 1+\sum_{n=1}^{\infty}\mathrm{E}\left(y^n\right)\frac{t^n}{n!}. \end{aligned} \tag{2.1.6}$$

Substituindo o resultado em (2.1.5) em (2.1.6) obtemos que

$$M(t)=\mathrm{E}(\mathrm{e}^{ty})=1+\sum_{n=1}^{\infty}\left(\prod_{r=0}^{n-1}\frac{a+r}{a+b+r}\right)\frac{t^n}{n!}. \tag{2.1.7}$$

Com o resultado (2.1.7) pode-se chegar a outros momentos úteis para obtenção de medidas importantes sobre a distribuição. Por exemplo, Johnson, Kotz e Balakrishnan (1995) apresentam expressões fechadas para a assimetria e curtose que são necessárias para o terceiro e quarto momentos da distribuição beta.

Ferrari e Cribari-Neto (2004) propuseram uma reparametrização da distribuição beta reescrevendo a PDF (2.1.2) por meio de novos parâmetros que representam a média e a precisão de y. Tal alteração objetivou definir uma estrutura de regressão para modelar a média μ_t de uma variável resposta y_t que seja distribuída segundo uma distribuição beta. Além disso, para viabilizar a modelagem da média μ_t foi necessário estabelecer um parâmetro ϕ que representasse a precisão da distribuição beta.

Nesse sentido, toma-se $\mu=\mathrm{E}(y)=a/(a+b)$ e $\phi=a+b$, resultando em $a=\mu\phi$ e $b=\phi-\mu\phi=(1-\mu)\phi$ e, consequentemente, na seguinte expressão para a PDF da distribuição beta reparametrizada:

$$f(y;\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},\ \ 0<y<1. \tag{2.1.8}$$

A CDF da distribuição beta reparametrizada é da forma

$$F(y;\mu,\phi)=\int_{-\infty}^{y}f(t;\mu,\phi)dt=\frac{B_Y(\mu\phi,(1-\mu)\phi)}{B(\mu\phi,(1-\mu)\phi)}.$$

Denotaremos por $y\sim\mathcal{B}(\mu,\phi)$ uma variável aleatória y que possui distribuição beta com PDF na forma (2.1.8). Conforme mencionado anteriormente, a distribuição beta é bas-

tante flexível, resultando em grande potencial para modelar dados limitados ao intervalo (0,1). Na Figura 1 são apresentadas curvas da PDF da distribuição beta, considerando diferentes valores para os parâmetros μ e ϕ. Percebe-se que as curvas podem apresentar diferentes formas a depender dos valores assumidos pelos parâmetros. Quando $\mu = 0{,}5$ e $\phi \neq 2$, as curvas apresentam formas simétricas e unimodais. Para $\mu \neq 0{,}5$ as formas apresentadas são assimétricas podendo ser unimodais, em formas de J ou J invertido. Para $\mu = 0{,}5$ e $\phi < 2$, a curva assume a forma de U. Quando $\mu = 0{,}5$ e $\phi = 2$, a função densidade da distribuição beta se reduz à da distribuição uniforme padrão.

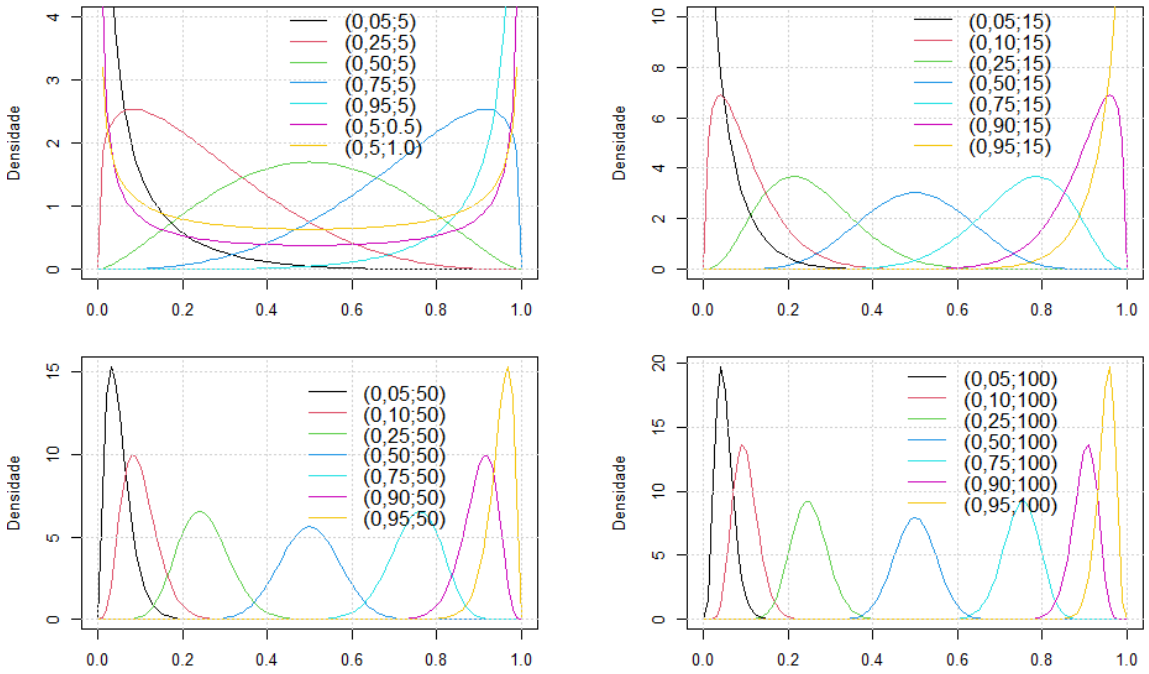

Figura 1 Curvas para a PDF da distribuição beta reparametrizada para diferentes valores de (μ, ϕ).

Sob a nova parametrização, a variância da variável aleatória y passa a ser

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

em que $V(\mu) = \mu(1-\mu)$. Nota-se que, para média μ fixa, a variância de y diminui à medida que o valor de ϕ aumenta. Em contrapartida, valores baixos de ϕ resultam em valores altos para a variância de y. Por esta razão, ϕ é tido como parâmetro de precisão da distribuição beta reparametrizada. Esse resultado também pode ser visualizado por

meio da Figura 2 em que estão representadas algumas curvas para a PDF (2.1.8), com μ fixado em 0,5 e diferentes valores do parâmetro de precisão ϕ.

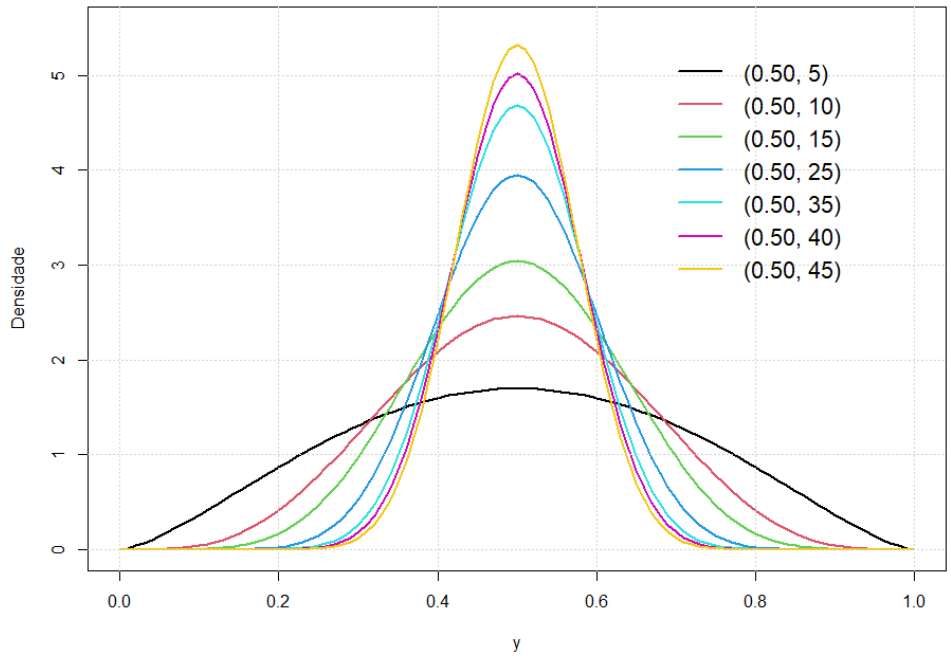

Figura 2 Curvas para a PDF da distribuição beta reparametrizada para μ fixo e diferentes valores de ϕ.

2.2 Regressão linear

A reparametrização da distribuição beta introduzida por Ferrari e Cribari-Neto (2004) viabilizou a sua utilização em modelos de regressão. Dada a expressão (2.1.8), segundo Ferrari e Cribari-Neto (2004) o modelo de regressão beta é obtido assumindo que, para n realizações independentes de uma variável aleatória y com distribuição beta, a média μ_t de cada observação y_t pode ser escrita como

$$g_\mu(\mu_t) = \sum_{i=1}^{p_1} x_{ti}\beta_i = \boldsymbol{X_t}^\top \boldsymbol{\beta}, \tag{2.2.1}$$

em que $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_{p_1})^\top \in \mathbb{R}^{p_1}$ é um vetor de parâmetros desconhecidos associados à média, $\boldsymbol{X}_t = (x_{t1}, x_{t2}, \ldots, x_{tp_1})^\top \in \mathbb{R}^{p_1}$ é o vetor de valores conhecidos das p_1 variáveis explicativas (covariáveis) para a t-ésima observação ($t = 1,2,\ldots,n$), e $g_\mu(\cdot)$ é uma função de ligação contínua, estritamente monótona e duas vezes diferenciável. O principal objetivo associado a $g_\mu(\cdot)$ é restringir μ_t ao suporte da distribuição beta que é o intervalo contínuo (0,1).

Existem diversas opções para a função de ligação $g_\mu(\cdot)$ que atendem aos requisitos mencionados. A rigor, a inversa da CDF de qualquer distribuição contínua poderia ser utilizada, entretanto, as funções de ligação mais citadas e utilizadas (FERRARI; CRIBARI-NETO, 2004; OSPINA, 2004; PEREIRA, 2010) estão a seguir:

- função logito: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$, cuja inversa corresponde à CDF da distribuição logística padrão;
- função probit: $g(\mu) = \Phi^{-1}(\mu)$, em que $\Phi(\cdot)$ é a CDF da distribuição normal padrão;
- função log-log: $g(\mu) = -\log[-\log(\mu)]$, em que $g^{-1}(\mu)$ é a CDF da distribuição Gumbel padrão (máximo), correspondente a uma das duas formas da distribuição do valor extremo padrão, tipo I (GUMBEL, 1954);
- função complementar log-log: $g(\mu) = \log[-\log(1-\mu)]$, em que $g^{-1}(\mu)$ é a CDF da distribuição Gumbel padrão (mínimo), correspondente a uma segunda forma da distribuição do valor extremo padrão, tipo I (GUMBEL, 1954);
- função Cauchit: $g(\mu) = \tan[\pi(\mu - 0{,}5)]$, cuja inversa corresponde à CDF da distribuição Cauchy padrão.

Apesar de heteroscedástico, o modelo proposto por Ferrari e Cribari-Neto (2004) considera que a precisão é constante para todas as observações, o que nem sempre será apropriado supor. Além disso, no contexto dos modelos lineares generalizados (*generalized linear models*; GLM) introduzidos por Nelder e Wedderburn (1972), existem trabalhos onde são considerados os GLMs duplos, nos quais a média e a precisão são modeladas simultaneamente (NELDER; LEE, 1991; SMYTH; VERBYLA, 1999). Nesse sentido, Smithson e Verkuilen (2006) propuseram uma extensão ao modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004). Sob essa nova abordagem, adicionou-se uma estrutura de regressão para modelar, simultaneamente à média, também o parâmetro de precisão ϕ, por meio da estrutura de regressão

$$g_\phi(\phi_t) = \sum_{j=1}^{p_2} z_{tj}\gamma_j = \boldsymbol{Z_t}^\top \boldsymbol{\gamma}, \tag{2.2.2}$$

em que $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_{p_2})^\top \in \mathbb{R}^{p_2}$ é um vetor de parâmetros desconhecidos associados à precisão, $\boldsymbol{Z_t} = (z_{t1}, z_{t2}, \ldots, z_{tp_2})^\top \in \mathbb{R}^{p_2}$ é o vetor de valores conhecidos das p_2 covariáveis da precisão para a t-ésima observação, e $g_\phi(\cdot)$ é uma função de ligação contínua, estritamente monótona e duas vezes diferenciável. Observa-se que, diferentemente do que ocorre com a média, que deve ser mapeada no domínio da variável resposta, o parâmetro de precisão deve assumir valores estritamente positivos, uma vez que $\mathrm{Var}(y)$ não pode

ser negativa. Dentre as funções de ligação que atendem a esses critérios, são citadas por Smithson e Verkuilen (2006) as funções de ligação abaixo:

- função logaritmo: $g_\phi(\phi) = \log(\phi)$.
- função raiz-quadrada: $g_\phi(\phi) = \sqrt{\phi}$.

O requisito de que as funções de ligação $g_\mu(\cdot)$ e $g_\phi(\cdot)$ sejam duas vezes diferenciáveis viabiliza o processo de estimação, em particular, a obtenção da matriz de informação de Fisher (*Fisher information matrix*; FIM). Tal matriz é necessária para dimensionar a variabilidade assintótica das estimativas dos parâmetros de regressão, conforme será visto mais adiante.

2.3 Regressão não linear

Os modelos propostos por Ferrari e Cribari-Neto (2004) e Smithson e Verkuilen (2006) consideram nas respectivas estruturas de regressão os preditores $\boldsymbol{X_t}^\top\boldsymbol{\beta}$ e $\boldsymbol{Z_t}^\top\boldsymbol{\gamma}$, que representam funções lineares de seus parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$. No entanto, ainda que a estrutura de regressão seja linear nos parâmetros, ao aplicar as funções de ligação mais utilizadas, obtém-se uma relação não linear entre a média da variável resposta y ($\mathrm{E}(y)$) e as variáveis explicativas. É importante ressaltar que o conceito de linearidade que será considerado neste trabalho diz respeito à estrutura de regressão vinculada aos parâmetros dos modelos, independentemente dessa relação não linear entre a média da variável resposta y e as covariáveis. Nesse sentido, chamaremos os modelos introduzidos por Ferrari e Cribari-Neto (2004) e Smithson e Verkuilen (2006) de regressão beta linear com precisão constante e regressão beta linear com precisão variável, respectivamente.

Simas, Barreto-Souza e Rocha (2010), apresentaram uma forma mais geral para o modelo de regressão beta com precisão variável, na qual as estruturas de regressão são descritas por relações não necessariamente lineares de seus parâmetros. Dessa forma, seja y_t, $t = 1, \ldots, n$, uma amostra aleatória tal que $y_t \sim \mathcal{B}(\mu_t,\phi_t)$, a média μ_t e a precisão ϕ_t podem ser escritas, respectivamente, como

$$\begin{aligned} g_{\boldsymbol{\mu}}(\mu_t) &= f_\mu(\boldsymbol{X}_t;\boldsymbol{\beta}) = \eta_{\mu t}, \\ g_{\boldsymbol{\phi}}(\phi_t) &= f_\phi(\boldsymbol{Z}_t;\boldsymbol{\gamma}) = \eta_{\phi t}, \end{aligned} \tag{2.3.1}$$

em que $f_\mu(\cdot;\cdot)$ e $f_\phi(\cdot;\cdot)$ são funções que relacionam os parâmetros e as covariáveis e que podem ser lineares ou não lineares. Observe que as estruturas lineares de regressão em (2.2.1) e (2.2.2) representam um caso particular da forma geral apresentada em (2.3.1), quando $f_\mu(\boldsymbol{X}_t;\boldsymbol{\beta}) = \boldsymbol{X_t}^\top\boldsymbol{\beta}$ e $f_\phi(\boldsymbol{Z}_t;\boldsymbol{\gamma}) = \boldsymbol{Z_t}^\top\boldsymbol{\gamma}$.

Os vetores de parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são desconhecidos e, portanto, devem ser estimados. Para este caso, Simas, Barreto-Souza e Rocha (2010) utilizaram o método da máxima verossimilhança, por meio do qual são estimados os valores dos parâmetros que maximizam a função densidade de probabilidade conjunta da amostra, obtendo-se os MLEs. Tomando $y_1, y_2, \ldots, y_n$ variáveis aleatórias independentes tal que $y_t \sim \mathcal{B}(\mu_t,\phi_t)$, a função de verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ é dada por (SIMAS; BARRETO-SOUZA; ROCHA, 2010)

$$\begin{aligned} \mathrm{L}(\boldsymbol{\theta}) &= \prod_{t=1}^{n} f(y_t; \mu_t,\phi_t) \\ &= \prod_{t=1}^{n} \left[\frac{\Gamma(\phi_t)}{\Gamma(\mu_t\phi_t)\Gamma((1-\mu_t)\phi_t)} y_t^{\mu_t\phi_t - 1}(1-y_t)^{(1-\mu_t)\phi_t - 1} \right], \end{aligned}$$

e o respectivo logaritmo da função de verossimilhança para $\boldsymbol{\theta}$ é

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \log(L(\boldsymbol{\theta})) \\ &= \sum_{t=1}^{n} \log(f(y_t; \mu_t,\phi_t)) \\ &= \sum_{t=1}^{n} \log \left[\frac{\Gamma(\phi_t)}{\Gamma(\mu_t\phi_t)\Gamma((1-\mu_t)\phi_t)} y_t^{\mu_t\phi_t - 1}(1-y_t)^{(1-\mu_t)\phi_t - 1} \right] \\ &= \sum_{t=1}^{n} \ell_t(\mu_t,\phi_t), \end{aligned} \tag{2.3.2}$$

em que

$$\begin{aligned} \ell_t(\mu_t,\phi_t) = \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) \\ + (\mu_t\phi_t - 1)\log(y_t) + [(1-\mu_t)\phi_t - 1]\log(1-y_t), \end{aligned}$$

com $\mu_t = g_{\boldsymbol{\mu}}^{-1}(\eta_{\mu t})$, uma função de $\boldsymbol{\beta}$ e $\boldsymbol{X}_t$, e $\phi_t = g_{\boldsymbol{\phi}}^{-1}(\eta_{\phi t})$, uma função de $\boldsymbol{\gamma}$ e $\boldsymbol{Z}_t$.

Para obter o valor de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ que maximiza a expressão em (2.3.2) podem-se calcular as derivadas parciais de $\ell_t(\mu_t,\phi_t)$ com relação a cada um dos parâmetros em $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ obtendo os vetores escore, aqui representados por $\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ e $\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta})$, respectivamente, e igualar a zero. Ressalta-se que, nesse caso, a média μ_t e a precisão ϕ_t são estimadas indiretamente através de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ por meio das estruturas de regressão associadas a μ_t e ϕ_t, conforme definido em (2.3.1).

As entradas do vetor escore para $\boldsymbol{\beta}$, $\boldsymbol{U}_{\beta_i}(\boldsymbol{\theta})$, em que $i = 1,2,\ldots,p_1$, são dadas

pela expressão

$$\begin{aligned}
\boldsymbol{U}_{\beta_i}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_i} \\
&= \sum_{t=1}^{n} \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{\mu t}} \frac{\partial \eta_{\mu t}}{\partial \beta_i} \\
&= \sum_{t=1}^{n} \Big\{ \phi_t \left[\log(y_t) - \log(1-y_t) - \psi(\mu_t\phi_t) + \psi((1-\mu_t)\phi_t)\right] \Big\} \frac{d\mu_t}{d\eta_{\mu t}} \frac{\partial \eta_{\mu t}}{\partial \beta_i} \\
&= \sum_{t=1}^{n} \phi_t ({y_t}^* - {\mu_t}^*) \frac{1}{g'_\mu(\mu_t)} \frac{\partial \eta_{\mu t}}{\partial \beta_i}, \qquad (2.3.3)
\end{aligned}$$

em que $y_t^* = \text{logito}(y_t) = \log(y_t/(1-y_t))$, $\mu_t^* = \text{E}(y_t^*) = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$, e $\psi(\lambda)$ denota a função digama, isto é, $\psi(\lambda) = d\log\Gamma(\lambda)/d\lambda$.

Seja X uma matriz de dimensão $n \times p_1$ em que cada coluna de X representa os valores conhecidos da i-ésima covariável, $i = 1, 2, \ldots, p_1$, $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^\top$, $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_n^*)^\top$ e $T = \text{diag}\{d\mu_1/d\eta_{\mu 1}, \ldots, d\mu_n/d\eta_{\mu n}\} = \text{diag}\{1/g'_\mu(\mu_1), \ldots, 1/g'_\mu(\mu_n)\}$. Segundo Espinheira, Santos e Cribari-Neto (2017), definindo $J_\mu = \partial\boldsymbol{\eta}_\mu/\partial\boldsymbol{\beta}$, uma matriz de dimensão $n \times p_1$, e $\Phi = \text{diag}\{\phi_1, \ldots, \phi_n\}$, então o vetor escore $\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ pode ser representado por

$$\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = J_\mu^\top \Phi T(\boldsymbol{y}^* - \boldsymbol{\mu}^*).$$

As entradas do vetor escore para $\boldsymbol{\gamma}$, $\boldsymbol{U}_{\gamma_j}(\boldsymbol{\theta})$, com $j = 1, 2, \ldots, p_2$, são dadas por

$$\begin{aligned}
\boldsymbol{U}_{\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} \\
&= \sum_{t=1}^{n} \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{\phi t}} \frac{\partial \eta_{\phi t}}{\partial \gamma_j} \\
&= \sum_{t=1}^{n} \Big\{ \mu_t[\log(y_t) - \log(1-y_t) - \psi(\mu_t\phi_t) \\
&\quad + \psi((1-\mu_t)\phi_t)] + \log(1-y_t) + \psi(\phi_t) - \psi[(1-\mu_t)\phi_t] \Big\} \frac{d\phi_t}{d\eta_{\phi t}} \frac{\partial \eta_{\phi t}}{\partial \gamma_j} \\
&= \sum_{t=1}^{n} \Big\{ \mu_t(y_t^* - \mu_t^*) + \log(1-y_t) + \psi(\phi_t) - \psi((1-\mu_t)\phi_t) \Big\} \frac{d\phi_t}{d\eta_{\phi t}} \frac{\partial \eta_{\phi t}}{\partial \gamma_j} \\
&= \sum_{t=1}^{n} a_t \frac{1}{g'_\phi(\phi_t)} \frac{\partial \eta_{\phi t}}{\partial \gamma_j}, \qquad (2.3.4)
\end{aligned}$$

em que $a_t = \mu_t(y_t^* - \mu_t^*) + \log(1-y_t) + \psi(\phi_t) - \psi((1-\mu_t)\phi_t)$.

Seja $J_\phi = \partial\boldsymbol{\eta}_\phi/\partial\boldsymbol{\gamma}$, uma matriz de dimensão $n \times p_2$, $H = \text{diag}\{d\phi_1/d\eta_{\phi 1}, \ldots,$

$d\phi_n/d\eta_{\phi n}\} = \text{diag}\{1/g'_\phi(\phi_1), \ldots, 1/g'_\phi(\phi_n)\}$ e $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$, temos que o vetor escore para $\boldsymbol{\gamma}$ é

$$\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) = J_\phi^\top H \boldsymbol{a}.$$

O MLE para $\boldsymbol{\theta}$, denotado por $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\gamma}}^\top)^\top$, pode ser obtido resolvendo o sistema de equações

$$\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{0},$$

$$\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) = \boldsymbol{0},$$

com relação a $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Observe que não é possivel explicitar os MLEs dos parâmetros de regressão $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, denotados por $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$, de forma analítica, sendo necessário recorrer a métodos iterativos de estimação, tais como o Broyden-Fletcher-Goldfarb-Shanno (BGFS). Maiores informações sobre métodos de otimização numérica podem sem consultados em Press et al. (1992). Tais métodos de otimização necessitam de estimativas iniciais para o procedimento iterativo. Simas, Barreto-Souza e Rocha (2010) sugerem obtê-las a partir do modelo de regressão não linear normal com estruturas de regressão

$$\begin{aligned} g_{\boldsymbol{\mu}}(\mu_t) &= f_\mu(\boldsymbol{X}_t; \boldsymbol{\beta}), \\ g_{\boldsymbol{\phi}}(\sigma_t^{-2}) &= f_\phi(\boldsymbol{Z}_t; \boldsymbol{\gamma}), \end{aligned} \tag{2.3.5}$$

para o qual é assumido que y_t segue umas distribuição normal com média μ_t e variância σ_t^2, ou seja, $y_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$. Os valores estimados de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ serão as estimativas iniciais $\hat{\boldsymbol{\beta}}^{(0)}$ e $\hat{\boldsymbol{\gamma}}^{(0)}$. Observe que o modelo em (2.3.5) representa uma regressão não linear normal, utilizando as funções de ligação $g_{\boldsymbol{\mu}}$ e $g_{\boldsymbol{\phi}}$. O ajuste do modelo (2.3.5) pode ser efetuado por meio da biblioteca `nlme` (PINHEIRO et al., 2017) do *software* `R` (R Core Team, 2024).

Conforme Espinheira, Santos e Cribari-Neto (2017), a FIM dos parâmetros, aqui denotada por K, é dada por

$$K = K(\boldsymbol{\theta}) = \begin{bmatrix} K_{\boldsymbol{\beta}\boldsymbol{\beta}} & K_{\boldsymbol{\beta}\boldsymbol{\gamma}} \\ K_{\boldsymbol{\gamma}\boldsymbol{\beta}} & K_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} -E\left(\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right) & -E\left(\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^\top}\right) \\ -E\left(\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^\top}\right) & -E\left(\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top}\right) \end{bmatrix}, \tag{2.3.6}$$

sendo $K_{\boldsymbol{\beta}\boldsymbol{\beta}} = J_\mu^\top \Phi W J_\mu$, $K_{\boldsymbol{\beta}\boldsymbol{\gamma}} = K_{\boldsymbol{\gamma}\boldsymbol{\beta}}^\top = J_\mu^\top C T H J_\phi$, $K_{\boldsymbol{\gamma}\boldsymbol{\gamma}} = J_\phi^\top D J_\phi$, $W = \text{diag}\{w_1, \ldots, w_n\}$,

$C = \text{diag}\{c_1, \ldots, c_n\}$, $D = \text{diag}\{d_1, \ldots, d_n\}$, e

$$
\begin{aligned}
w_t &= \phi_t^2 \left[\psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)\right] \left(\frac{1}{g'_\mu(\mu_t)}\right)^2, \\
c_t &= \phi_t\{\psi'(\mu_t\phi_t)\mu_t - \psi'((1-\mu_t)\phi_t)(1-\mu_t)\}, \\
d_t &= \left[\psi'(\mu_t\phi_t){\mu_t}^2 + \psi'((1-\mu_t)\phi_t)(1-\mu_t)^2 - \psi'(\phi_t)\right] \left(\frac{1}{g'_\phi(\phi_t)}\right)^2.
\end{aligned}
$$

Sob algumas condições de regularidade e para n suficientemente grande, Simas, Barreto-Souza e Rocha (2010) mencionam que a distribuição conjunta aproximada de $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\gamma}}^\top)^\top$ é normal $(k+q)$-variada, isto é

$$
\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} \underset{n\to\infty}{\overset{D}{\longrightarrow}} N_{p_1+p_2}\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix}; K^{-1}\right), \tag{2.3.7}
$$

em que K^{-1} é a inversa da FIM, que é da forma

$$
K^{-1} = K^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} K^{\beta\beta} & K^{\beta\gamma} \\ K^{\gamma\beta} & K^{\gamma\gamma} \end{bmatrix}
$$

em que

$$
\begin{aligned}
K^{\boldsymbol{\beta\beta}} &= \left(K_{\boldsymbol{\beta\beta}} - K_{\boldsymbol{\beta\gamma}}K_{\boldsymbol{\gamma\gamma}}^{-1}K_{\boldsymbol{\gamma\beta}}\right)^{-1} = \left(J_\mu^\top \Phi W J_\mu - J_\mu^\top C T H J_\phi (J_\phi^\top D J_\phi)^{-1} J_\mu^\top C T H J_\phi\right)^{-1} \\
K^{\boldsymbol{\beta\gamma}} &= (K^{\boldsymbol{\gamma\beta}})^\top = -K^{\boldsymbol{\beta\beta}}K_{\boldsymbol{\beta\beta}}K_{\boldsymbol{\gamma\gamma}}^{-1} = -K^{\boldsymbol{\beta\beta}} J_\mu^\top \Phi W J_\mu (J_\phi^\top D J_\phi)^{-1} \\
K^{\boldsymbol{\gamma\beta}} &= (K^{\boldsymbol{\beta\gamma}})^\top = -K_{\boldsymbol{\gamma\gamma}}^{-1}K_{\boldsymbol{\beta\beta}}K^{\boldsymbol{\beta\beta}} = -(J_\phi^\top D J_\phi)^{-1} J_\mu^\top \Phi W J_\mu K^{\boldsymbol{\beta\beta}} = K^{\boldsymbol{\beta\gamma}}, \\
K^{\boldsymbol{\gamma\gamma}} &= K_{\boldsymbol{\gamma\gamma}}^{-1} + K_{\boldsymbol{\gamma\gamma}}^{-1}K_{\boldsymbol{\gamma\beta}}K^{\boldsymbol{\beta\beta}}K_{\boldsymbol{\beta\gamma}}K_{\boldsymbol{\gamma\gamma}}^{-1} \\
&= (J_\phi^\top D J_\phi)^{-1} + (J_\phi^\top D J_\phi)^{-1} J_\mu^\top C T H J_\phi K^{\boldsymbol{\beta\beta}} J_\mu^\top C T H J_\phi (J_\phi^\top D J_\phi)^{-1} \\
&= (J_\phi^\top D J_\phi)^{-1}[I_{p_2} + (J_\mu^\top C T H J_\phi) K^{\boldsymbol{\beta\beta}} J_\mu^\top C T H J_\phi (J_\phi^\top D J_\phi)^{-1}]
\end{aligned}
$$

com I_{p_2} denotando uma matriz identidade de ordem p_2.

Por meio da propriedade de normalidade assintótica dos MLEs é possível dimensionar, também de forma assintótica, a variabilidade desses estimadores. Com isso, sob condições usuais de regularidade, pode-se demonstrar que, conforme definido em (2.3.7), para a s-ésima componente de $\hat{\boldsymbol{\theta}}$, $\hat{\theta}_s$, obtemos que

$$
\left(\hat{\theta}_s - \theta_s\right)\left[K(\boldsymbol{\theta})^{ss}\right]^{-1/2} \underset{n\to\infty}{\overset{D}{\longrightarrow}} N(0,1),
$$

em que $K(\boldsymbol{\theta})$ é a FIM de $\boldsymbol{\theta}$ e $K(\boldsymbol{\theta})^{ss}$ é o (s,s)-ésimo elemento de $K(\boldsymbol{\theta})^{-1}$. Esse resultado permite a construção de intervalos de confiança aproximados para diversas grandezas relacionadas aos modelos, como por exemplo, para cada parâmetro estimado. Dessa

forma, considerando um nível de confiança de $100(1-\alpha)\%$, um intervalo de confiança aproximado para a s-ésima componente do vetor de parâmetros $\boldsymbol{\theta}$ é

$$\left(\hat{\theta}_s - z_{1-\frac{\alpha}{2}}\sqrt{\widehat{K(\hat{\boldsymbol{\theta}})}^{ss}} \;;\; \hat{\theta}_s + z_{1-\frac{\alpha}{2}}\sqrt{\widehat{K(\hat{\boldsymbol{\theta}})}^{ss}}\right),$$

em que $z_{1-\frac{\alpha}{2}}$ representa o quantil da distribuição Normal padrão tal que $P(Z \leq z_{1-\frac{\alpha}{2}}) = 1-\alpha/2$, com $Z \sim N(0,1)$.

3 Medidas de robustez

Considerando o ajuste por meio do método da máxima verossimilhança, os modelos apresentados no Capítulo 2 se mostram bastante úteis para resolver uma grande quantidade de problemas e têm sido amplamente utilizados para modelagem de dados contínuos limitados, especialmente aqueles contidos no intervalo unitário. Não obstante, conforme já mencionado anteriormente, os MLEs sob alguns modelos probabilísticos podem ser desproporcionalmente influenciados pela presença de observações discrepantes nos dados, situação na qual não são considerados robustos.

Neste capítulo, que é baseado nos trabalhos de Ribeiro (2020) e Queiroz (2022), serão apresentados alguns conceitos referentes à inferência robusta que serão utilizados na sequência deste trabalho, objetivando melhor caracterizar o conceito de robustez que está sendo considerado.

3.1 Conceitos preliminares

Sejam $y_t, t = 1, \ldots, n$, variáveis aleatórias independentes e identicamente distribuídas (*independent and identically distributed*; IID) segundo uma família de distribuições paramétricas $\mathcal{F}_{\boldsymbol{\Theta}} = \{F_{\boldsymbol{\theta}}, \theta \in \Theta \subset \mathbb{R}^p\}$, $p > 1$, em que Θ é o espaço paramétrico de $\boldsymbol{\theta}$ e $f_{\boldsymbol{\theta}}$ é a PDF de y_t. Seja $\mathcal{I}(\cdot)$ a função indicadora, consideremos estimadores para o parâmetro $\boldsymbol{\theta}$ que dependam dos dados y_t somente por meio da função de distribuição empírica (*empirical distribution function*; EDF), dada por

$$F_n(v) = \frac{1}{n}\sum_{t=1}^{n} \mathcal{I}(y_t < v),$$

isto é satisfazem a relação

$$\boldsymbol{T}_n(y_1, \ldots, y_n) = \boldsymbol{T}(F_n). \tag{3.1.1}$$

Neste caso, chamamos $\boldsymbol{T}(F_n)$ de estimador funcional para $\boldsymbol{\theta}$. Adicionalmente, dizemos que $\boldsymbol{T}(F_n)$ também é um estimador Fisher-consistente para $\boldsymbol{\theta}$ se, além de satisfazer a relação em (3.1.1), também satisfaz (KALLIANPUR; RAO, 1955)

$$\boldsymbol{T}(F_{\boldsymbol{\theta}}) = \boldsymbol{\theta}, \quad \forall \ \boldsymbol{\theta} \in \Theta.$$

Portanto, a propriedade de Fisher-consistência de um estimador assegura que o mesmo atingirá o verdadeiro valor do parâmetro estimado, no caso $\boldsymbol{\theta}$, quando este é calculado sob a distribuição populacional dos dados, ou seja, sob $F_{\boldsymbol{\theta}}$.

3.2 Função de influência

A função de influência (*influence function*; IF) é uma medida muito conhecida e utilizada para avaliar a robustez de um estimador funcional. Segundo Hampel et al. (2011), seja $F_{h,y} = (1-h)F_{\boldsymbol{\theta}} + h\Delta_y$ a CDF contaminada após a introdução de uma perturbação infinitesimal h no ponto y, então a IF do estimador $\boldsymbol{T}$ em $F_{\boldsymbol{\theta}}$ é dada por

$$\begin{aligned}\text{IF}(y;\boldsymbol{T},F_{\boldsymbol{\theta}}) &= \frac{\partial}{\partial h}\left[\boldsymbol{T}(F_{h,y})\right]|_{h=0} \\ &= \lim_{h\to 0}\frac{\boldsymbol{T}(F_{h,y}) - \boldsymbol{T}(F_{\boldsymbol{\theta}})}{h} \\ &= \lim_{h\to 0}\frac{\boldsymbol{T}((1-h)F_{\boldsymbol{\theta}} + h\Delta_y) - \boldsymbol{T}(F_{\boldsymbol{\theta}})}{h},\end{aligned} \tag{3.2.1}$$

em que $\boldsymbol{T}(F)$ é o estimador para $\boldsymbol{\theta}$ avaliado sob a CDF F e Δ_y é a medida de probabilidade que coloca toda a massa em y.

Considerando a expressão em (3.2.1), pode-se interpretar a IF como sendo o efeito causado ao estimador $\boldsymbol{T}$ após uma contaminação infinitesimal h no ponto y. Desse modo, se mesmo uma perturbação mínima, tendendo a zero, for suficiente para afetar desproporcionalmente o estimador $\boldsymbol{T}$, então poderá ser um indicativo de que este estimador é sensível a pequenas variações no ponto y e, portanto, não robusto a observações atípicas. Conforme Hampel et al. (2011), a IF quantifica o viés assintótico no estimador $\boldsymbol{T}$ causado pela perturbação nos dados, e este será considerado qualitativamente robusto se possui IF limitada para todo y pertencente ao suporte da distribuição.

Uma medida de robustez desenvolvida a partir da IF é a sensibilidade a erro grosseiro não padronizada (*unstandardized gross-error sensitivity*; UGES). Tal medida é dada por

$$\gamma_u^* = \sup_y \parallel \text{IF}(y;\boldsymbol{T},F_{\boldsymbol{\theta}}) \parallel,$$

em que $\parallel \cdot \parallel$ denota a norma euclidiana. A medida γ_u^* representa o viés máximo causado no estimador $\boldsymbol{T}$ em decorrência da contaminação infinitesimal introduzida. Assim, a medida de UGES pode ser interpretada como um limite superior para o viés do estimador $\boldsymbol{T}$ sob contaminação, e é desejável que tal medida seja finita. Além disso, observa-se que se o estimador $\boldsymbol{T}$ contiver ao menos uma entrada cuja respectiva IF divirja, então γ_u^* será infinito e $\boldsymbol{T}$ não será considerado robusto. Estimadores que possuem a UGES finita são denominados B-Robustos (ROUSSEEUW, 1981).

Conforme pontuado por Ribeiro (2020), a medida em γ_u^* não leva em consideração a escala das covariáveis, podendo gerar confundimento quando utilizada em um contexto

de modelos de regressão. Sobre isso, Hampel et al. (2011) apresentou duas propostas de padronização dessa medida, por meios das quais obtém-se valores que são invariantes à escala das covariáveis. A primeira delas é a sensibilidade auto-padronizada (*self-standardized sensitivity*; SSS), que é definida por

$$\gamma_s^* = \sup_y \left\{ \mathrm{IF}(y; \boldsymbol{T}, F_{\boldsymbol{\theta}})^\top \mathrm{V}(\boldsymbol{T}, F_{\boldsymbol{\theta}})^{-1} \mathrm{IF}(y; \boldsymbol{T}, F_{\boldsymbol{\theta}}) \right\}^{\frac{1}{2}}, \tag{3.2.2}$$

em que $\mathrm{V}(\boldsymbol{T}, F_{\boldsymbol{\theta}})$ é a matriz de covariâncias assintótica de $\boldsymbol{T}$. Observa-se que, de fato, a expressão em (3.2.2) conduz a uma padronização da medida γ_u^*, efetuada por meio de $\mathrm{V}(\boldsymbol{T}, F_{\boldsymbol{\theta}})$. Caso o estimador $\boldsymbol{T}$ possua baixa eficiência assintótica (*asymptotic efficiency*; AE), então serão obtidos baixos valores para γ_s^*.

A segunda proposta de Hampel et al. (2011) introduz a medida de sensibilidade padronizada pela informação (*information-standardized sensitivity*; ISS), que é expressa por

$$\gamma_{is}^* = \sup_y \left\{ \mathrm{IF}(y; \boldsymbol{T}, F_{\boldsymbol{\theta}})^\top \mathrm{K}(\boldsymbol{T}(F_{\boldsymbol{\theta}}), F_{\boldsymbol{\theta}})^{-1} \mathrm{IF}(y; \boldsymbol{T}, F_{\boldsymbol{\theta}}) \right\}^{\frac{1}{2}}, \tag{3.2.3}$$

em que $\mathrm{K}(\boldsymbol{T}(F_{\boldsymbol{\theta}}), F_{\boldsymbol{\theta}})$ é a matriz de covariâncias assintótica do MLE para $\boldsymbol{\theta}$, avaliada sob o estimador $\boldsymbol{T}$. Ressalta-se que se $\boldsymbol{T}$ for o MLE para $\boldsymbol{\theta}$, então as medidas em (3.2.2) e (3.2.3) serão iguais.

3.3 M-Estimadores

Com o objetivo de obter um estimador que fosse robusto à presença de observações atípicas, Huber (1964) desenvolveu um método de estimação que se baseou na generalização do procedimento de estimação por máxima verossimilhança. Segundo o método de máxima verossmilhança, dada $y_1, \ldots, y_n$ uma amostra aleatória tal que $y_t, t = 1, \ldots, n$, possua densidade $f_{\boldsymbol{\theta}}(y_t)$, $\boldsymbol{\theta} \in \Theta$, o logaritmo da função de verossimilhança é expresso por

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log(f_{\boldsymbol{\theta}}(y_t)),$$

em que o MLE denotado por $\hat{\boldsymbol{\theta}}$ é equivalente ao valor que maximiza $\ell(\boldsymbol{\theta})$. Portanto, $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} [\ell(\boldsymbol{\theta})]$ ou, equivalentemente, $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} [-\ell(\boldsymbol{\theta})]$. Tal método consiste em substituir a contribuição individual $-\log(f_{\boldsymbol{\theta}}(y_t))$ da t-ésima observação por uma função $\rho(y_t, \boldsymbol{\theta})$, de modo que o estimador proposto seja obtido por meio da relação

$$\widehat{\boldsymbol{\theta}}_M = \boldsymbol{T}(F_n) = \arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{t=1}^{n} \rho(y_t, \boldsymbol{\theta}). \tag{3.3.1}$$

em que $\rho(\cdot, \boldsymbol{\theta})$ é uma função diferenciável tal que $\rho : \mathcal{X} \times \Theta \longrightarrow \mathbb{R}$, com $\mathcal{X}$ denotando o conjunto suporte. Maiores informações sobre as condições adicionais que a função $\rho(\cdot, \boldsymbol{\theta})$ referente a um M-estimador de localização ou escala deve satisfazer podem ser obtidas em Maronna et al. (2019).

A classe de estimadores resultantes desse método e que, portanto, satisfazem (3.3.1), foi denominada de M-estimadores. A equação de estimação associada ao M-estimador $\widehat{\boldsymbol{\theta}}_M$ é dada por

$$\sum_{t=1}^{n} \psi(y_t, \boldsymbol{T}(F_n)) = \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \rho(y_t, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{T}(F_n)} = \mathbf{0}.$$

Observe que ao considerar $\rho(y_t, \boldsymbol{\theta}) = -\log(f_{\boldsymbol{\theta}}(y_t))$, então $\psi(y_t, \boldsymbol{\theta})$ será o negativo da função escore e, portanto, teremos o caso particular referente ao MLE. Conforme Ribeiro (2020), a IF para $\widehat{\boldsymbol{\theta}}_M$ é dada por

$$\begin{aligned} \mathrm{IF}(y; \boldsymbol{T}, F_{\boldsymbol{\theta}}) &= \left[- \int \frac{\partial}{\partial \boldsymbol{\theta}} [\psi(y, \boldsymbol{\theta})]|_{\boldsymbol{\theta}=\boldsymbol{T}(F_{\boldsymbol{\theta}})} dF_{\boldsymbol{\theta}}(y) \right]^{-1} \psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}})) \\ &= \mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1} \psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}})), \end{aligned}$$

em que

$$\mathrm{M}(\psi, F_{\boldsymbol{\theta}}) = \frac{\partial}{\partial \boldsymbol{\theta}} [\psi(y, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{T}(F_{\boldsymbol{\theta}})} dF_{\boldsymbol{\theta}}(y).$$

Nota-se que se algum componente de $\psi(y, \boldsymbol{\theta})$ não for limitado, então a sua IF também não será limitada e, consequentemente, a sua UGES não será finita. Portanto, o estimador relacionado não será considerado B-Robusto. Tomando $\boldsymbol{T} = \hat{\boldsymbol{\theta}}$, tal que $\hat{\boldsymbol{\theta}}$ seja o MLE para $\boldsymbol{\theta}$, a IF fica expressa por

$$\mathrm{IF}(y; \hat{\boldsymbol{\theta}}, F_{\boldsymbol{\theta}}) = K(\boldsymbol{\theta}, F_{\boldsymbol{\theta}})^{-1} \boldsymbol{U}(y, \boldsymbol{\theta}), \tag{3.3.2}$$

em que $K(\boldsymbol{\theta}, F_{\boldsymbol{\theta}})$ é a FIM de $\boldsymbol{\theta}$ sob $F_{\boldsymbol{\theta}}$ e $\boldsymbol{U}(\cdot, \boldsymbol{\theta})$ é o vetor escore para $\boldsymbol{\theta}$. Da mesma forma, se algum componente do vetor escore não for limitado, então a IF não será limitada e, portanto, o estimador $\hat{\boldsymbol{\theta}}$ não será B-Robusto.

Segundo Ribeiro (2020, p. 16, cap. 2), a matriz de variâncias e covariâncias as-

sintóticas de um M-estimador $\boldsymbol{T}$ é dada por

$$\begin{aligned}
\mathrm{V}(\boldsymbol{T}; F_{\boldsymbol{\theta}}) &= \int \mathrm{IF}(y, \boldsymbol{T}, F_{\boldsymbol{\theta}})\mathrm{IF}(y, \boldsymbol{T}, F_{\boldsymbol{\theta}})^{\top} dF_{\boldsymbol{\theta}}(y) \\
&= \int \mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1}\psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}})) \left[\mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1}\psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}}))\right]^{\top} dF_{\boldsymbol{\theta}}(y) \\
&= \int \mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1}\psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}})) \left[\psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}}))\right]^{\top} \left[\mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1}\right]^{\top} dF_{\boldsymbol{\theta}}(y) \\
&= \mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1} \left[\int \psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}})) \left[\psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}}))\right]^{\top} dF_{\boldsymbol{\theta}}(y)\right] \left[\mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1}\right]^{\top} \\
&= \mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1}\mathrm{Q}(\psi, F_{\boldsymbol{\theta}}) \left[\mathrm{M}(\psi, F_{\boldsymbol{\theta}})^{-1}\right]^{\top},
\end{aligned}$$

em que

$$\mathrm{Q}(\psi, F_{\boldsymbol{\theta}}) = \int \psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}})) \left[\psi(y, \boldsymbol{T}(F_{\boldsymbol{\theta}}))\right]^{\top} dF_{\boldsymbol{\theta}}(y).$$

Hampel et al. (2011) demonstra que os M-estimadores gozam da propriedades análogas aos MLEs no que se refere à sua distribuição assintótica. Assim, sob condições de regularidade e para uma amostra n suficientemente grande, vale que

$$\sqrt{n}(\boldsymbol{T}(F_n) - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{D} \mathrm{N}(\mathbf{0}, \mathrm{V}(\boldsymbol{T}, F_{\boldsymbol{\theta}})).$$

Esse resultado permite a obtenção de erros-padrão assintóticos e estimativas intervalares para o M-estimador $\boldsymbol{T}$.

4 Inferência robusta

Ribeiro e Ferrari (2023) demonstraram que o procedimento de estimação por máxima verossimilhança não é robusto para os parâmetros do modelo de regressão beta especificado na Seção 2.2. Tomando o modelo definido na Seção 2.3, temos que a sua IF sob o método de estimação por máxima verossimilhança é dada pela expressão em (3.3.2), ou seja, o produto entre a inversa da FIM, definida em (2.3.6), e o vetor escore para $\boldsymbol{\theta}$, obtido a partir das expressões em (2.3.3) e (2.3.4). Logo, sob a regressão beta não linear, o vetor escore para $\boldsymbol{\theta}$ correspondente a uma única observação y_t é

$$\boldsymbol{U}(y_t, \boldsymbol{\theta}) = \left(\phi_t({y_t}^* - {\mu_t}^*) \frac{1}{g'_\mu(\mu_t)} \frac{\partial \eta_{\mu t}}{\partial \boldsymbol{\beta}^\top}, \quad a_t \frac{1}{g'_\phi(\phi_t)} \frac{\partial \eta_{\phi t}}{\partial \boldsymbol{\gamma}^\top} \right)^\top . \tag{4.0.1}$$

Para que o MLE sob o modelo de regressão beta não seja robusto, basta que um dos componentes da sua IF não seja limitado. Para a regressão beta linear, os limites dos componentes do vetor escore divergem quando y_t tende para os limites do suporte da distribuição beta, ou seja, quando $y_t \to 0$ ou $y_t \to 1$ (RIBEIRO; FERRARI, 2023). Observa-se que esse mesmo resultado também é válido para o caso não linear. Portanto, isto é suficiente para a conclusão de que a respectiva IF não é limitada e, consequentemente, a medida UGES y_u^* não é finita. Portanto, sob o modelo de regressão beta não linear, o MLE não é considerado B-Robusto.

Desse modo, pode-se afirmar que na presença de observações atípicas, não há garantia de que a inferência via máxima verossimilhança produza estimativas robustas para os parâmetros do modelo de regressão beta não linear. Como alternativa à estimação via máxima verossimilhança sob a regressão beta linear, Ribeiro e Ferrari (2023) propuseram o SMLE, obtido com base na maximização de uma reparametrização da função de L_q-verossimilhança introduzida por Ferrari e Yang (2010). Posteriormente, objetivando superar limitações do SMLE, Maluf, Ferrari e Queiroz (2025) propuseram o LSMLE, que consiste em uma adaptação do método introduzido por Ribeiro e Ferrari (2023), porém aplicando a transformação logito na variável resposta.

Neste capítulo, que tem como base os trabalhos de Ribeiro e Ferrari (2023) e Maluf, Ferrari e Queiroz (2025), será apresentado o desenvolvimento do SMLE e do LSMLE para os modelos de regressão beta com precisão variável considerando a generalização das estruturas de regressão para o caso não linear.

4.1 Estimação via L_q-verossimilhança reparametrizada

Sejam $y_t, t = 1, \ldots, n$, observações independentes obtidas a partir do modelo de regressão beta com precisão variável definido na Seção 2.3, indexado por um parâmetro desconhecido $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{p_1+p_2}$. Segundo Ferrari e Yang (2010), a função de L_q-verossimilhança, aqui denotada por $\ell_q(\boldsymbol{\theta})$, é definida por

$$\ell_q(\boldsymbol{\theta}) = \sum_{t=1}^{n} L_q(f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)),$$

em que $f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)$ é a PDF assumida para y_t, e

$$L_q(u) = \begin{cases} (u^{1-q} - 1)/(1-q) & \text{se } q \neq 1, \\ \log(u), & \text{se } q = 1, \end{cases} \tag{4.1.1}$$

é a transformação de Box-Cox (BOX; COX, 1964), com $q \in (0,1]$ sendo uma constante denominada de constante de afinação, que será melhor discutida nas seções seguintes. Observe que o parâmetro da transformação de Box-Cox em (4.1.1), representado por α, é $\alpha = 1 - q$. Dessa forma, o estimador de máxima L_q-verossimilhança (*maximum L_q-likelihood estimator*; ML_qE), aqui denotado por $\overline{\boldsymbol{\theta}}_q$, é obtido a partir da maximização de $\ell_q(\boldsymbol{\theta})$, ou seja, $\overline{\boldsymbol{\theta}}_q = \underset{\boldsymbol{\theta}\in\Theta}{\arg\max}\,[\ell_q(\boldsymbol{\theta})]$ ou, de forma análoga, $\overline{\boldsymbol{\theta}}_q = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min}\,[-\ell_q(\boldsymbol{\theta})]$. Ressalta-se que o MLE é um caso particular do ML_qE, uma vez que para $q = 1$, obtém-se

$$\ell_q(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log(f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)),$$

e, portanto, $\overline{\boldsymbol{\theta}}_q$ será igual ao MLE $\widehat{\boldsymbol{\theta}}$ usual.

Considerando que

$$f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)^{1-q} = \exp\left\{(1-q)\log(f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t))\right\},$$

segue que a equação de estimação associada ao ML_qE fica dada por

$$\boldsymbol{U}_q(\boldsymbol{y}, \boldsymbol{\theta}) = \sum_{t=1}^{n} \boldsymbol{U}(y_t, \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)^{1-q} = \mathbf{0}, \tag{4.1.2}$$

em que $\boldsymbol{U}(y_t, \boldsymbol{\theta})$ é o vetor escore referente à t-ésima observação definido em (4.0.1). Observe que a expressão em (4.1.2) corresponde a um processo de M-estimação tal qual o descrito na Seção 3.3, pois a contribuição individual de cada observação está sendo substituída pela função $\boldsymbol{U}(y_t, \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)^{1-q}$, que corresponde à contribuição individual no MLE, porém ponderada por $f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)^{1-q}$. Assim, a escolha do valor para contante q

controla a ponderação atribuída para cada observação. Nesse sentido, temos que quanto mais distante de um for o valor escolhido para q, mais robusto será o procedimento de estimação, uma vez que as observações tidas como atípicas receberão uma menor ponderação e, consequentemente, contribuirão menos para o processo de estimação.

Observa-se que, exceto para $q = 1$, a função de estimação em (4.1.2) é enviesada, ou seja, $\mathrm{E}[\boldsymbol{U}_q(\boldsymbol{y}, \boldsymbol{\theta})] \neq \mathbf{0}$. Desse modo, o estimador não será Fisher-consistente. Para contornar esse problema e construir um estimador que seja Fisher-consistente, Ribeiro e Ferrari (2023) utilizaram uma reparametrização da função de L_q-verossimilhança introduzida por Ferrari e La Vecchia (2012), que se baseou em uma função de calibração para reescalonar as estimativas de $\boldsymbol{\theta}$. Assim, se a família de distribuições postulada aos dados for fechada sob a transformação potência, então temos garantida a Fisher-consistência do estimador. Dada uma PDF h e uma constante $\omega > 0$, a transformação potência é definida como

$$h^{(\omega)}(y) = \frac{h(y)^\omega}{\int h(y)^\omega dy} \propto h(y)^\omega, \forall\ y \text{ no suporte}, \tag{4.1.3}$$

desde que $\int h(y)^\omega dy < \infty$. Para a família de densidades $\{h_{\boldsymbol{\theta}}(\cdot), \boldsymbol{\theta} \in \Theta\}$, que é fechada sob a transformação potência em (4.1.3), considere uma função contínua inversível $\tau_\omega(\boldsymbol{\theta}) : \Theta \to \Theta$ que satisfaz

$$h_{\tau_\omega(\boldsymbol{\theta})}(y) = h^{(\omega)}_{\boldsymbol{\theta}}(y),$$

para todo y no suporte da distribuição postulada, sendo que este não depende de $\boldsymbol{\theta}$. Assim, a aplicação da transformação potência à densidade $h_{\boldsymbol{\theta}}$ tem como resultado uma densidade $h^{(\omega)}_{\boldsymbol{\theta}}$ pertencente à mesma família de distribuições da qual $h_{\boldsymbol{\theta}}$ pertence, porém, sob uma parametrização diferente, no caso $\tau_\omega(\boldsymbol{\theta})$. Ribeiro e Ferrari (2023) mostraram que a PDF da distribuição beta é fechada sob a transformação potência, desde que $\mu_t\phi_t > 1$ e $(1 - \mu_t)\phi_t > 1$. Em outras palavras, a distribuição beta é fechada sob a transformação potência para todo $\omega > 0$ se a densidade beta $f_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)$ for limitada.

Ferrari e La Vecchia (2012) demonstraram que apesar de $\overline{\boldsymbol{\theta}}_q$ não ser Fisher-consistente para $\boldsymbol{\theta}$, é possível obter um outro estimador por meio do mesmo processo, porém sob a parametrização $\tau_q^{-1}(\boldsymbol{\theta}) = \tau_{1/q}(\boldsymbol{\theta})$ que atende à propriedade de Fisher-consistência. Desse modo, o novo estimador $\widehat{\boldsymbol{\theta}}_q$ é obtido a partir da maximização da função L_q-verossimilhança sob a parametrização $\tau_{1/q}(\boldsymbol{\theta})$, aqui denotada por $\ell^*_q(\boldsymbol{\theta})$, que, sob o modelo de regressão beta não linear, é definida por

$$\ell^*_q(\boldsymbol{\theta}) = \sum_{t=1}^{n} L_q(f_{\tau_{1/q}(\boldsymbol{\theta})}(y_t; \mu_t; \phi_t)) = \sum_{t=1}^{n} L_q(f^{(1/q)}_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)), \tag{4.1.4}$$

em que $f_{\boldsymbol{\theta}}^{(1/q)}(y_t;\mu_t;\phi_t) = f_{\boldsymbol{\theta}}(y_t;\mu_{t,q^{-1}};\phi_{t,q^{-1}})$, para $q \in (0,1)$, com

$$\begin{aligned}\mu_{t,q} &= \phi_{t,q}^{-1}[q(\mu_t\phi_t - 1) + 1] \text{ e}\\ \phi_{t,q} &= q(\phi_t - 2) + 2,\end{aligned}$$

desde que $\mu_{t,q^{-1}} \in (0,1)$ e $\phi_{t,q^{-1}} > 0$, ou, equivalentemente, $\mu_t\phi_t > 1-q$ e $(1-\mu_t)\phi_t > 1-q$. Logo, $f_{\boldsymbol{\theta}}(y_t;\mu_t;\phi_t)$ satisfaz a transformação potência definida em (4.1.3) para todo $\omega = 1/q > 0$, se $\mu_t\phi_t \geq 1$ e $(1-\mu_t)\phi_t \geq 1$, ou seja, se a densidade $f_{\boldsymbol{\theta}}(y_t;\mu_t;\phi_t)$ for limitada.

A expressão equivalente à densidade $f_{\boldsymbol{\theta}}^{(1/q)}(y_t;\mu_t;\phi_t)$ para o caso linear corresponde ao modelo de regressão beta modificado definido por meio da densidade especificada em (2.1.8), com as estruturas de regressão associadas aos parâmetros μ_t e ϕ_t obtidas a partir de (2.2.1) e (2.2.2), respectivamente. Considerando o caso não linear especificado em (2.3.1), os submodelos da média e da precisão associados ao modelo de regressão beta modificado, que aqui será denotado por $f_{\boldsymbol{\theta}}^*(y_t;\mu_t;\phi_t)$, são dados por

$$\begin{aligned}g_\mu^*(\mu_t) &= g_\mu(\mu_{t,q}) = f_\mu(\boldsymbol{X}_t;\boldsymbol{\beta}) = \eta_{\mu t},\\ g_\phi^*(\phi_t) &= g_\phi(\phi_{t,q}) = f_\phi(\boldsymbol{Z}_t;\boldsymbol{\gamma}) = \eta_{\phi t}.\end{aligned} \tag{4.1.5}$$

O estimador proposto por Ribeiro e Ferrari (2023) é obtido por meio da maximação da função L_q-verossimilhança reparametrizada definida em (4.1.4). Neste trabalho, também maximizaremos a expressão em (4.1.4) mas considerando a flexibilização das estruturas de regressão associadas a μ_t e ϕ_t, obtendo a equação de estimação

$$\sum_{t=1}^{n} \boldsymbol{U}^*(y_t,\boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(y_t;\mu_t;\phi_t)^{1-q} = \mathbf{0}, \tag{4.1.6}$$

em que $\boldsymbol{U}^*(y_t,\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\log\left[f_{\boldsymbol{\theta}}^*(y_t;\mu_t;\phi_t)\right]$, com $\nabla\boldsymbol{\theta}$ denotando o gradiente relativo a $\boldsymbol{\theta}$, é o vetor escore modificado para $\boldsymbol{\theta}$ referente à t-ésima observação, dado por

$$\boldsymbol{U}^*(y_t,\boldsymbol{\theta}) = \left(q^{-1}\frac{\phi_{t,q}({y_t}^* - {\mu_t}^*)}{g'_\mu(\mu_{t,q})}\frac{\partial\eta_{\mu t}}{\partial\boldsymbol{\beta}^\top},\quad q^{-1}\frac{\mu_{t,q}(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)}{g'_\phi(\phi_{t,q})}\frac{\partial\eta_{\phi t}}{\partial\boldsymbol{\gamma}^\top}\right)^\top, \tag{4.1.7}$$

em que $y_t^\dagger = \log(1-y_t)$ e $\mu_t^\dagger = \mathrm{E}(y_t^\dagger) = \psi((1-\mu_t)\phi_t) - \psi(\phi_t)$. Observa-se que o fator $f_{\boldsymbol{\theta}}^*(y_t;\mu_t;\phi_t)^{1-q}$ em (4.1.6) funciona como uma ponderação no processo de estimação. Se $q = 1$, o mesmo peso é aplicado a todas as observações e, então, teremos como resultado o MLE usual. Por outro lado, quando $q < 1$, então as observações tidas como atípicas em relação ao esperado para o modelo de regressão beta postulado receberão pesos menores e terão menor influência no estimador final $\widehat{\boldsymbol{\theta}}_q$.

O estimador $\widehat{\boldsymbol{\theta}}_q$, denominado SMLE, goza de propriedades úteis para o processo inferencial, a exemplo da Fisher-consistente para $\boldsymbol{\theta}$ e normalidade assintótica. Nesse

sentido, por se tratar de um M-estimador, então a sua distribuição é assintoticamente normal, com $\widehat{\boldsymbol{\theta}}_q \overset{a}{\sim} \mathrm{N}(\boldsymbol{\theta}, V_{1,q}(\boldsymbol{\theta}))$, em que

$$V_{1,q}(\boldsymbol{\theta}) = J_{1,q}(\boldsymbol{\theta})^{-1} K_{1,q}(\boldsymbol{\theta}) \left[J_{1,q}(\boldsymbol{\theta})^{-1}\right]^{\top}, \tag{4.1.8}$$

em que

$$J_{1,q}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{\nabla_{\theta^{\top}} \left[\boldsymbol{U}^*(y_t, \boldsymbol{\theta}) f^*_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)^{1-q}\right]\right\},$$
$$K_{1,q}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{\boldsymbol{U}^*(y_t, \boldsymbol{\theta}) \boldsymbol{U}^*(y_t, \boldsymbol{\theta})^{\top} f^*_{\boldsymbol{\theta}}(y_t; \mu_t; \phi_t)^{2(1-q)}\right\}.$$

Ribeiro e Ferrari (2023) apresentaram as matrizes $J_{1,q}(\boldsymbol{\theta})$ e $K_{1,q}(\boldsymbol{\theta})$ para o modelo de regressão beta linear e demonstraram, ainda, que, para distribuições beta limitadas, o vetor escore modificado expresso em (4.1.7) é limitado para todo y no suporte da distribuição e que a sua derivada também é limitada. Isto implica que, sob a condição mencionada, o SMLE é B-robusto e que eventuais observações atípicas possuem pouca influência sobre o valor estimado de sua matriz de covariâncias assintótica. Neste trabalho não apresentaremos as matrizes $J_{1,q}(\boldsymbol{\theta})$ e $K_{1,q}(\boldsymbol{\theta})$ correspondentes aos modelos de regressão beta não lineares, uma vez que, conforme será abordado mais adiante, foi utilizado outro método para obtenção das estimativas dos erros padrão dos parâmetros.

4.2 Estimação via transformação da variável resposta

Modelos de regressão beta ajustados por meio do SMLE podem ser úteis para modelar dados provenientes de diversos fenômenos e situações. Entretanto, conforme pontuado da Seção 4.1, não existe garantia de que este estimador seja bem definido para distribuições beta que sejam não limitadas. Distribuições beta ilimitadas, que são aquelas cujas curvas para a densidade apresentam formas de J, J invertido ou U, são raras de serem observadas na prática, entretanto podem ocorrer para alguns conjuntos de dados limitados ao intervalo (0,1). Diante disso, Maluf, Ferrari e Queiroz (2025) introduziram novos estimadores que são bem definidos e preservam propriedades de robustez mesmo para distribuições beta cujas densidades não sejam limitadas, ou seja, que tendem ao infinito em um ou ambos os extremos do suporte da distribuição.

O processo de obtenção desses novos estimadores se baseou nos trabalhos de Ghosh (2019) e Ribeiro e Ferrari (2023), e consiste em replicar os métodos empregados nos trabalhos citados, porém, considerando a distribuição de uma transformação na variável resposta, aqui referida por y^*. Aqui nos limitaremos ao estimador decorrente do SMLE. Assim, seja $y \sim \mathcal{B}(\mu,\phi)$, considere y^* como a transformação logito de y, ou seja, $y^* =$

$\text{logito}(y) = \log[y/(1-y)]$. A PDF de y^* é dada por

$$\begin{aligned} s_\theta(y^*;\mu,\phi) &= \frac{1}{B(\mu\phi,(1-\mu)\phi)}\frac{e^{-y^*(1-\mu)\phi}}{(1+e^{-y^*})^\phi}, \\ &= \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\frac{\mathrm{e}^{-\mathrm{y}^*(1-\mu)\phi}}{(1+e^{-y^*})^\phi}, \quad y\in\mathbb{R}. \end{aligned}$$

A distribuição de y^* é chamada de beta exponencial generalizada do segundo tipo (*exponential generalized beta of the second type*; EGB) e escrevemos $y^* \sim \text{EGB}(\mu,\phi)$ (MCDONALD; XU, 1995). Ao contrário da distribuição beta convencional, a distribuição EGB é fechada sob a transformação potência para todo y^* no suporte da distribuição, pois

$$s_\theta(y^*,\mu,\phi)^\xi \propto s_\theta(y^*,\mu,\phi\xi),$$

para todo $y^*\in\mathbb{R}$, $\mu\in(0,1)$, $\phi>0$ e $\xi>0$.

Sejam $y_t^* = \log[y_t/(1-y_t)], t=1,\ldots,n$, onde cada y_t é uma observação independente obtida a partir do modelo de regressão beta com precisão variável definido em (2.3.1), e $s_{\boldsymbol{\theta}}(y^*,\mu_t,\phi_t)$ a PDF de y^*. A função L_q-verossimilhança referente à densidade $s_{\boldsymbol{\theta}}(\cdot,\mu_t,\phi_t)$, é dada por

$$\ell_q^\dagger(\boldsymbol{\theta}) = \sum_{t=1}^n L_q(s_{\boldsymbol{\theta}}(y_t^*;\mu_t;\phi_t)), \tag{4.2.1}$$

em que $\text{L}_q(\cdot)$ é a transformação definida em (4.1.1). Como já mencionado, a maximação da função em (4.2.1) conduz a estimadores que não são Fisher-consistentes. Assim, considerando que a família de distribuições EGB é fechada sob a transformação potência, estimadores Fisher-consistentes podem ser obtidos a partir da reparametrização $\tau_q^{-1}(\boldsymbol{\theta}) = \tau_{1/q}(\boldsymbol{\theta})$ de $\ell_q^\dagger(\boldsymbol{\theta})$. O LSMLE é obtido a partir da maximização da reparametrização da função L_q-verossimilhança definida em (4.2.1), dada por

$$\ell_q^{\dagger *}(\boldsymbol{\theta}) = \sum_{t=1}^n L_q(s_{\tau_{1/q}(\boldsymbol{\theta})}(y_t^*;\mu_t;\phi_t)) = \sum_{t=1}^n L_q(s_{\boldsymbol{\theta}}^{(1/q)}(y_t^*;\mu_t;\phi_t)).$$

em que $s_{\boldsymbol{\theta}}^{1/q}(y_t^*;\mu_t;\phi_t) = s_{\boldsymbol{\theta}}(y_t^*;\mu_t;\phi_{t,q^{-1}})$, para $q\in(0,1)$, e μ_t e ϕ_t satisfazendo (2.2.1) e (2.2.2), respectivamente. Para o caso não linear expresso em (2.3.1), os submodelos da média e da precisão associados a um modelo de regressão beta modificado, que aqui será denotado por $s_{\boldsymbol{\theta}}^*(y_t^*;\mu_t;\phi_t)$, são dados por

$$\begin{aligned} g_\mu^*(\mu_t) &= g_\mu(\mu_t) = f_\mu(\boldsymbol{X}_t;\boldsymbol{\beta}) = \eta_{\mu t}, \\ g_\phi^*(\phi_t) &= g_\phi(\phi_{t,q}) = f_\phi(\boldsymbol{Z}_t;\boldsymbol{\gamma}) = \eta_{\phi t}. \end{aligned} \tag{4.2.2}$$

e o LSMLE, que aqui será denotado por $\widetilde{\boldsymbol{\theta}}_q$, é obtido a partir da maximização da função

$$\ell_q^{\dagger *}(\boldsymbol{\theta}) = \sum_{t=1}^{n} L_q(s_{\boldsymbol{\theta}}^*(y_t^*; \mu_t; \phi_t)).$$

A equação de estimação associada ao processo de obtenção do LSMLE é

$$\sum_{t=1}^{n} \boldsymbol{U}^{\dagger *}(y_t^*, \boldsymbol{\theta}) s_{\boldsymbol{\theta}}^*(y_t^*; \mu_t; \phi_t)^{1-q} = \mathbf{0}, \tag{4.2.3}$$

em que $\boldsymbol{U}^{\dagger *}(y_t^*, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \log\left[s_{\boldsymbol{\theta}}^*(y_t^*; \mu_t; \phi_t)\right]$ é o vetor escore modificado para $\boldsymbol{\theta}$ referente à t-ésima observação, que para o caso não linear é dado por

$$\boldsymbol{U}^{\dagger *}(y_t^*, \boldsymbol{\theta}) = \left(\phi_t \frac{(y_t{}^* - \mu_t{}^*)}{g'_\mu(\mu_t)} \frac{\partial \eta_{\mu t}}{\partial \boldsymbol{\beta}^\top}, \quad q^{-1} \frac{\mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)}{g'_\phi(\phi_{t,q})} \frac{\partial \eta_{\phi t}}{\partial \boldsymbol{\gamma}^\top} \right)^\top,$$

Maluf, Ferrari e Queiroz (2025) demonstraram que o estimador $\widetilde{\boldsymbol{\theta}}_q$ também é Fisher-consistente para $\boldsymbol{\theta}$. Além disso, considerando que o LSMLE pertence à classe dos M-estimadores, temos que $\widetilde{\boldsymbol{\theta}}_q \overset{a}{\sim} \mathrm{N}(\boldsymbol{\theta}, V_{2,q}(\boldsymbol{\theta}))$, em que

$$V_{2,q}(\boldsymbol{\theta})) = J_{2,q}(\boldsymbol{\theta})^{-1} K_{2,q}(\boldsymbol{\theta}) \left[J_{2,q}(\boldsymbol{\theta})^{-1}\right]^\top, \tag{4.2.4}$$

sendo

$$J_{2,q}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{\nabla_{\theta^\top} \left[\boldsymbol{U}^{\dagger *}(y_t^*, \boldsymbol{\theta}) s_{\boldsymbol{\theta}}^*(y_t^*; \mu_t; \phi_t)^{1-q}\right]\right\} \text{ e}$$

$$K_{2,q}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathrm{E}\left\{\boldsymbol{U}^{\dagger *}(y_t^*, \boldsymbol{\theta}) \boldsymbol{U}^{\dagger *}(y_t^*, \boldsymbol{\theta})^\top s_{\boldsymbol{\theta}}^*(y_t^*; \mu_t; \phi_t)^{1-q}\right\}.$$

As expressões para as matrizes $J_{2,q}(\boldsymbol{\theta})$ e $K_{2,q}(\boldsymbol{\theta})$ para o modelo de regressão beta linear são apresentadas por Maluf, Ferrari e Queiroz (2025), que também ressaltam que $V_{2,q}(\boldsymbol{\theta})$ é bem definida para todo $\alpha = 1-q \in [0,1)$, e que para $\alpha = 0$ $(q = 1)$ a matriz de covariâncias assintóticas de $\widetilde{\boldsymbol{\theta}}_q$ equivale à matriz de covariãncias assintóticas do MLE. Além disso, tem-se que a IF do LSMLE é sempre limitada, o que implica que $\widetilde{\boldsymbol{\theta}}_q$ é B-robusto. Também não apresentaremos as matrizes $J_{2,q}(\boldsymbol{\theta})$ e $K_{2,q}(\boldsymbol{\theta})$ referentes aos modelos de regressão beta não lineares. Conforme será detalhado adiante, foi utilizado outro método para obtenção das estimativas dos erros padrão dos parâmetros do modelo.

4.3 Estimativas iniciais

Ressalta-se que não é possível explicitar ambos estimadores SMLE e LSMLE de forma analítica, sendo necessário recorrer a métodos de otimização numérica, a exemplo do BGFS (WRIGHT; NOCEDAL, 1999, p. 136, cap. 6), para maximizar as funções $\ell_q^*(\boldsymbol{\theta})$ e $\ell_q^{\dagger *}(\boldsymbol{\theta})$, respectivamente. Esses métodos exigem estimativas iniciais para que o processo iterativo seja realizado. Dessa forma, sugerimos, para os modelos não lineares aqui propostos, a obtenção desses valores iniciais a partir das estimativas de máxima verossimilhança do modelo de regressão beta não linear detalhado na Subseção 2.3, utilizando como f_μ e f_ϕ em (2.3.1) as mesmas funções de ligação a serem utilizadas nos submodelos da média e da precisão, respectivamente, dos ajustes sob os estimadores robustos.

Conforme será detalhado nas experimentações efetuadas nos estudos de simulação e aplicações do Capítulo 5, foram obtidos bons resultados na obtenção dos SMLEs e LSMLEs ao gerar as estimativas iniciais a partir do modelo de regressão beta não linear ajustado com o MLE.

4.4 Estimativas para os erros padrão via método *bootstrap*

Não obstante a possibilidade de obter as estimativas dos erros padrão de $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\gamma}}^\top)^\top$, nas estruturas de regressão em (2.3.1) por meio das matrizes de covariâncias assintóticas do SMLE e LSMLE em (4.1.8) e (4.2.4), respectivamente, para este trabalho optamos por calculá-las por meio de processo *bootstrap*.

O *bootstrap* consiste em uma abordagem computacional baseada em reamostragem, que permite estimar a distribuição de uma estatística de interesse e realizar inferências sobre ela, oferecendo uma alternativa prática e intuitiva para estimar erros padrão, construir intervalos de confiança, obter vieses de estimadores, simular distribuições amostrais de estatísticas, entre outras finalidades (EFRON; TIBSHIRANI, 1994; LIMA, 2017).

Desde sua proposição por Efron (1979), o método *bootstrap* consolidou-se como uma técnica estatística versátil e poderosa, amplamente utilizada em diversas áreas do conhecimento, se destacando por sua aplicabilidade em situações em que métodos estatísticos tradicionais não são totalmente viáveis ou são difíceis de implementar. Situações como essas podem ocorrer, por exemplo, em cenários com tamanhos amostrais pequenos ou quando não existem maiores informações sobre a real distribuição dos dados de interesse.

O procedimento utilizado para obtenção do erro padrão *bootstrap* é o descrito no Algoritmo 1 (EFRON; TIBSHIRANI, 1994), com o o erro padrão de $\hat{\boldsymbol{\theta}}$, denotado por

$\widehat{\mathbf{SE}_{\mathbf{boot}}}(\hat{\boldsymbol{\theta}})$, dado por

$$\widehat{\mathbf{SE}_{\mathbf{boot}}}(\hat{\boldsymbol{\theta}}) = \sqrt{\frac{1}{B_1 - 1}\sum_{b=1}^{B_1}\left(\hat{\boldsymbol{\theta}}^{(b)} - \overline{\hat{\boldsymbol{\theta}}}\right)^2},$$

em que B_1 é a quantidade de réplicas de $\hat{\boldsymbol{\theta}}$ geradas e $\overline{\hat{\boldsymbol{\theta}}} = \left(\sum_{b=1}^{B_1}\hat{\boldsymbol{\theta}}^{(b)}\right)/B_1$. Observe que aqui não está sendo feita nenhuma suposição sobre a distribuição de $\hat{\boldsymbol{\theta}}$. Com base em estudos e simulações efetuadas, Efron e Tibshirani (1994) mostram que para a maioria dos casos uma quantidade $B_1 = 200$ réplicas é suficiente para se alcançar bons resultados.

Algoritmo 1 Cálculo do erro padrão dos estmadores via *bootstrap*

Entrada: Vetor $\boldsymbol{y}$ contendo as n realizações da variável resposta, quantidade B_1 de reamostragens de $\boldsymbol{y}$, vetores de estimativas $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\gamma}}$ e matrizes de covariáveis X e Z utilizadas no ajuste do modelo.

para $b = 1$ até B_1 **faça**
gere uma réplica $\boldsymbol{y}^{(k)}$ da resposta a partir do modelo postulado utilizando as estimativas originais $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$ para os parâmetros $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, respectivamente.
Estime $\tilde{\boldsymbol{\theta}}^{(k)} = (\tilde{\boldsymbol{\beta}}^{(k)}, \tilde{\boldsymbol{\gamma}}^{(k)})$ com base em $\boldsymbol{y}^{(k)}$.
Armazene $\tilde{\boldsymbol{\theta}}^{(k)}$.
fim para
Calcule o erro padrão amostral de $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ com base nas B_1 estimativas $\tilde{\boldsymbol{\theta}}^{(k)}$ por meio da formula

$$\widehat{\mathbf{SE}_{\mathbf{boot}}}(\hat{\boldsymbol{\theta}}) = \sqrt{\frac{1}{B_1 - 1}\sum_{k=1}^{B_1}\left(\tilde{\boldsymbol{\theta}}^{(k)} - \overline{\tilde{\boldsymbol{\theta}}}\right)^2}.$$

Saída: Vetor $\widehat{\mathbf{SE}_{\mathbf{boot}}}(\hat{\boldsymbol{\theta}})$ contendo os erros padrão *bootstrap* dos componentes de $\boldsymbol{\theta}$.

4.5 Teste de hipóteses robusto

Além das estimações pontuais discutidas nas Seções anteriores, também é importante avaliar os coeficientes de regressão dos modelos para, considerando a amostra em estudo, avalair se as covariáveis são relevantes para explicar o comportamento da variável resposta. Para essa finalidade, usualmente são utilizados testes de hipóteses baseados na função de verossimilhança e que dependem do processo de estimação por máxima verossimilhança. Entretanto, é esperado que, sob contaminação nos dados, testes de hipóteses baseados no MLE também sejam sensíveis observações atípicas e, portanto, tenham o seu desempenho prejudicado.

Seguindo as ideias do trabalho de Heritier e Ronchetti (1994), Ribeiro e Ferrari (2023) propuseram um teste de hipóteses robusto baseado na estatística de Wald para o modelo de regressão beta em estudo. A nova estatística de teste, referida por estatística tipo-Wald, é obtida a partir da mesma fórmula da estatística de Wald, porém com a substituição do MLE e do correspondente erro padrão (*standard error*; SE) pelo SMLE e por seu SE assintótico obtido a partir da matriz de covariâncias em (4.1.8).

Sejam $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ o vetor de parâmetros, $\widehat{\boldsymbol{\theta}}_q = (\widehat{\theta}_{1q}, \ldots, \widehat{\theta}_{pq})^\top$ o vetor das respectivas estimativas, e $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)} \cdots, \theta_p^{(0)})^\top$ um vetor de valores dados, segundo Ribeiro e Ferrari (2023), o teste referente a um único parâmetro considera as hipóteses $H_0 : \theta_k = \theta_k^{(0)}$ contra $H_1 : \theta_k \neq \theta_k^{(0)}$, $1 \leq k \leq p$, e a correspondente estatística do teste tipo-Wald é definida por

$$W_{0,q} = \frac{(\widehat{\theta}_{kq} - \theta_k^{(0)})^2}{\mathrm{SE}\left(\widehat{\theta}_{kq}\right)^2}, \tag{4.5.1}$$

em que $\widehat{\theta}_{kq}$ é a estimativa de θ_k e $\mathrm{SE}\left(\widehat{\theta}_{kq}\right)$ é o erro padrão assintótico obtido a partir da matriz de covariâncias assintóticas expressa em (4.1.8). Maluf, Ferrari e Queiroz (2025) utilizam a mesma estatística para testar coeficientes da regressão sob o LSMLE, porém substituindo o MLE e o correspondente SE pelo LSMLE e seu SE assintótico obtido a partir da matriz de covariâncias em (4.2.4). Sob H_0 e condições usuais de regularidade, mostra-se que a estatística do teste em (4.5.1) possui distribuição aproximada qui-quadrado com 1 grau de liberdade ($W_{0,q} \overset{a}{\sim} \chi^2_1$). Nesse sentido, considerando um nível de significância α, rejeitamos H_0 em favor de H_1 quando a estatística $W_{0,q}$ for maior do que o quantil de ordem $(1-\alpha)$ da distribuição χ^2_1, ou seja, se $W_{0,q} \geq \chi^2_{1,1-\alpha}$, com $\mathrm{P}(\chi^2_1 \leq \chi^2_{1,1-\alpha}) = 1-\alpha$.

Entretanto, a utilização do erro padrão *bootstrap* pode tornar inadequada a utilização da distribuição assintótica da estatística original do teste tipo-Wald, expressa em (4.5.1). Dessa forma, o p-valor do teste de hipóteses robusto tratado na Subseção 4.5 foi obtido por meio da distribuição empírica da citada estatística, gerada também via *bootstrap* a partir do processo detalhado no Algoritmo 2 a seguir. A partir desse processo foi possível obter estimativas para os p-valores referentes ao teste tipo-Wald. Observe que, para esse caso, também não está sendo feita qualquer suposição em relação a distribuição de estatística $W_{0,q}$. Com base em Efron e Tibshirani (1994) e Efron (1992), foram utilizados os valores $B_1 = 200$ e $B_2 = 500$ para as contantes necessárias para inicialização do Algoritmo 2.

Algoritmo 2 Cálculo do p-valor via *Bootstrap* do teste de hipóteses tipo-Wald

Entrada: Número de réplicas B_2 da estatística tipo-Wald, número B_1 de reamostragens para obtenção do erro padrão, vetores de estimativas $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\gamma}}$ e seus respectivos erros padrão $\widehat{\mathbf{SE}_{\mathbf{boot}}(\hat{\boldsymbol{\beta}})}$ e $\widehat{\mathbf{SE}_{\mathbf{boot}}(\hat{\boldsymbol{\gamma}})}$, vetor $\boldsymbol{y}$ de realizações da variável resposta e matrizes de covariáveis X e Z.

Calcule as estatísticas tipo-Wald observadas:

$$\boldsymbol{W}_{\boldsymbol{\beta}}^{\text{obs}} = \left(\frac{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}}{\widehat{\text{SE}_{\text{boot}}(\hat{\boldsymbol{\beta}})}}\right)^2, \quad \boldsymbol{W}_{\boldsymbol{\gamma}}^{\text{obs}} = \left(\frac{\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^{(0)}}{\widehat{\text{SE}_{\text{boot}}(\hat{\boldsymbol{\gamma}})}}\right)^2.$$

para $j = 1$ até B_2 **faça**

Gere uma réplica $\boldsymbol{y}^{(j)}$ da variável resposta a partir do modelo postulado, porém sob a hipótese nula, ou seja, utilizando $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)^\top}, \boldsymbol{\gamma}^{(0)^\top})^\top$, $\boldsymbol{\beta}^{(0)^\top} = (0, \ldots, 0)^\top \in \mathbb{R}^{p_1}$ e $\boldsymbol{\gamma}^{(0)^\top} = (0, \ldots, 0)^\top \in \mathbb{R}^{p_2}$.

Estime $\hat{\boldsymbol{\theta}}^{(j)} = (\hat{\boldsymbol{\beta}}^{(j)}, \hat{\boldsymbol{\gamma}}^{(j)})$ com base em $\boldsymbol{y}^{(j)}$.

Armazene $\hat{\boldsymbol{\theta}}^{(j)}$.

para $k = 1$ até B_1 **faça**

Gere nova réplica $\boldsymbol{y}^{(j,k)}$ da resposta a partir do modelo postulado, porém utilizando os vetores $\hat{\boldsymbol{\beta}}^{(j)}$, $\hat{\boldsymbol{\gamma}}^{(j)}$ para os parâmetros $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, respectivamente.

Estime $\hat{\boldsymbol{\theta}}^{(j,k)} = (\hat{\boldsymbol{\beta}}^{(j,k)}, \hat{\boldsymbol{\gamma}}^{(j,k)})$ com base em $\boldsymbol{y}^{(j,k)}$.

Armazene $\hat{\boldsymbol{\theta}}^{(j,k)}$.

fim para

Obtenha o erro padrão amostral de $\hat{\boldsymbol{\theta}}^{(j)} = (\hat{\boldsymbol{\beta}}^{(j)}, \hat{\boldsymbol{\gamma}}^{(j)})$ com base nas B_1 estimativas $\hat{\boldsymbol{\theta}}^{(j,k)}$ por meio da formula

$$\widehat{\mathbf{SE}_{\mathbf{boot}}(\hat{\boldsymbol{\theta}})}^{(j)} = \sqrt{\frac{1}{B_1 - 1}\sum_{k=1}^{B_1}\left(\hat{\boldsymbol{\theta}}^{(j,k)} - \overline{\hat{\boldsymbol{\theta}}^{(j)}}\right)^2}.$$

Calcule as réplicas das estatísticas tipo-Wald observadas:

$$\boldsymbol{W}_{\boldsymbol{\beta}}^{(j)} = \left(\frac{\widehat{\boldsymbol{\beta}}^{(j)}}{\widehat{\mathbf{SE}_{\mathbf{boot}}(\hat{\boldsymbol{\beta}}^{(j)})}}\right)^2, \quad \boldsymbol{W}_{\boldsymbol{\gamma}}^{(j)} = \left(\frac{\widehat{\boldsymbol{\gamma}}^{(j)}}{\widehat{\mathbf{SE}_{\mathbf{boot}}(\hat{\boldsymbol{\gamma}}^{(j)})}}\right)^2.$$

fim para

Para cada $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, calcular o *p-valor* bootstrap:

$$\boldsymbol{p}_\beta = \frac{1}{B_2}\sum_{j=1}^{B_2}\mathcal{I}\left(\boldsymbol{W}_{\boldsymbol{\beta}} > \boldsymbol{W}_{\boldsymbol{\beta}}^{\text{obs}}\right), \quad \boldsymbol{p}_\gamma = \frac{1}{B_2}\sum_{j=1}^{B_2}\mathcal{I}\left(\boldsymbol{W}_{\boldsymbol{\gamma}} > \boldsymbol{W}_{\boldsymbol{\gamma}}^{\text{obs}}\right),$$

em que $\mathcal{I}$ é a função indicadora, e $\boldsymbol{W}_{\boldsymbol{\beta}}$ e $\boldsymbol{W}_{\boldsymbol{\gamma}}$ são os vetores contendo as B_2 réplicas das estatísticas tipo-Wald correspondentes a $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ respectivamente.

Saída: Vetores $\boldsymbol{p}_{\boldsymbol{\beta}} \in \mathbb{R}^{p_1}$ e $\boldsymbol{p}_{\boldsymbol{\gamma}} \in \mathbb{R}^{p_2}$ com os p-valores *bootstrap* referentes aos componentes dos parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, respectivamente.

4.6 Constante de afinação

Vimos nas seções anteriores que as equações de estimação associadas ao SMLE e LSMLE, dadas por (4.1.6) e (4.2.3), respectivamente, dependem diretamente da definição de um valor para a constante de afinação, denotada nessas equações pela letra q, tal que $q \in (0,1]$. Tal constante é de suma importância para o processo de estimação, pois seu valor controla o balanceamento entre eficiência assintótica e robustez dos estimadores. Desse modo, valores menores para q privilegiam a robustez do estimador em detrimento da eficiência.

A escolha de um valor ótimo para q constitui um problema adicional no processo de estimação, considerando a sua importância e o fato de que este deve ser fixado a priori. LA VECCHIA, Camponovo e Ferrari (2015) sugerem utilizar um valor para q que seja mais próximo de 1, ou seja, de modo que o estimador obtido fique próximo ao MLE convencional para que as estimativas dos parâmetros sejam suficientemente estáveis na presença de contaminação e proporcionem eficiência completa na ausência de contaminação nos dados. Nos trabalhos de Ghosh e Basu (2016) e Ghosh (2019), por exemplo, a escolha de um valor para q é sugerida com base em estudos de simulação e comparações dos valores das estimativas dos parâmetros de regressão para diferentes valores da constante, todos próximos de 1.

Ribeiro e Ferrari (2023) propuseram um método orientado a dados baseado na proposta de LA VECCHIA, Camponovo e Ferrari (2015), porém utilizando uma padronização das estimativas, objetivando remover o efeito do tamanho amostral e da magnitude das estimativas de parâmetros distintos. Esse método tem se mostrado bastante eficaz e foi utilizado com sucesso para seleção da constante de afinação em outros trabalhos que envolvem regressão robusta, tais como Queiroz (2022) e Maluf, Ferrari e Queiroz (2025).

Neste trabalho, propomos uma adaptação ao método de seleção introduzido por Ribeiro e Ferrari (2023), objetivando dar estabilidade ao processo de seleção da constante de afinação. No algoritmo original, a medida utilizada para definição do critério de parada do procedimento, o vetor de variações quadráticas padronizadas (*standardized quadratric variations*; SQV), é padronizada pela variabilidade das estimativas dos parâmetros, sendo utilizado o erro padrão assintótico para tanto. Conforme detalhado na Subseção 4.4, para este trabalho o erro padrão será obtido por meio do processo de *bootstrap* detalhado no Algoritmo 1, e a sua utilização no método de seleção original poderia resultar na obtenção de diferentes valores de q para um mesmo conjunto de dados caso não seja utilizada uma semente para a reprodutibilidade do processo computacional, o que não é desejável. Além disso, a mudança tornou o processo computacional mais rápido e eficiente, reduzindo consideravelmente a quantidade de cálculos necessários e, consequentemente, o tempo total de execução.

A adaptação proposta consiste em substituir a medida SQV pela norma euclidiana dos vetores das estimativas dos parâmetros (*euclidian norm of the parameter estimate vectors*; ENPEV), dada por

$$\text{ENPEV}_{q_k} = \left\| \widehat{\theta}_{q_k} - \widehat{\theta}_{q_{k+1}} \right\|, \tag{4.6.1}$$

em que $\widehat{\theta}_{q_k} = (\widehat{\theta}^1_{q_k}, \ldots, \widehat{\theta}^p_{q_k})$, e q_k é o k-ésimo valor testado para a constante de afinação q. A k-ésima estimativa do vetor de parâmetros será considerada suficientemente estável se o valor da ENPEV for menor do que o produto entre uma constante B pré-fixada e a mediana dos valores da medida ENPEV obtidos no passo inicial do algoritmo. Observe que, dessa forma, sempre obteremos um critério de parada relativo à magnitude das estimativas dos parâmetros, tornando desnecessário utilizar o erro padrão para padronizar essas estimativas. Portanto, seja $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$, $p = p_1 + p_2$, o vetor dos parâmetros a serem estimados e q_k, $k = 1, \cdots, m$, os m valores a serem atribuídos para a constante de afinação q, então o método proposto segue os passos descritos no Algoritmo 3, a seguir.

Observa-se que o Algoritmo 3, recomendado para seleção da constante de afinação para ambos os estimadores robustos tratados neste trabalho, escolhe um valor ótimo para q que seja o mais próximo possível de 1 ou, caso a estabilidade não seja alcançada, escolhe $q = 1$ e o estimador resultante será o MLE. Quanto às contantes necessárias como entrada para o Algoritmo 3, foram utilizadas as sugestões de Ribeiro e Ferrari (2023) para considerar o tamanho das grades em $m = 3$ (exceto na primeira execução do algoritmo), valor mínimo de q em $q_{\min} = 0{,}5$ e o espaçamento da grade em 0,02. Observe que essa configuração, em especial o valor sugerido para o q_{min}, garante a escolha de um valor ótimo para q que seja mais próximo de 1 do que de 0, o que privilegia a estabilidade e a AE do estimador. Em relação à constante B a ser utilizada para cálculo do valor limitante da condição de estabilidade, a partir de experimentos com amostras simuladas sugerimos utilizar o valor $B = 2{,}1$. Observe que a adaptação aqui proposta manteve uma das principais características do método, que é a seleção da constante q com base nos próprios dados a serem utilizados no ajuste do modelo de regressão.

4.7 Implementação computacional

Todos os cálculos e avaliações numéricas relacionadas às estimações dos parâmetros dos modelos, bem como os gráficos gerados ao longo desse trabalho, foram realizados com suporte computacional utilizando a linguagem de programação R e o *software* estatístico de mesmo nome, em sua versão 4.4.0. O software R (R Core Team, 2024) é de domínio público e está disponível gratuitamente para *download* no endereço eletrônico http://www.r-project.org/.

Algoritmo 3 Seleção do valor ótimo para a constante de afinação q

Entrada: Conjunto de dados para o qual se pretende ajustar o modelo, multiplicador $B = 2{,}1$ para condição de estabilidade, tamanho $m = 3$ das grades a partir da segunda grade, valor mínimo $q_{min} = 0{,}5$ da constante de afinação e espaçamento $s = 0{,}02$ dos valores da grade

Defina uma grade ordenada de forma decrescente de valores para q que sejam igualmente espaçados com distância s entre si, ou seja, $q_0 > q_1 > q_2 > \cdots > q_{m_1}$, tal que $q_0 = 1$, e $q_{m_1} = 0{,}8$
repita
  para cada q_k na grade **faça**
    Calcule as estimativas dos parâmetros obtendo $\widehat{\boldsymbol{\theta}}_{q_k} = (\widehat{\boldsymbol{\theta}}^1_{q_k}, \cdots, \widehat{\boldsymbol{\theta}}^p_{q_k})$
    Calcule o vetor das medidas ENPEV, conforme definido em (4.6.1)
  fim para
  se grade inicial **então**
    Obtenha a mediana dos valores do vetor de medidas ENPEV obtido no passo anterior, aqui denotada por med(ENPEV)
  fim se
  se todas as $\text{ENPEV}_{q_k} < B * \text{med(ENPEV)}$ **então**
    Defina $q^* = \max_{q_m < q_k < q_0} (q_k)$
    Pare
  senão
    Identifique o menor q_k tal que $\text{ENPEV}_{q_k} \geq B * \text{med(ENPEV)}$
    Defina q_{start} como o próximo ponto na grade após o q_k definido no passo anterior
    Construa nova grade decrescente com m novos valores para q com espaçamento s entre si a partir de q_{start}
  fim se
até a estabilidade ser alcançada, ou seja, todas as $\text{ENPEV}_{q_k} < B * \text{med(ENPEV)}$ ou $q_{\text{start}} = q_{\text{min}}$
se $q_{\text{start}} = q_{\text{min}}$, significa que a estabilidade não foi alcançada **então**
  Defina $q^* = 1$
fim se

Saída: Valor ótimo da constante de afinação q: q^*

Estudos de simulação com diferentes cenários, e aplicações a dados simulados e reais foram realizadas para ilustrar a aplicabilidade da metodologia que está sendo proposta neste trabalho. Para tanto, os processos de obtenção dos estimadores SMLE e LSMLE foram implementados computacionalmente por meio de adaptações efetuadas nas funções e métodos disponibilizados na biblioteca `robustbetareg` (QUEIROZ; MALUF, 2022) do *software* R. A biblioteca `robustbetareg` permite obter diretamente os referidos estimadores para a regressão beta linear, sendo que as adaptações visaram adequar os cálculos para contemplar as estruturas de regressão com as formas não lineares utilizadas neste trabalho para o modelo de regressão beta.

Além disso, para viabilizar a comparação dos referidos estimadores com métodos

não robustos, foi efetuada a implementação computacional do processo de estimação via o MLE descrito na Subseção 2.3, contemplando as mesmas estruturas não lineares de regressão utilizadas para o SMLE e LSMLE sob o modelo beta não linear.

Conforme já mencionado nas Subseções 4.4 e 4.5, as estimativas dos erros padrão dos coeficientes e a distribuição da estatística do teste tipo-Wald, respectivamente, foram obtidas por meio de processos *bootstrap*. Assim, foi efetuada a implementação computacional dos procedimentos detalhados nos Algoritmos 1 e 2.

Para a realização dos estudos de simulação que serão apresentados na Subseção 5.1 e para a aplicação com dados simulados em 5.2.1 foram utilizados processos *bootstrap* paramétricos para geração das réplicas das amostras utilizadas para ajuste dos modelos de regressão beta não lineares robustos avaliados. Para esses casos as réplicas de Monte Carlo foram geradas partindo do pressuposto de que a variável resposta possui distribuição beta na forma espressa em (2.3.1).

Os códigos em R, o conjunto de dados utilizado e os resultados obtidos nesse trabalho estão disponíveis em repositório `Github`[1]. Por meio do material disponibilizado no repositório é possível reproduzir os estudos de simulação e as aplicações aqui efetuadas, bem como utilizar as implementações efetuadas para outros estudos e análises de dados.

[1]Disponível em: ⟨https://github.com/eddusousa/nlrobustbetareg⟩.

5 Resultados e discussões

5.1 Estudos de simulação

Para avaliar o desempenho e comparar os modelos de regressão beta não lineares sob os estimadores MLE, SMLE e LSMLE, foram realizados estudos de simulação de Monte Carlo baseados em 1000 réplicas e considerando amostras com e sem contaminação nos dados. Os tamanhos amostrais considerados são $n = 40$, 80, 160, e 320. Os valores das covariáveis foram obtidos para o tamanho amostral $n = 40$ e replicados duas, quatro e oito vezes para obter as matrizes de covariáveis correspondentes aos tamanhos amostrais $n = 80$, 160 e 320, respectivamente. Segundo Espinheira, Santos e Cribari-Neto (2017), esse método garante que o grau de heteroscedasticidade seja constante para todos os tamanhos amostrais. Em todos os cenários foram consideradas como ligação as funções logito nos submodelos da média e logarítmica para os submodelos da precisão sob as estrutura de regressão apresentadas em (2.3.1). Todos os submodelos contêm interceptos e as covariáveis são obtidas a partir de variáveis aleatórias com distribuição uniforme padrão e mantidas constantes ao longo das amostras simuladas. Para o cenário com precisão variável, a mesma covariável é utilizadas no submodelo da média e precisão. A porcentagem de contaminação na amostra para todos os cenários foi fixada em 5%. A seleção da constante de afinação q para os modelos sob os estimadores SMLE e LSMLE foi efetuada utilizando o Algoritmo 3 para seleção. Além disso, em todos os cenários foi aplicada a função exponencial nos termos correspondentes à variável explicativa dos submodelos da média para obtenção da estrutura não linear de regressão, resultando na forma

$$g_\mu(\mu_t) = \beta_1 + \mathrm{e}^{\beta_2 \mathrm{x_{t1}}}.$$

No cenário de precisão variável foi considerada uma estrutura de regressão linear para o submodelo da precisão. Além disso, distintas configurações dos valores dos parâmetros e diferentes padrões de contaminação nos dados foram considerados para cada um dos quatro cenários.

Para cada cenário simulado, o experimento consistiu nos seguintes passos:

Passo 1. Foram geradas amostras considerando o modelo de regressão beta não linear especificado para o cenário.

Passo 2. Os dados foram contaminados conforme padrão descrito em cada cenário, de modo que para cada uma das réplicas de Monte Carlo foi obtida uma versão sem contaminação e outra contaminada.

Passo 3. Foram ajustados modelos de regressão beta não lineares sob os três estimadores considerados para a amostra contaminada e não contaminada de cada uma das réplicas.

Passo 4. Por meio de análises numéricas e gráficas dos erros e dos valores das estimativas, efetuou-se uma comparação dos resultados dos modelos descritos no Passo 3.

Os cenários considerados estão descritos a seguir.

Cenário 1: Modelo de regressão beta não linear com precisão variável e valores da média da variável resposta próximos a 0,85. Os valores dos parâmetros foram fixados em $\beta_1 = -1{,}7$, $\beta_2 = 1{,}2$, $\gamma_1 = 2{,}5$ e $\gamma_2 = 3{,}5$, de modo que, para as amostras geradas, as médias de μ e ϕ ficaram próximas a 0,80 e 110, respectivamente. A amostra contaminada substitui as observações geradas com as 5% maiores médias da resposta por observações geradas a partir de um modelo de regressão beta com média $\mu_t' = (2 - 1{,}5\mu_t)/2$ e precisão ϕ_t. Por exemplo, se $\mu_t \approx 0{,}85$, então $\mu_t' \approx 0{,}36$.

Cenário 2: Modelo de regressão beta não linear com precisão constante e valores da média da variável resposta por volta de 0,75. Os valores dos parâmetros foram fixados em $\beta_1 = -1{,}0$, $\beta_2 = 1{,}0$ e $\gamma_1 = 6{,}5$, de modo que para as amostras geradas as médias de μ ficaram próximas a 0,80 e ϕ ficou igual a 665. A amostra contaminada substitui 5% das observações, sendo 2,5% das observações geradas com os maiores valores de μ e 2,5% daquelas com os menores valores de μ. A contaminação é gerada por meio de um modelo de regressão beta com precisão ϕ e média $\mu_t^{(1)} = a_1 c_t/(1 + a_1 c_t)$ e $\mu_t^{(2)} = a_2 c_t/(1 + a_2 c_t)$, respectivamente, em que $c_t = \mu_t/(1 - \mu_t)$, $a_1 = 0{,}2$ e $a_2 = 2{,}4$. Assim, se se $\mu_t \approx 0{,}75$, então $\mu_t^{(1)} \approx 0{,}38$ e $\mu_t^{(2)} \approx 0{,}88$.

Cenário 3: Modelo de regressão beta não linear com precisão constante e valores da média da variável resposta próximos a 0,4. Os valores dos parâmetros foram fixados em $\beta_1 = -1{,}0$, $\beta_2 = -1{,}4$ e $\gamma_1 = 6{,}0$, de modo que para as amostras geradas as médias de μ ficaram próximas a 0,39 e ϕ ficou igual a 403. A amostra contaminada substitui as observações geradas com as 5% menores médias da resposta por observações geradas a partir de um modelo de regressão beta com média $\mu_t' = (2 - 1{,}7\mu_t)/2$ e precisão ϕ. Nesse caso, se $\mu_t \approx 0{,}4$, então $\mu_t' \approx 0{,}66$.

Cenário 4: Modelo de regressão beta não linear com precisão constante e valores da média da variável resposta próximos a 0.8. Os valores dos parâmetros foram fixados em $\beta_1 = -1{,}7$, $\beta_2 = 1{,}2$ e $\gamma_1 = 4{,}7$, de modo que para as amostras geradas as médias de μ ficaram próximas a 0,8 e ϕ ficou igual a 110. A amostra contaminada substitui as observações geradas com as 5% maiores médias da resposta por observações geradas a partir de um modelo de regressão beta com média $\mu_t' = (2 - 2\mu_t)/2$ e precisão ϕ. Portanto, se $\mu_t \approx 0{,}8$, então $\mu_t' \approx 0{,}2$.

A Figura 3 ilustra as diferentes formas de contaminação descritas acima para os 4 cenários, considerando uma única amostra de tamanho 80, em que as observações contaminadas estão destacadas em vermelho.

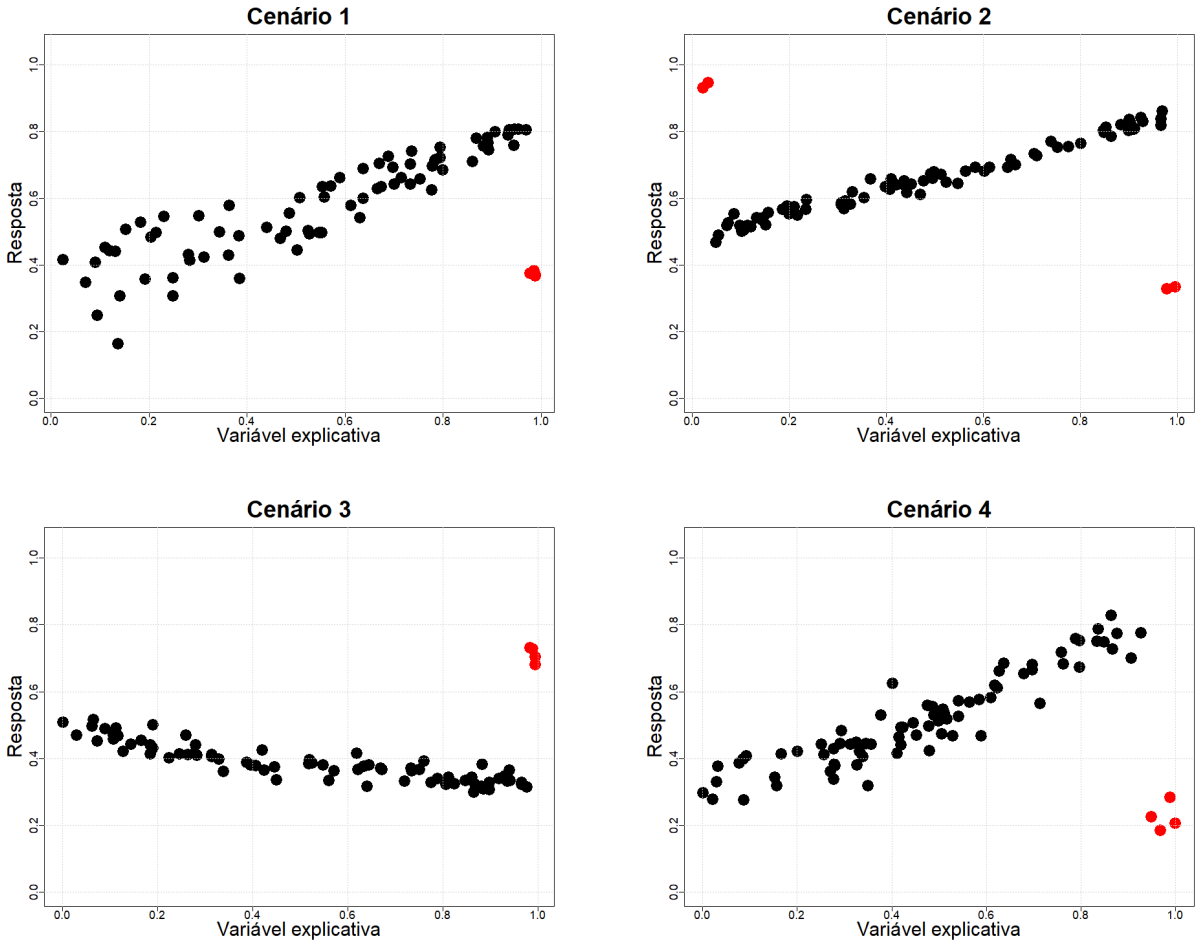

Figura 3: Ilustração dos quatro padrões de contaminação utilizados para uma amostra de tamanho $n = 80$. Os pontos em vermelho correspondem às observações contaminadas introduzidas na amostra.

Na Tabela 1 são feitas comparações da eficiência dos estimadores por meio das razões entre os erros quadráticos médios totais (*total mean squared errors*; TMSE) do MLE, SMLE e LSMLE sob os cenários considerados. Observa-se que para os dados não contaminados, a eficiência dos três estimadores é igual a 1 na maioria dos casos e próximas desse valor nos demais, indicando similaridade nos ajustes, conforme esperado. Isto indica que a escolha ótima da constante q na ausência de contaminação está funcionando. Já nos casos onde há contaminação, constata-se que os estimadores robustos são muito mais eficientes do que o MLE, uma vez que seus TMSEs são consideravelmente menores. Em em todos os cenários essa discrepância tende a aumentar conforme se aumenta o tamanho amostral. Por exemplo, considerando o Cenário 2, o TMSE é cerca de de 79 vezes maior do que os TMSEs do SMLE e do LSMLE para as amostras com 40 observações. Nesse mesmo cenário, essa razão aumenta para aproximadamente 162, 326 e 612 para os tamanhos

amostrais 80, 160 e 320, respectivamente. Além disso, observa-se que apenas sob o Cenário 1 tivemos diferença de desempenho entre o SMLE e LSMLE, ainda que pequena.

Tabela 1: Razão entre os TMSEs das estimativas sob os Cenários 1, 2, 3 e 4.

	Cenário 1						**Cenário 2**					
	Sem contaminação			**Com contaminação**			**Sem contaminação**			**Com contaminação**		
n	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$
40	0,92	0,91	0,99	9,20	8,17	0,89	0,98	0,98	1,00	78,83	78,77	1,00
80	0,99	0,99	1,00	23,76	22,73	0,96	0,99	0,99	0,99	161,58	161,78	1,00
160	1,00	1,00	1,00	56,80	56,01	0,99	0,98	1,00	1,00	325,70	325,25	1,00
320	1,00	1,00	1,00	117,78	116,37	0,99	1,00	1,00	1,00	612,75	612,39	1,00
	Cenário 3						**Cenário 4**					
	Sem contaminação			**Com contaminação**			**Sem contaminação**			**Com contaminação**		
n	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$	$\frac{\text{MLE}}{\text{SMLE}}$	$\frac{\text{MLE}}{\text{LSMLE}}$	$\frac{\text{SMLE}}{\text{LSMLE}}$
40	0,99	0,99	1,00	65,92	65,93	1,00	0,97	0,97	1,00	54,72	54,75	1,00
80	1,00	1,00	1,00	134,08	134,01	1,00	1,00	1,00	1,00	111,73	112,17	1,00
160	1,00	1,00	1,00	295,81	295,99	1,00	1,00	1,00	1,00	246,64	247,63	1,00
320	1,00	1,00	1,00	592,24	592,91	1,00	1,00	1,00	1,00	512,43	513,67	1,00

Nas Figuras 4, 5, 6 e 7 são apresentados os *boxplots* das estimativas dos parâmetros utilizando o MLE, SMLE e LSMLE sob os Cenários 1, 2, 3 e 4, respectivamente, para os dados na presença e ausência de contaminação. Na ausência de contaminação os estimadores possuem desempenho praticamente idêntico, e isto se deve ao método de escolha da constante selecionar $q = 1$ na grande maioria das vezes. Para todos os cenários observa-se que as estimativas dos parâmetros sob o MLE foram fortemente influenciadas pela contaminação introduzida nas amostras, gerando estimativas bastante divergentes dos verdadeiros valores desses parâmetros. Entretanto, nota-se que os estimadores robustos tem desempenhos nas amostras contaminadas bastante próximos aos do MLE nas amostras sem contaminação, indicando que os processos robustos de estimação funcionaram bem em todos os cenários simulados. Inclusive, percebe-se que as medianas das estimativas robustas nas amostras contaminadas ficaram todas centradas em torno dos verdadeiros valores dos parâmetros, porém, essa robustez é alcançada às custas de uma maior variabilidade das estimativas, situação que se acentua para os menores tamanhos amostrais. Além disso, verifica-se que, de uma forma geral, o SMLE e o LSMLE apresentam desempenhos muito próximos em todas as situações, conforme já era esperado.

Na Figura 8 são utilizados *boxplots* para ilustrar a distribuição dos valores escolhidos, via algoritmo de seleção, para a constante de afinação q para o SMLE e LSMLE em todos os cenários. É possível verificar que o processo de seleção das constantes de afinação funcionou de forma adequada, uma vez que os valores ótimos de q ficaram iguais a 1 para a grande maioria das amostras sem contaminação, resultando no próprio MLE e garantindo eficiência assintótica total. Em contrapartida, nas amostras contaminadas foram selecionados, quase na totalidade das réplicas, valores diferentes de 1 para a constante q, resultando em estimadores robustos. Nesses casos os valores de q ficaram centrados em

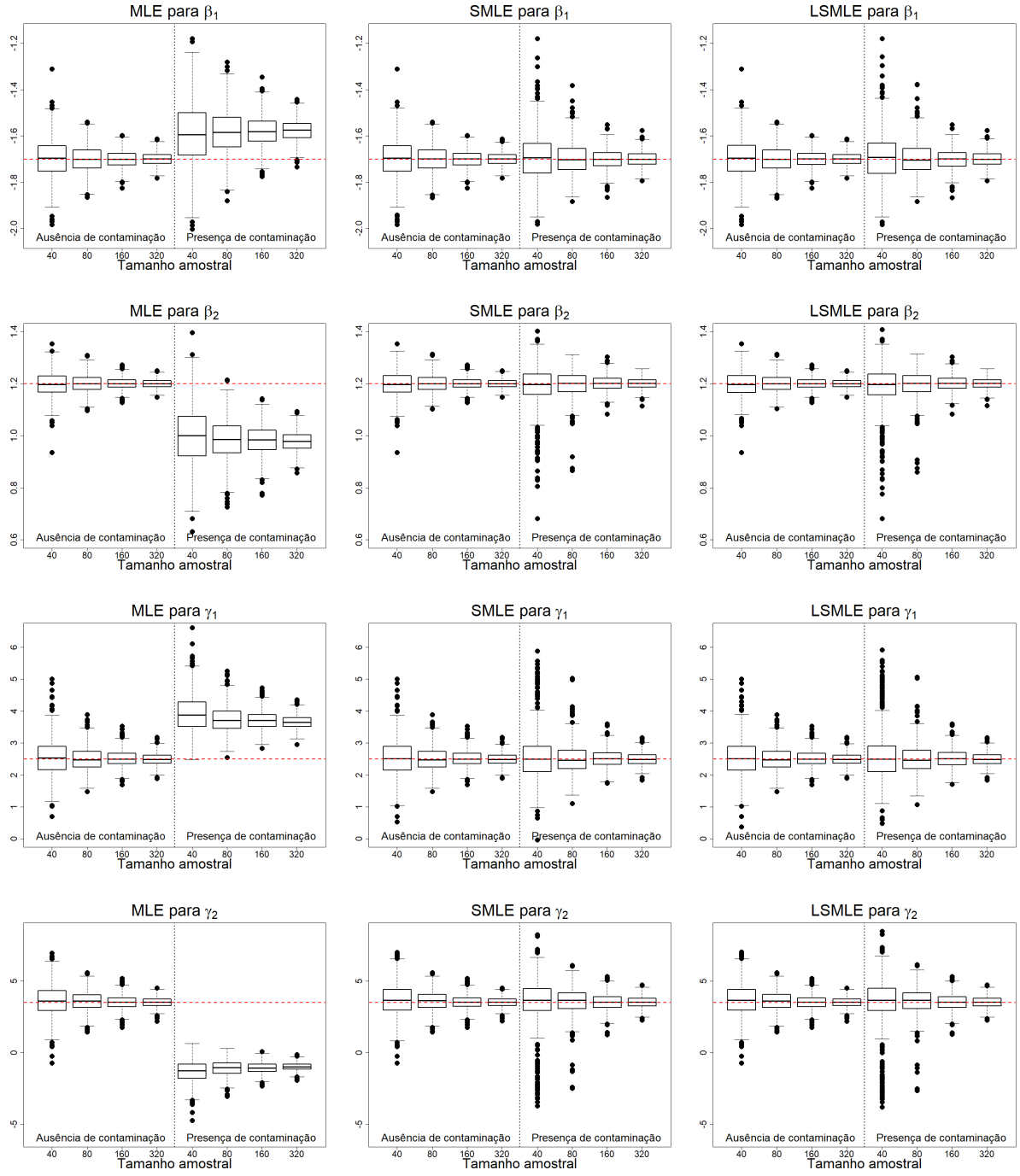

Figura 4: *Boxplots* das estimativas dos parâmetros β_1, β_2, γ_1 e γ_2 sob o Cenário 1: MLE (esquerda), SMLE (centro) e LSMLE (direita). A linha tracejada em vermelho representa o verdadeiro valor do parâmetro.

0,84, 0,94 nos Cenários 1 e 3, respectivamente, e em 0,92 nos Cenários 2 e 4.

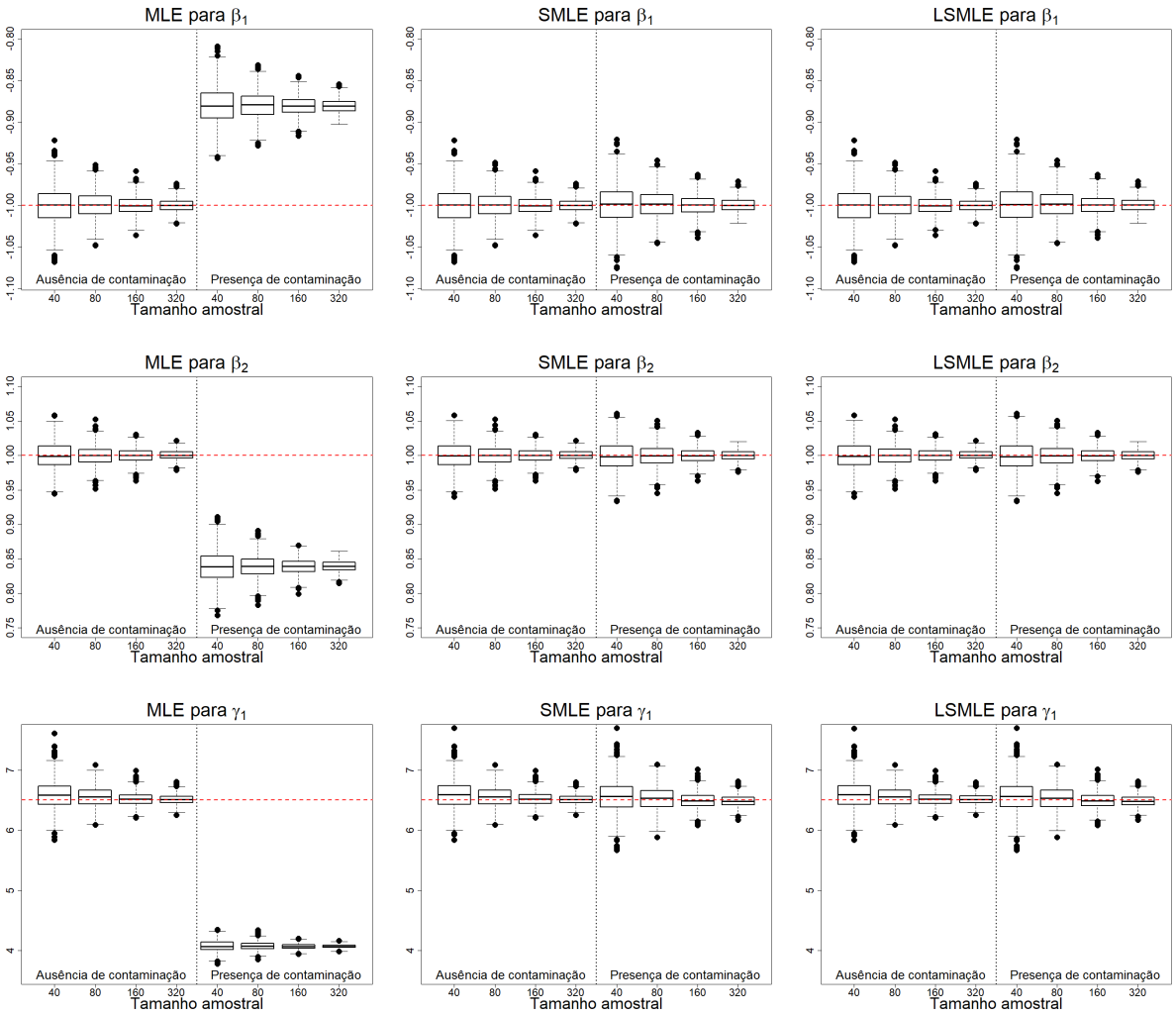

Figura 5: *Boxplots* das estimativas dos parâmetros β_1, β_2 e γ_1 sob o Cenário 2: MLE (esquerda), SMLE (centro) e LSMLE (direita). A linha tracejada em vermelho representa o verdadeiro valor do parâmetro.

5.2 Aplicações

Para complementar a avaliação efetuada nos estudos de simulação e melhor ilustrar a aplicabilidade dos modelos de regressão beta não lineares robustos propostos neste trabalho, foram realizadas, ainda, duas aplicações, sendo uma com dados simulados e outra com dados reais. Para todos os modelos ajustados sob os estimadores SMLE e LSMLE, a seleção da constante de afinação q foi efetuada utilizando o Algoritmo 3, e foram consideradas como ligação as funções logito nos submodelos da média e logarítmica para os submodelos precisão nas estruturas em (2.3.1). Além disso, para fins de diagnóstico dos modelos foram utilizados, também em todos os casos, os resíduos quantis aleatorizados introduzidos por Dunn e Smyth (1996). Para os testes de hipóteses aplicados, a menos que expressamente indicado o contrário, está sendo considerado um nível de 5% de significância.

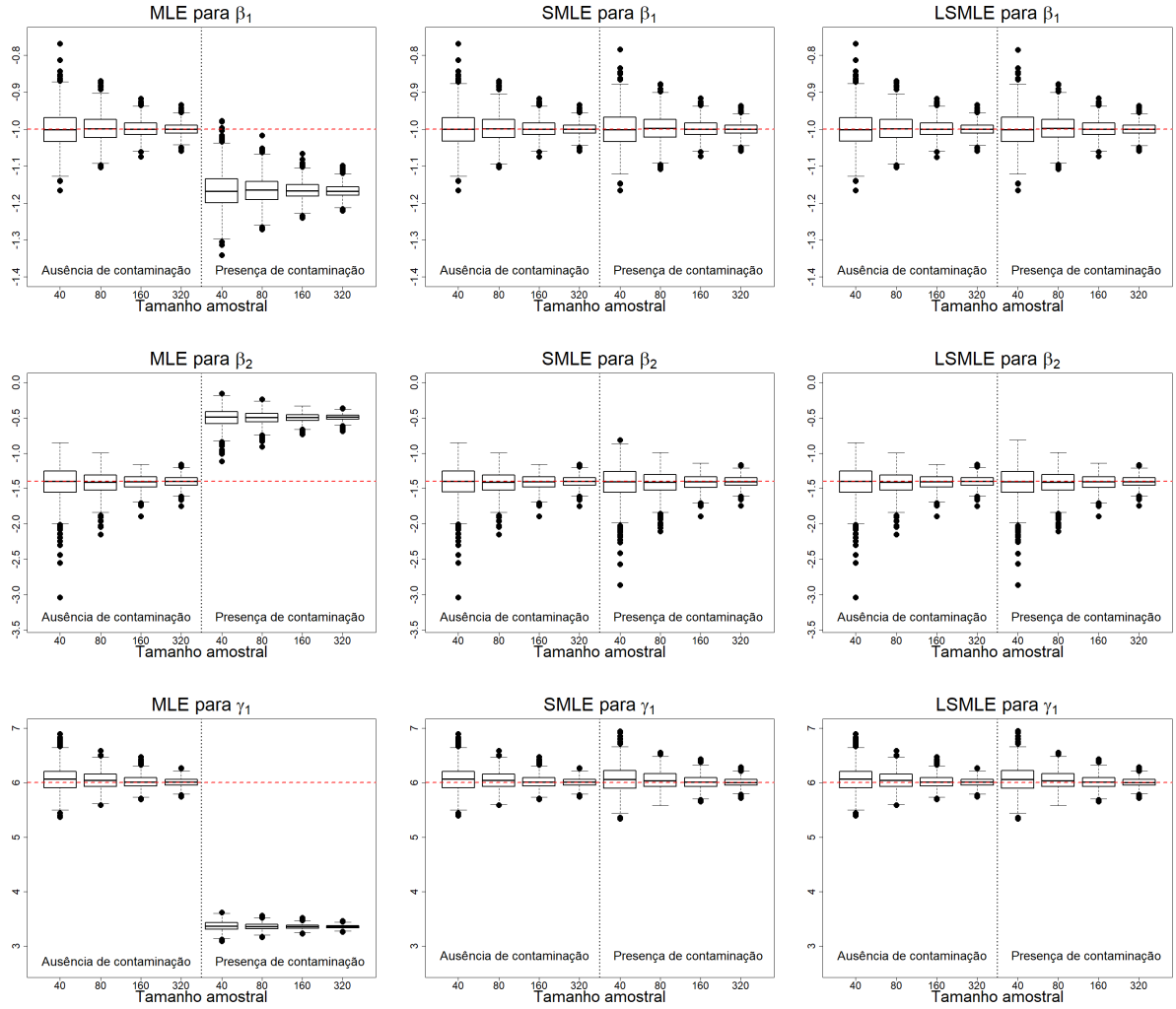

Figura 6: *Boxplots* das estimativas dos parâmetros β_1, β_2 e γ_1 sob o Cenário 3: MLE (esquerda), SMLE (centro) e LSMLE (direita). A linha tracejada em vermelho representa o verdadeiro valor do parâmetro.

5.2.1 Aplicação com dados simulados

Para esta aplicação foram geradas 4 amostras de tamanhos $n = 40$, 80, 160, e 320, considerando o modelo de regressão beta não linear com precisão constante

$$\begin{aligned} g_\mu(\mu_t) &= \beta_1 + x_{t1}^{\beta_2}, \\ g_\phi(\phi) &= \gamma_1, \end{aligned} \tag{5.2.1}$$

em que x_{t1} é valor da covariável associada ao submodelo da média para a t-ésima observação. Os valores dos parâmetros foram fixados em $\beta_1 = -0{,}6$, $\beta_2 = 0{,}8$ e $\gamma_1 = 3{,}9$, de modo que para as amostras geradas as médias de μ ficaram próximas a 0,50 e ϕ é igual a 49,4. A forma não linear para o submodelo da média em (5.2.1) foi utilizada anteriormente por Espinheira, Santos e Cribari-Neto (2017).

Os dados foram gerados de forma que para cada uma das amostras foi obtida uma

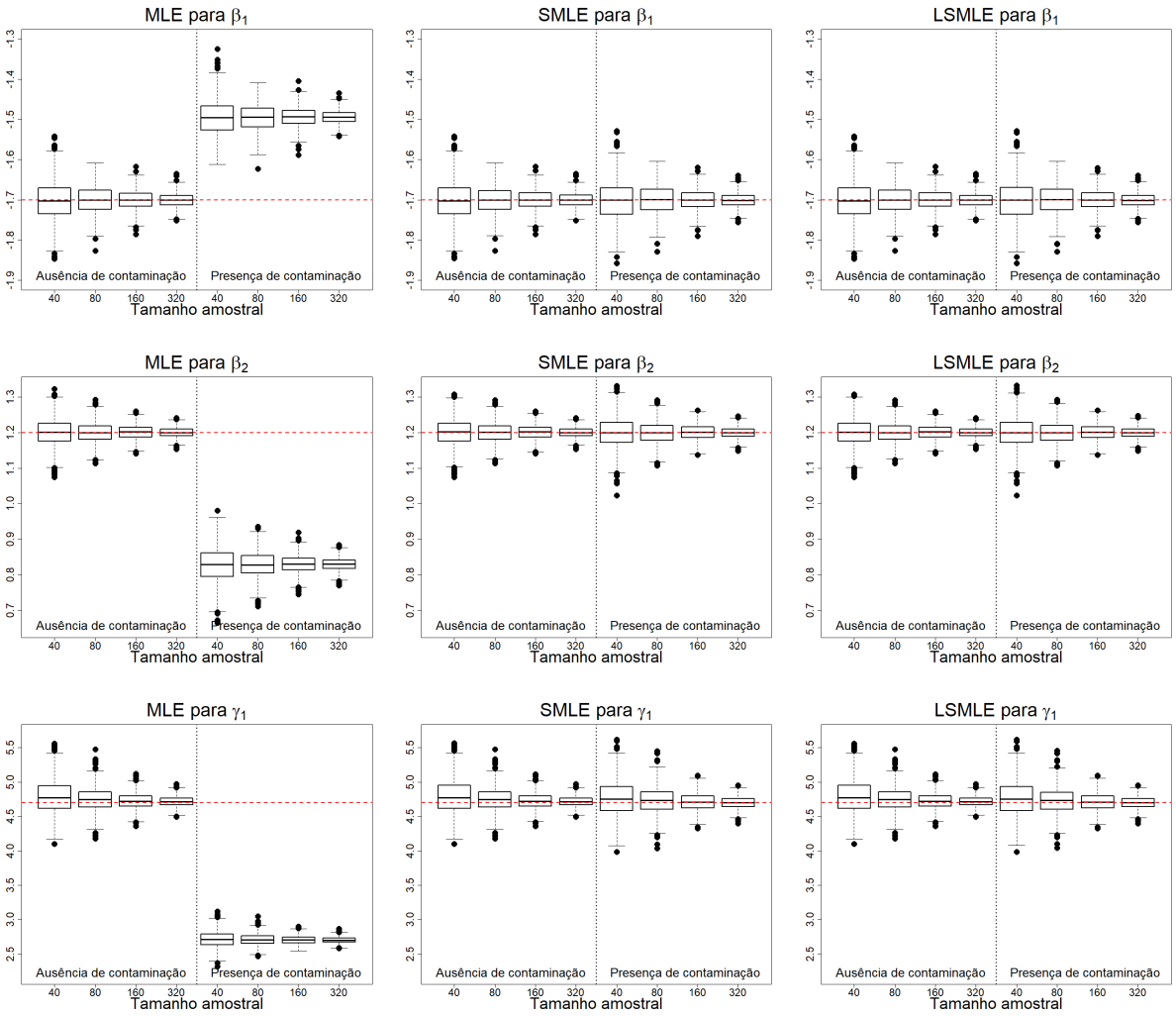

Figura 7: *Boxplots* das estimativas dos parâmetros β_1, β_2 e γ_1 sob o Cenário 4: MLE (esquerda), SMLE (centro) e LSMLE (direita). A linha tracejada em vermelho representa o verdadeiro valor do parâmetro.

versão sem contaminação e outra contaminada. As amostras contaminadas substituem 5% das observações, sendo 2,5% das observações geradas com os maiores valores de μ e 2,5% daquelas com os menores valores de μ. A contaminação é gerada por meio de um modelo de regressão beta com precisão ϕ e média $\mu_t^{(1)} = a_1 c_t/(1 + a_1 c_t)$ e $\mu_t^{(2)} = a_2 c_t/(1 + a_2 c_t)$, respectivamente, em que $c_t = \mu_t/(1 - \mu_t)$, $a_1 = 0{,}1$ e $a_2 = 15$. Assim, se se $\mu_t \approx 0{,}50$, então $\mu_t^{(1)} \approx 0{,}09$ e $\mu_t^{(2)} \approx 0{,}94$. Observe que esta configuração é similar à do Cenário 2 da simulação, onde a contaminação ocorre em ambos os extremos do intervalo considerado para a variável resposta.

A exemplo do procedimento adotado nos estudo de simulação, para garantir estabilidade no grau de heteroscedasticidade, os valores das covariáveis foram obtidos para o tamanho amostral $n = 40$ e replicados duas, quatro e oito vezes para obter as matrizes de covariáveis correspondentes aos demais tamanhos amostrais. A Figura 9 ilustra as amostras utilizadas para os 4 tamanhos amostrais destacando as observações contaminadas, que estão em vermelho.

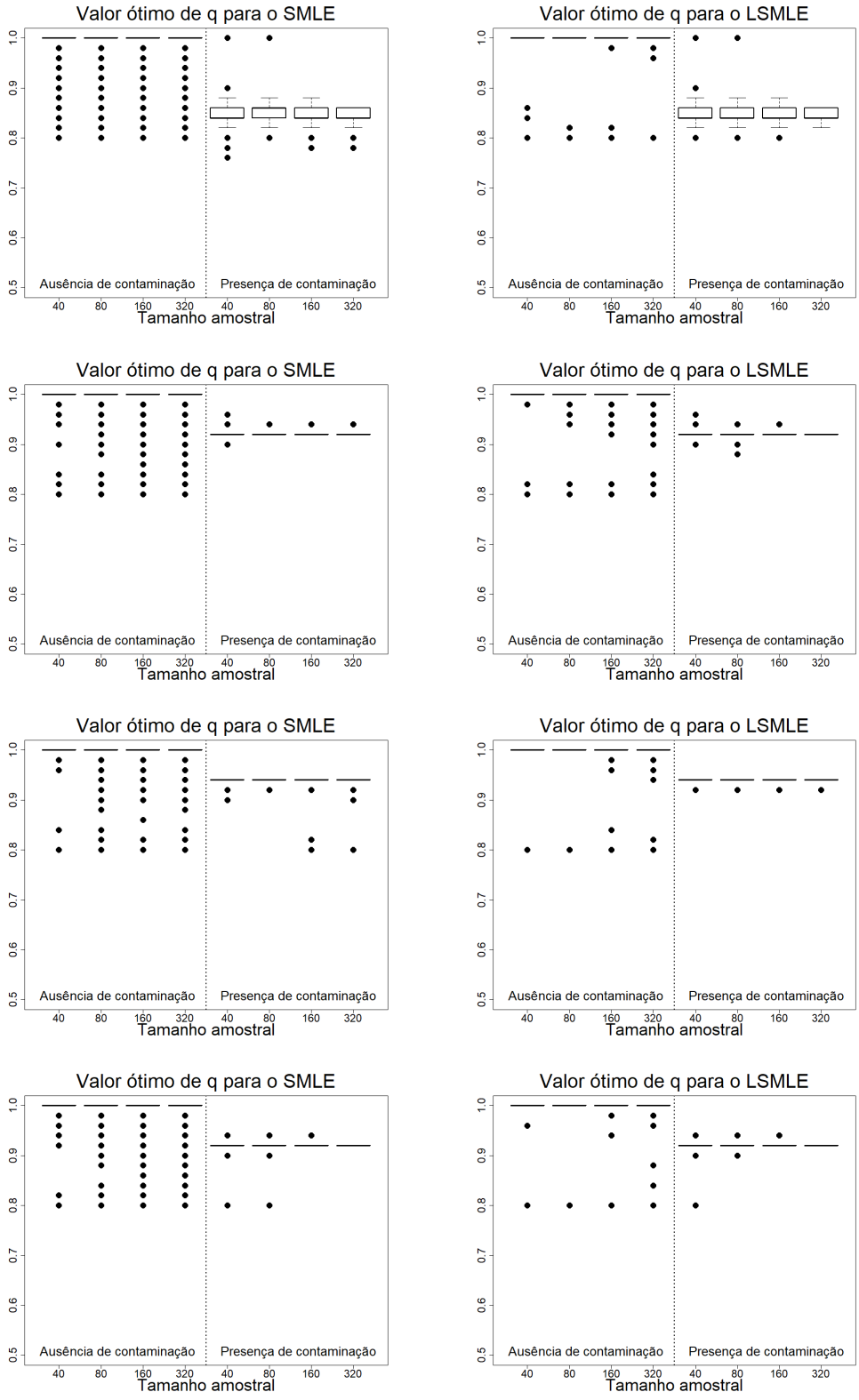

Figura 8: *Boxplots* dos valores ótimos para a constante de afinação q para o SMLE (esquerda), e LSMLE (direita), sob os cenários 1 (primeira linha), 2 (segunda linha), 3 (terceira linha) e 4 (quarta linha)).

Foram ajustados modelos de regressão beta não lineares com precisão constante sob os três estimadores considerados para os dados com e sem contaminação em todos

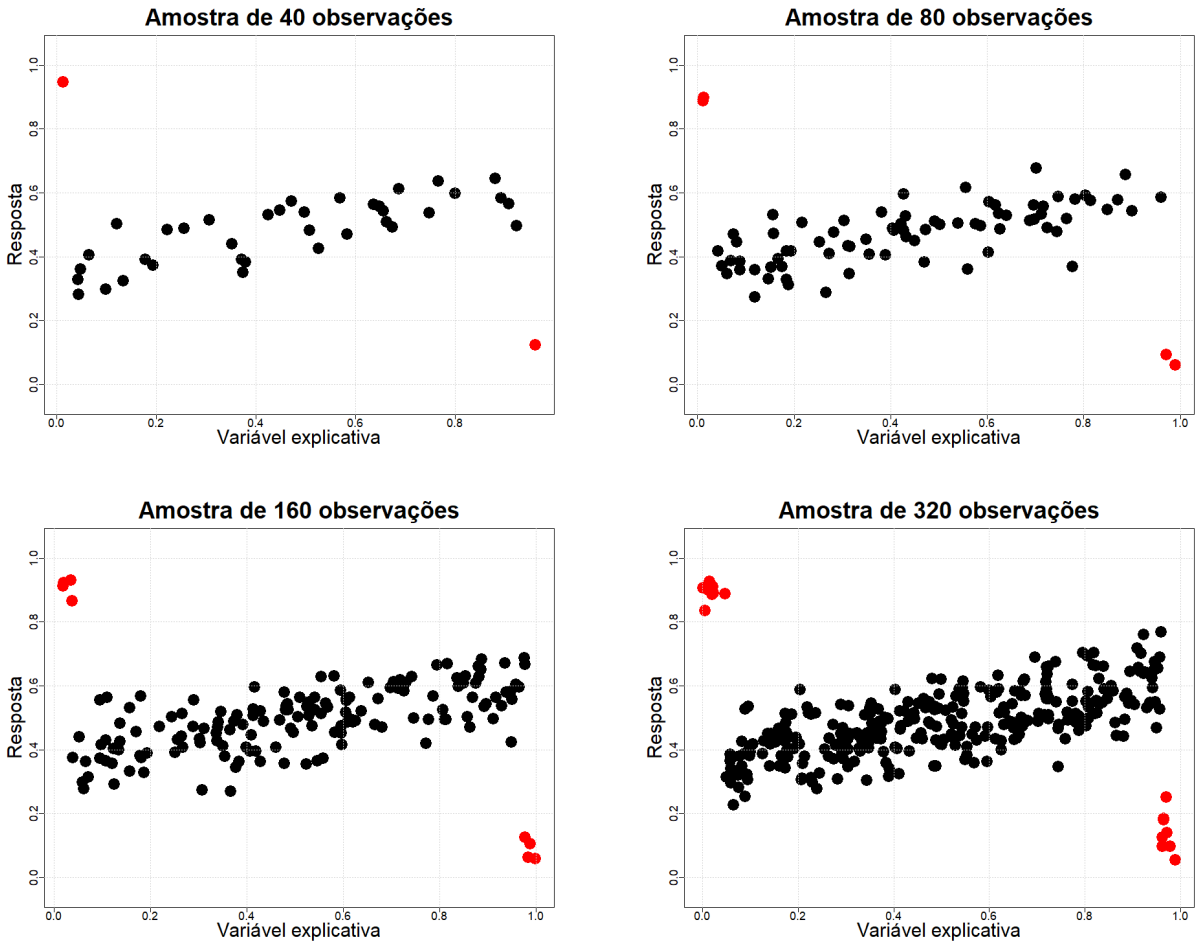

Figura 9: Gráfico de dispersão das amostras geradas para a aplicação. Os pontos em vermelho correspondem às observações contaminadas introduzidas na amostra.

os tamanhos amostrais. A Figura 10 apresenta o diagrama de dispersão da variável explicativa usada no submodelo da média versus a variável resposta contaminada para as 4 amostras geradas, juntamente com as respectivas curvas ajustadas sob o MLE para os dados contaminados e não contaminados e sob o SMLE e LSMLE para os dados contaminados. A partir destes gráfiicos é perceptível o quanto as observações atípicas introduzidas conduzem a mudanças significativas nas curvas de regressão ajustadas, ocasionando, nesse caso, uma inversão de sentido na relação entre as variáveis. No entanto, os ajustes sob o SMLE e LSMLE produzem curvas de regressão bem ajustadas e praticamente indistinguíveis entre si e também quando comparadas à curva ajustada sob o MLE para as amostras não contaminadas. Além disso, todos os ajustes sob os estimadores robustos nas amostras sem contaminação conduziram ao valor $q = 1$ para a constante de afinação, resultando no MLE.

A Tabela 2 elenca os valores das estimativas dos parâmetros, erros padrão, estatísticas tipo-Wald e respectivos p-valores obtidos via *bootstrap* para os modelos ajustados sob o MLE. Observa-se que, a exemplo do que já havia sido identificado na análise gráfica

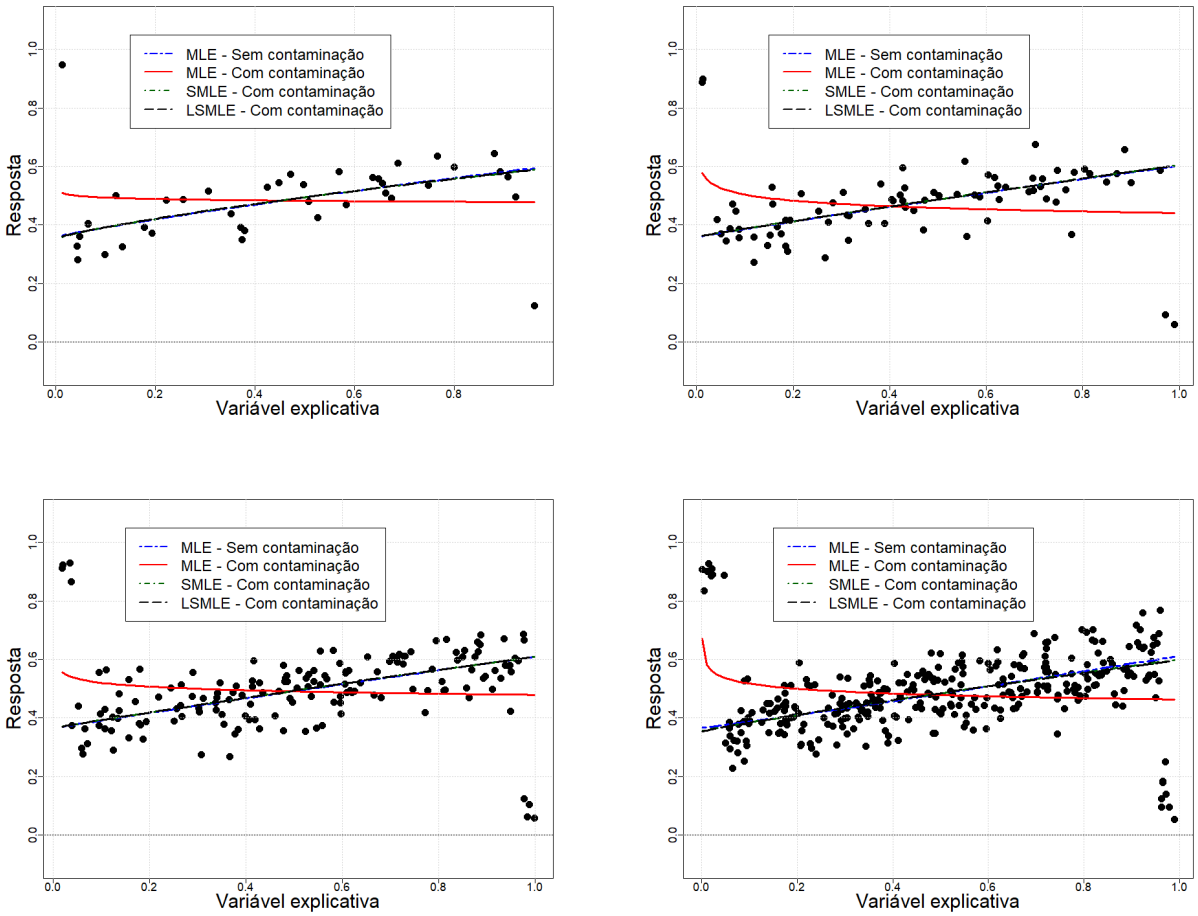

Figura 10: Gráfico de dispersão das amostras contaminadas geradas para a aplicação e as curvas ajustadas para cada cenário considerado.

anterior, as estimativas sob o MLE ficaram desproporcionalmente influenciadas pela contaminação introduzida nos dados, resultando em estimativas relativamente distantes dos valores reais dos parâmetros em todos os casos. Nota-se, inclusive, estimativas com sinal invertido para o coeficiente associado à covariável do submodelo da média. Além disso, os p-valores referentes aos testes tipo-Wald efetuados para avaliar a significância do parâmetro associado à covariável ficaram mais altos nos tamanhos amostrais de 40, 80 e 160 dos dados contaminados. No caso das amostras de 40 e 160 observações, caso seja considerado um nível de significância de 5%, por exemplo, a conclusão do teste seria de não significância desse coeficiente de regressão. Diante desses achados e considerando os valores reais dos parâmetros e as estimativas para os dados não contaminados, percebe-se uma inadequação do ajuste sob o MLE quando existe contaminação nas amostras.

Na Tabela 3 são relacionados valores obtidos para as estimativas dos parâmetros, erros padrão, estatísticas tipo-Wald e respectivos p-valores obtidos via *bootstrap* para os modelos ajustados sob o SMLE e LSMLE nas amostras contaminadas. Comparando os valores das estimativas dos coeficientes de regressão com os obtidos para o MLE nas

Tabela 2: Estimativas, erros-padrão *bootstrap*, estatísticas w e valores-p *bootstrap* para regressão beta não linear com precisão constante ajustada com o MLE nas amostras com e sem contaminação.

Amostra de 40 obs.	Sem contaminação				Com contaminação			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,587	0,110	28,490	-	−1,087	0,251	18,726	-
Covariável	0,831	0,395	4,437	0,036	−0,027	25,831	0,000	0,860
submodelo da precisão								
Intercepto	4,157	0,267	242,439	-	2,353	0,218	117,012	-
Amostra de 80 obs.	Sem contaminação				Com contaminação			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,585	0,087	45,033	-	−1,236	0,092	179,084	-
Covariável	0,911	0,310	8,635	$< 0{,}002$	−0,098	0,051	3,716	0,012
submodelo da precisão								
Intercepto	4,166	0,158	695,766	-	2,520	0,158	256,956	-
Amostra de 160 obs.	Sem contaminação				Com contaminação			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,558	0,054	106,142	-	−1,083	0,067	260,321	-
Covariável	0,929	0,190	23,910	$< 0{,}002$	−0,067	0,045	2,202	0,068
submodelo da precisão								
Intercepto	3,929	0,114	1189,219	-	2,414	0,104	537,649	-
Amostra de 320 obs.	Sem contaminação				Com contaminação			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,546	0,048	129,317	-	−1,147	0,037	944,503	-
Covariável	1,065	0,194	30,221	$< 0{,}002$	−0,086	0,023	14,155	$< 0{,}002$
submodelo da precisão								
Intercepto	3,839	0,075	2641,484	-	2,532	0,078	1058,767	-

amostras não contaminadas constantes da Tabela 2, verifica-se uma boa adequação do SMLE e LSMLE aos dados após contaminação. Inclusive os valores das estimativas ficaram próximos dos valores reais dos parâmetros usados na geração das amostras, mesmo para os menores tamanhos amostrais, e apresentam tendência de se aproximar ainda mais desses valores quanto maior for a amostra utilizada. Os valores obtidos para a constante de afinação q nas amostras contaminadas sob o SMLE foram 0,92 para o tamanho amostral 40, 0,90 para os tamanhos 80 e 160, e 0,88 para a amostra de 320 observações. Sob o LSMLE os valores selecionados para q foram 0,92 para as amostras de 40 e 80 observações e 0,90 para as amostras de tamanho 160 e 320. No caso das amostras sem contaminação, os ajustes dos modelos sob os métodos de estimação robustos retornaram $q = 1$ em todos os casos, evidenciando o funcionamento satisfatório do método de seleção da referida constante. Quanto aos p-valores, apesar de ficarem um pouco mais altos na amostra de 40 observações para ambos estimadores, ainda sim indicam significância do parâmetro relacionado à covariável no submodelo da média em todos os tamanhos amostrais quando consideramos um nível de 5% para o teste tipo-Wald. Então, ao contrário da conclusão obtida sob os modelos ajustados com o MLE, os ajustes sob o SMLE e LSMLE se mostraram adequados tanto para os cenários onde existe contaminação na amostra quanto em situações onde não há contaminação.

Um método gráfico muito útil para avaliação da qualidade do ajuste em relação à distribuição de probabilidade assumida para a variável resposta é o gráfico normal de probabilidades com envelope simulado, introduzido por Atkinson (1985). Este método

Tabela 3: Estimativas, erros-padrão *bootstrap*, estatísticas w e valores-p *bootstrap* para regressão beta robusta com precisão constante ajustada com SMLE e LSMLE nas amostras com contaminação.

Amostra de 40 obs.	SMLE				LSMLE			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,612	0,109	31,276	-	−0,610	0,111	30,010	-
Covariável	0,760	0,352	4,665	0,006	0,764	0,356	4,612	0,018
submodelo da precisão								
Intercepto	4,023	0,232	300,350	-	4,036	0,226	318,189	-
Amostra de 80 obs.	SMLE				LSMLE			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,578	0,090	41,415	-	−0,579	0,086	44,995	-
Covariável	0,921	0,314	8,597	0,002	0,918	0,316	8,446	0,010
submodelo da precisão								
Intercepto	4,144	0,159	682,140	-	4,070	0,166	602,532	-
Amostra de 160 obs.	SMLE				LSMLE			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,554	0,058	91,043	-	−0,553	0,062	78,762	-
Covariável	0,923	0,196	22,082	$< 0{,}002$	0,927	0,238	15,210	$< 0{,}002$
submodelo da precisão								
Intercepto	3,798	0,113	1126,014	-	3,809	0,105	1307,397	-
Amostra de 320 obs.	SMLE				LSMLE			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−0,593	0,048	155,100	-	−0,603	0,048	157,092	-
Covariável	0,914	0,156	34,344	$< 0{,}002$	0,882	0,165	28,623	$< 0{,}002$
submodelo da precisão								
Intercepto	3,712	0,076	2385,736	-	3,616	0,077	2181,941	-

consiste na inclusão, em um gráfico normal de probabilidades, de bandas obtidas por meio de amostras geradas pelo método de Monte Carlo a partir do modelo ajustado. Assim, com o auxílio de tais bandas pode-se identificar possíveis afastamentos entre valores realizados da variável resposta e a distribuição de probabilidades teórica assumida. Para modelos de regressão beta, têm sido comum a utilização de gráficos de probabilidade normal dos resíduos com envelope simulado para avaliação da qualidade dos ajustes (ESPINHEIRA; FERRARI; CRIBARI-NETO, 2008; OSPINA; FERRARI, 2012; ESPINHEIRA; SANTOS; CRIBARI-NETO, 2017), além de demonstrar o funcionamento dos estimadores robustos em relação às observações atípicas (RIBEIRO; FERRARI, 2023; MALUF; FERRARI; QUEIROZ, 2025).

As Figuras 11 e 12 apresentam os gráficos de probabilidade normal dos resíduos quantílicos dos modelos com envelope simulado considerando um nível de 90% de confiança, para os tamanhos amostrais 40, 80, 160 e 320. Esses gráficos revelam que, enquanto o ajuste sob o MLE nos dados sem contaminação apresentaram resultado satisfatório, ocorreu uma certa falta de ajuste do MLE nas amostras contaminadas, evidenciada pela quantidade considerável de pontos localizados ligeiramente fora das bandas dos envelopes em todos os tamanhos amostrais. Quanto a isso, observa-se que mesmo os resíduos referentes às observações atípicas decorrentes da contaminação introduzida não apresentam grande distância em relação às bandas do envelope, indicando que esses pontos realmente tiveram peso no processo de estimação dos parâmetros. Em outras palavras, pelo fato das observações atípicas terem peso igual no processo de estimação, então o modelo tentou se

ajustar também a estas observações, produzindo resíduos não tão aberrantes, apesar de influentes. Diferentemente disso, os ajustes efetuados sob os estimadores robustos apresentaram resíduos dentro dos limites da banda em sua maioria e aqueles correspondentes aos *outliers* bem distantes, indicando que tiveram pouco peso no processo. Ressalta-se que, conforme pondera Ribeiro (2020), esse é o comportamento esperado para estimadores robustos, ou seja, o método tem por objetivo ajustar bem a maioria dos dados, mas não necessariamente as observações atípicas.

Na Figura 13 são apresentados os gráficos correspondentes a todas os tamannhos amostrais das ponderações estimadas versus os resíduos dos ajustes sob o MLE para os dados com e sem contaminação e sob o SMLE e LSMLE somente para as amostras contaminadas. Nas imagens observa-se que, para o MLE os pesos são constantes e iguais a 1 para todas as observações, além disto o ajuste na amostra contaminada produz resíduos ligeiramente discrepante dos demais para as observações atípicas. Para os estimadores robustos, verifica-se que o peso atribuído varia conforme a observação, sendo mais próximos de zero para os *outliers* e próximos a 1 para as demais observações. Portanto, considerando esse comportamento, que se repetiu em todos tamanhos amostrais considerados, constata-se que os ajustes produziram resultados adequados e dentro do esperado.

5.2.2 Aplicação com dados reais

Para esta aplicação, estão sendo utilizados dados disponibilizados por Monllor-Hurtado, Pennino e Sanchez-Lizaso (2017), que foram obtidos a partir de um estudo que teve por objetivo avaliar o impacto do aquecimento dos oceanos na pesca global. A partir da verificação de que nos últimos anos houve um aumento nas capturas de espécies de peixes de águas mais quentes em latitudes mais altas, e de que houve redução nas capturas de espécies tropicais e subtropicais em áreas delimitadas pelos trópicos, os autores levantaram e analisaram a hipótese de que o aquecimento dos oceanos está afetando a pesca no mundo, e que isso pode ser um indicativo de que populações de peixes estão se movimentando em direção aos polos em resposta à elevação das temperaturas dos oceanos. Segundo os autores da pesquisa, o estudo se concentrou no atum tropical uma vez que a sua distribuição ao longo dos oceanos é fortemente condicionada à temperatura da superície do mar, o que torna a distribuição dessa espécie um bom indicador do efeito da mudança climática.

Os dados originais contém observações referentes a 19.019 tentativas individuais de capturas de peixes com um palangre entre 1967 e 2011 nos Oceanos Índico, Pacífico e Atlântico. O palangre é uma estrutura constituída por uma linha principal, forte e comprida, de onde partem outras linhas secundárias mais curtas, em grande número e em intervalos regulares, com um anzol ao final de cada uma delas (WIKIPEDIA, 2025).

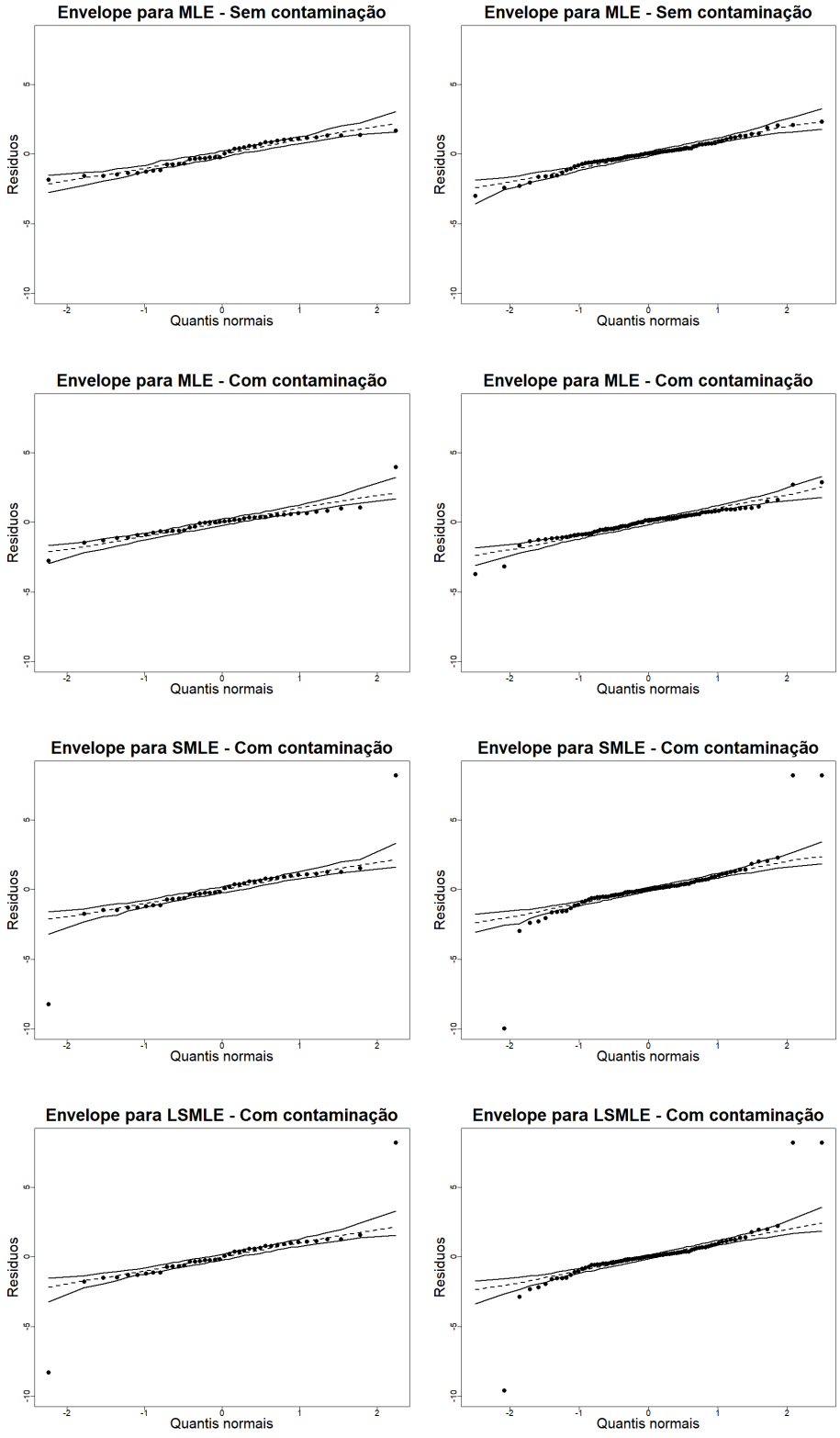

Figura 11: Gráficos de probabilidade normal e envelope simulado dos resíduos dos modelos ajustados para as amostras de tamanhos 40 (coluna à esquerda) e 80 (coluna à direita).

Entretanto, aqui está sendo considerado um subconjunto dos dados disponíveis preparado e já analisado por Ribeiro e Ferrari (2023), contendo 77 observações referentes a pescas

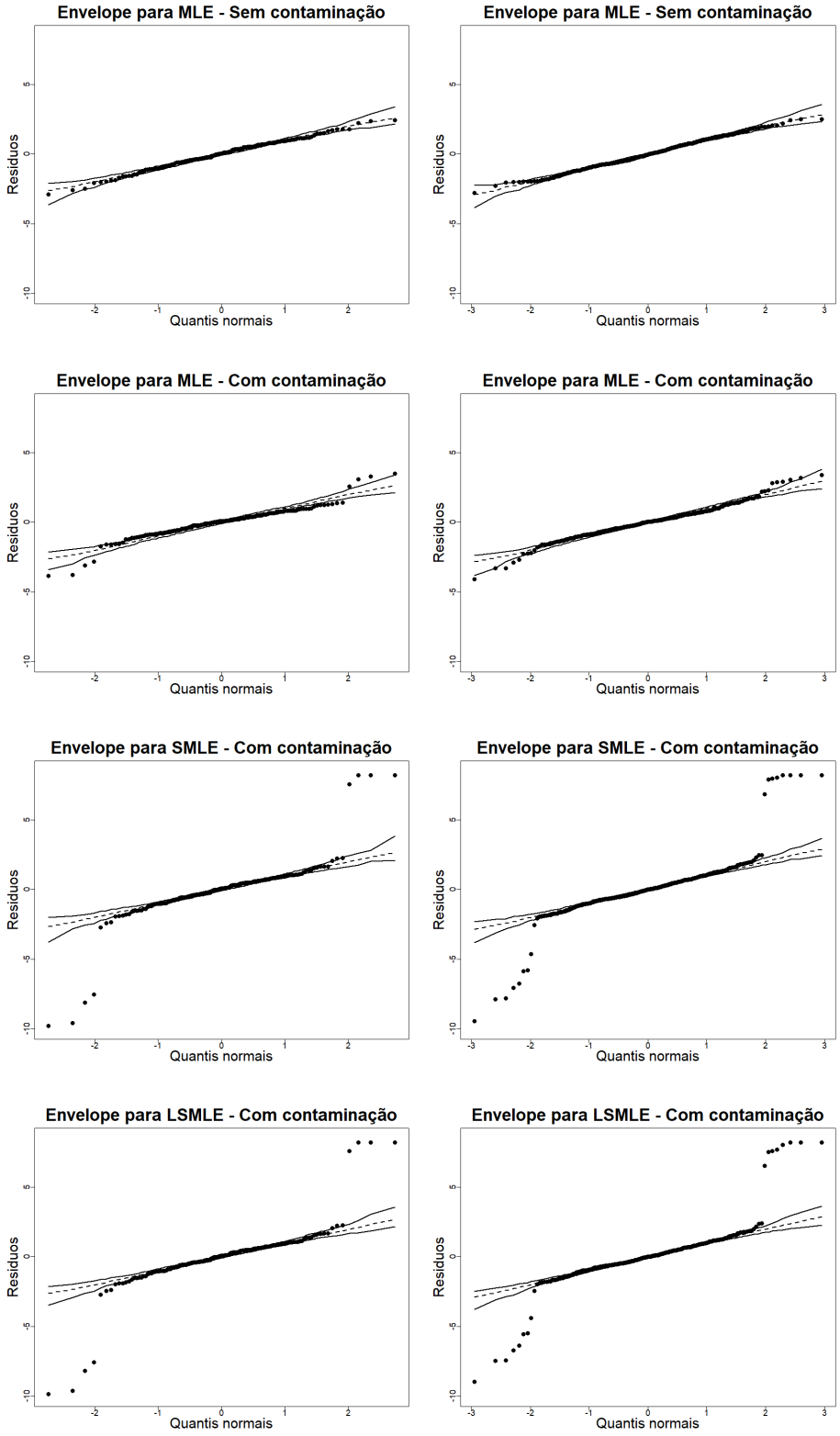

Figura 12: Gráficos de probabilidade normal e envelope simulado dos resíduos dos modelos ajustados para as amostras de tamanhos 160 (coluna à esquerda) e 320 (coluna à direita).

efetuadas em diversos pontos do oceano índico no ano de 2000. A variável resposta é a porcentagem de atum tropical (*tropical tuna percentage*; TTP) e a variável explicativa

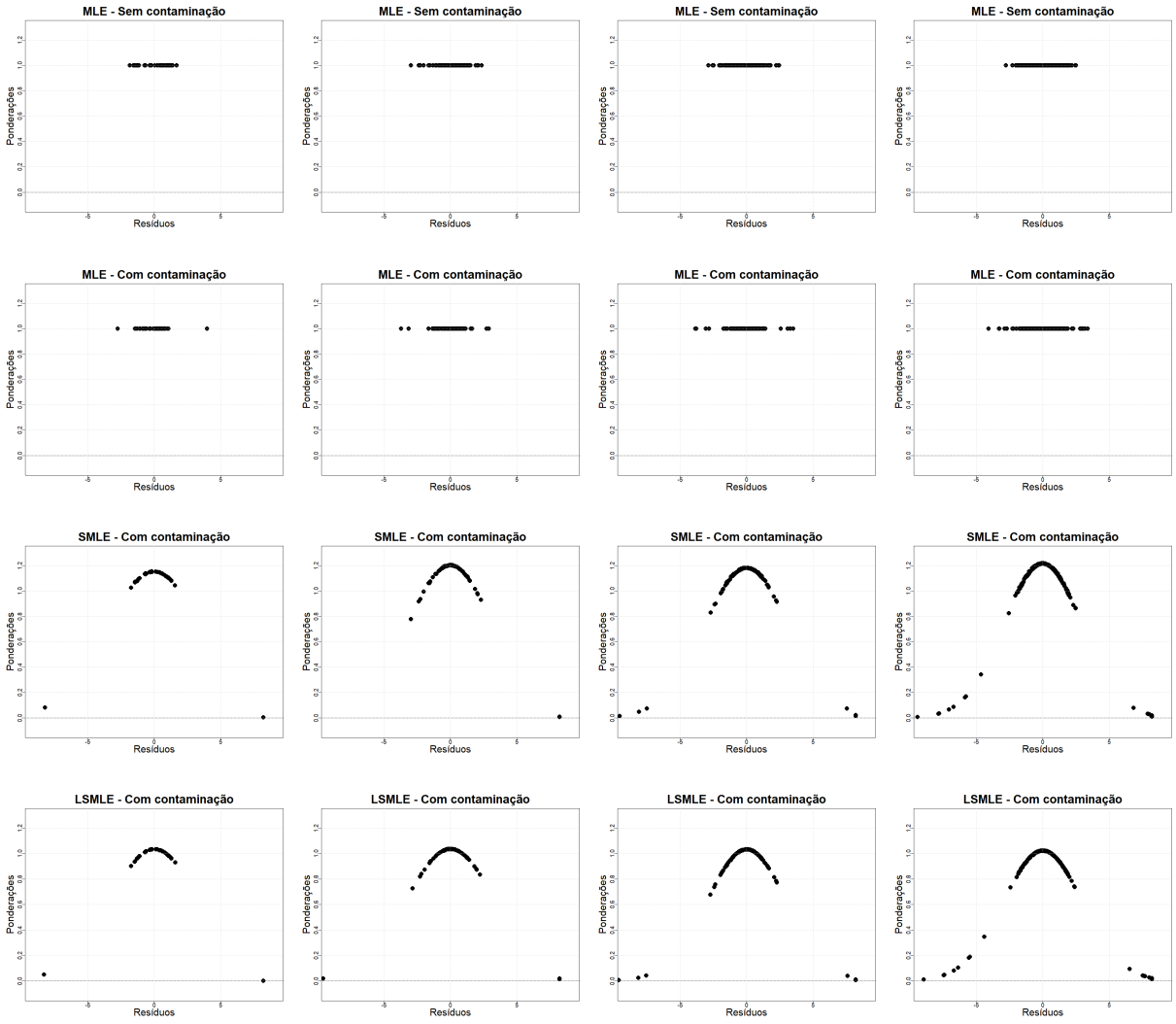

Figura 13: Gráficos das ponderações estimadas correspondentes aos ajustes efetuados para as amostras de tamanhos 40 (primeira coluna), 80 (segunda coluna), 160 (terceira coluna) e 320 (quarta coluna).

utilizada é a temperatura da superfície do mar (*sea surface temperature*; SST). Uma das observações da TTP no subconjunto de dados é igual a 1, indicando que a totalidade dos peixes fisgados na respectiva tentativa é atum tropical. Assim, considerando que o modelo de regressão beta aqui tratado é inadequado para tratar observações cuja resposta está nos limites do intervalo unitário, então, para deixar essa observação dentro do suporte admitido para a distribuição, o valor 1 foi substituído por 0,999. Com isso, essa observação passa a ser um *outlier* em relação às demais presentes no subconjunto de dados, podendo influenciar desproporcionalmente o ajuste do modelo se utilizado um método de estimação não robusto, a exemplo do MLE.

Admitindo-se que as realizações da resposta (TTP) são variáveis aleatórias independentes tal que cada y_t, $t = 1, \ldots, 77$, tem distribuição beta na forma expressa em (2.1.8), com parâmetros μ_t e ϕ, considere o modelo de regressão beta com precisão cons-

tante dados por

$$\begin{aligned} g_\mu(\mu_t) &= \beta_1 + x_{t1}^{\beta_2}, \\ g_\phi(\phi) &= \gamma_1, \end{aligned} \tag{5.2.2}$$

em que μ_t e ϕ são a média e a precisão de TTP e x_{t1} é o valor de SST para a t-ésima observação. Além disso foram utilizadas como ligação a função logit para $g_\mu(\cdot)$ e a função logarítmica para $g_\phi(\cdot)$. Observe que a especificação do modelo é análoga à utilizada para a aplicação com dados simulados, a qual foi inspirada no trabalho de Espinheira, Santos e Cribari-Neto (2017). Foram ajustados modelos de regressão beta não lineares com a especificação em (5.2.2) sob o MLE, SMLE e LSMLE com os dados completos. Na sequência os mesmos ajustes foram efetuados no conjunto de dados após exclusão da observação citada como discrepante (46ª observação), uma vez que, por meio do ajuste sob o MLE nos dados completos, foi possível identificar que essa observação apresentou um resíduo muito superior aos demais, indicando se tratar de uma observação atípica. A Figura 14 apresenta, em ambos os gráficos, o diagrama de dispersão entre a covariável SST e a resposta TTP, juntamente com as curvas obtidas nos dados completos (à esquerda) e nos dados sem a observação 46 (à direita). Nos dados completos sob SMLE e o LSMLE, foi selecionado o valor de 0,96 para a constante q, e nos dados sem o *outlier* o procedimento de seleção retornou $q = 1$ em ambos os casos, resultando no próprio MLE. No gráfico à esquerda verifica-se que enquanto os ajustes sob os estimadores robustos, cujas curvas estão quase indistinguíveis entre si, parecem se ajustar melhor à maior parte dos dados, o ajuste sob o MLE aparenta estar deslocado em direção à observação 46, identificada na imagem. Os ajustes efetuados considerando os dados reduzidos, incluindo o do MLE, apresentam posicionamentos quase idênticos aos dos ajustes do SMLE e LSMLE nos dados completos. Assim, percebe-se que, de fato, a observação 46 está afetando desproporcionalmente o ajuste do modelo sob o MLE e que, aparentemente, os modelos sob os estimadores robustos foram pouco afetados por ela.

Na Tabela 4 constam os valores obtidos para as estimativas dos parâmetros, erros padrão, estatísticas tipo-Wald e respectivos p-valores obtidos via *bootstrap* para os modelos sob o MLE, SMLE e LSMLE para os dados completos, além do modelo sob o MLE para os dados em a observação 46. Os valores das estimativas dos coeficientes de regressão e dos erros padrão *bootstrap* obtidos para os dois estimadores robustos nos dados completos ficaram muito próximos entre si e também em relação aos valores do ajuste do MLE nos dados reduzidos. Isso confirma que os estimadores robustos cumpriram muito bem o seu papel de melhor se ajustarem às observações não discrepantes e de atribuir pouco peso àquelas atípicas. Quanto aos p-valores, os testes tipo-Wald indicaram, em todos os casos, significância do parâmetro relacionado à covariável no submodelo da média.

A Figura 15 mostra os gráficos de probabilidade normal dos resíduos dos mo-

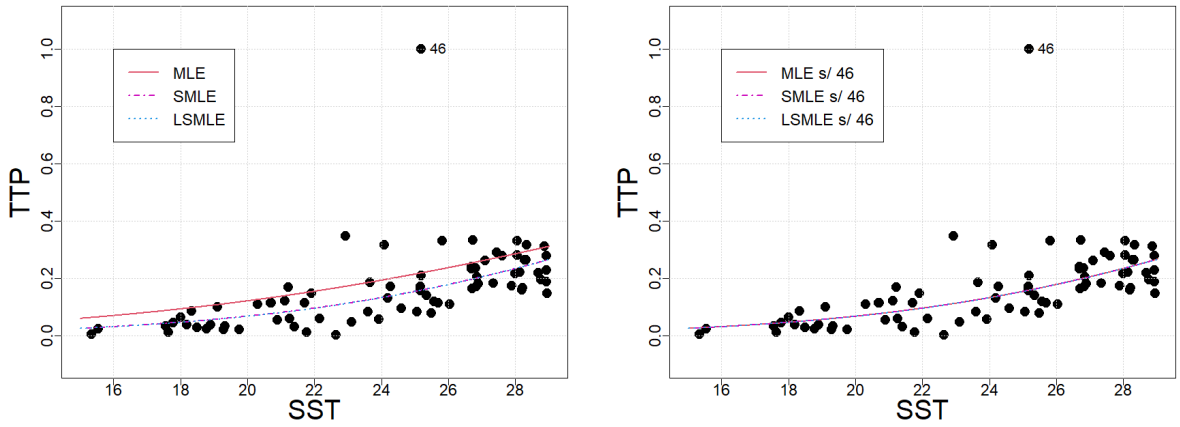

Figura 14: Gráficos de dispersão entre a resposta TTP e a covariável SST juntamente com as curvas ajustadas com os modelos com precisão constante sob os três estimadores para os dados completos (à esquerda) e os dados após exclusão da observação discrepante (à direita).

Tabela 4: Estimativas, erros-padrão *bootstrap*, estatísticas w e valores-p *bootstrap* para os modelos ajustados nos dados completos e nos dados sem a observação 46.

	MLE - Dados completos				MLE - Dados sem a obs. 46			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−7,185	0,861	69,634	-	−8,888	0,527	283,927	-
SST	0,551	0,046	142,193	$< 0{,}002$	0,613	0,022	802,236	$< 0{,}002$
submodelo da precisão								
Intercepto	1,725	0,164	110,429	-	3,304	0,157	440,795	-

	SMLE - Dados completos				LSMLE - Dados completos			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−8,861	0,488	329,475	-	−8,871	0,545	264,732	-
SST	0,612	0,021	877,787	$< 0{,}002$	0,612	0,023	710,366	$< 0{,}002$
submodelo da precisão								
Intercepto	3,318	0,148	500,577	-	3,327	0,155	460,275	-

delos com envelope simulado a uma nível de 90% de confiança para os dados completos (MLE, SMLE e LSMLE), e para os dados após exclusão da observação 46 (somente MLE). Inicialmente verifica-se que o ajuste do MLE para os dados completos se mostrou inadequado, uma vez que os resíduos estão, em sua maioria, fora dos limites das bandas do envelope. Diferentemente disso, nos dados reduzidos, apesar de alguns poucos desvios que não comprometem a conclusão sobre a adequabilidade do modelo, o ajuste sob o mesmo MLE passa a apresentar um comportamento mais próximo do esperado para um bom ajuste, com poucos pontos extrapolando o envelope. Quanto aos modelos nos quais foram usados métodos robustos de estimação, percebe-se também uma boa adequação dos ajustes. Nesses casos, observa-se que o resíduo correspondente à observação 46 permaneceu consideravelmente mais alto que os demais, entretanto, os modelos se ajustaram melhor à maioria dos dados, atribuindo pouca importância à contribuição dessa observação em específico para o processo de estimação.

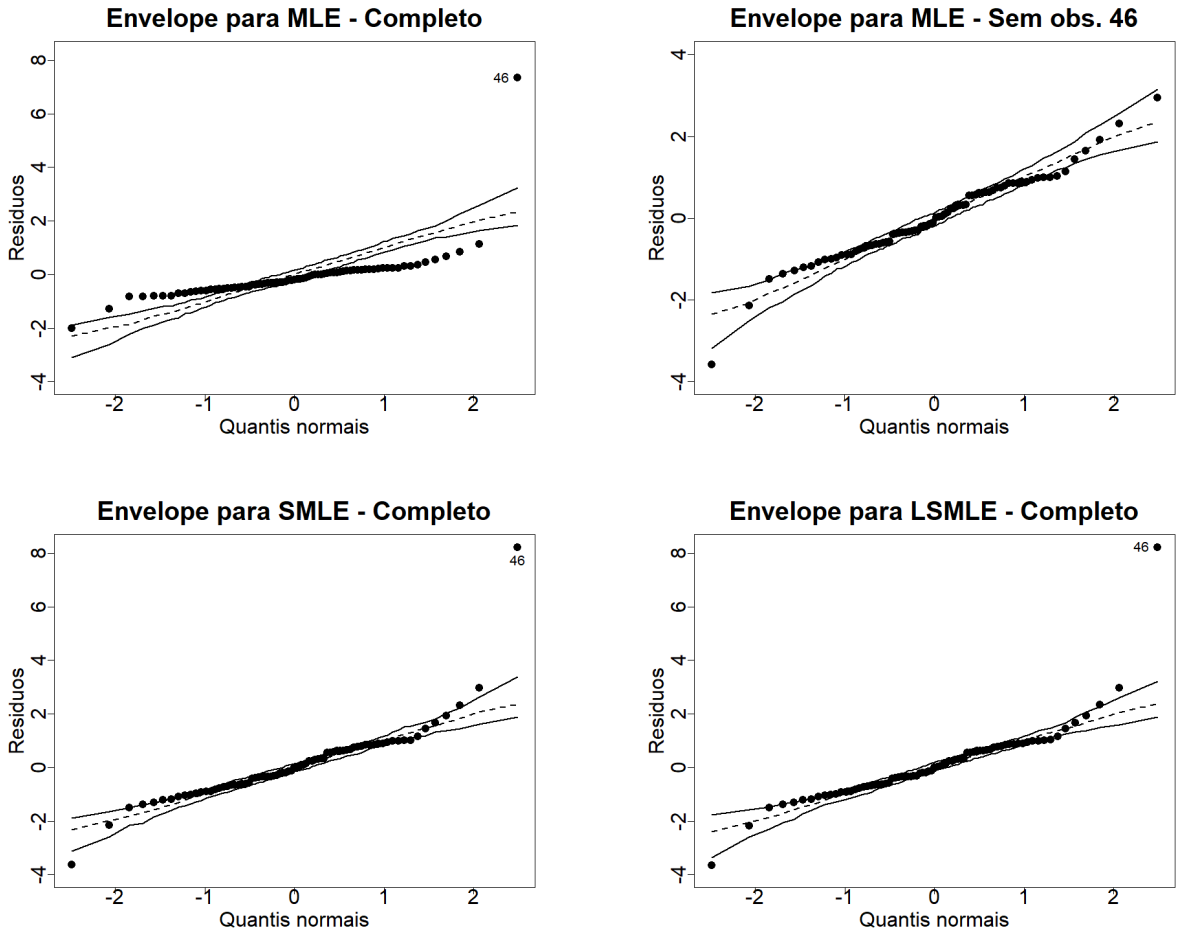

Figura 15: Gráficos de probabilidade normal e envelope simulado dos resíduos dos modelos com precisão constante ajustados para os dados completos e para os dados após a exclusão da observação atípica.

Na Figura 16 são apresentados os gráficos de dispersão das ponderações estimadas versus os resíduos dos ajustes sob o MLE para os dados completos e reduzidos, e sob o SMLE e LSMLE somente para os dados completos. Nota-se, no caso do MLE para os dados completos, a evidente discrepância entre o resíduo da observação 46 e os demais. Além disso, constata-se que esse método não robusto considera igualmente a contribuição de todas as observações, atribuindo o peso igual a 1 no processo de obtenção das estimativas. Nos casos do SMLE e LSMLE, os pesos atribuídos são diferentes para cada observação e ficaram muito próximos de zero para a observação atípica e mais próximos de 1 para as demais observações. Portanto, isso corrobora com as análises anteriores no que se refere à conclusão de que os estimadores robustos produziram bons ajustes, a despeito da observação atípica que influenciou fortemente o ajuste sob o estimador não robusto.

Também é importante avaliar se essa influência causada pela observação atípica é melhor tratada quando utilizamos uma estrutura de regressão com covariável para modelar também a precisão. Portanto, agora será considerado o modelo de regressão beta não

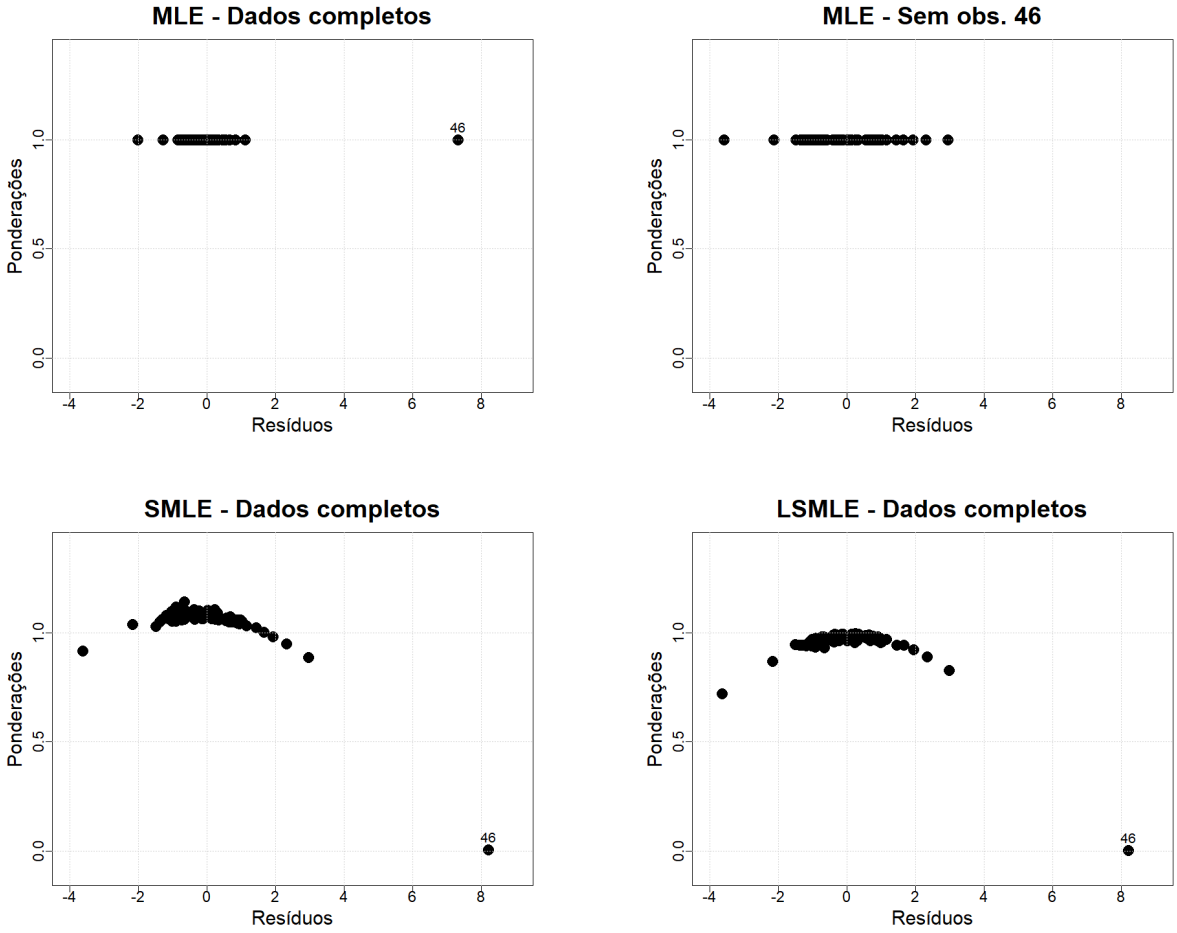

Figura 16: Gráficos das ponderações estimadas versus resíduos correspondentes aos modelos com precisão constante ajustados para os dados completos e para os dados após a exclusão da observação atípica.

linear com precisão variável definido por

$$\begin{aligned} g_\mu(\mu_t) &= \beta_1 + x_{t1}^{\beta_2}, \\ g_\phi(\phi) &= \gamma_1 + \gamma_2 x_{t1}, \end{aligned} \tag{5.2.3}$$

mantendo as mesmas especificações do modelo em (5.2.2), exceto quanto ao submodelo da precisão, no qual foi incluída uma estrutura linear de regressão associada à precisão ϕ que também utiliza como covariável a SST.

O modelo em (5.2.3) foi aplicado aos dados completos e aos dados sem a 46ª observação, contemplando o MLE, o SMLE e o LSMLE. A Figura 17 exibe o gráfico de dispersão entre a SST e a TTP conjuntamente com as curvas geradas a partir dos ajustes nos dados completos e reduzidos. O resultado é muito semelhante ao observado para o modelo com precisão constante, onde, para os dados completos, a curva referente ao modelo sob o MLE aparenta ter sido fortemente influenciada pela observação discrepante, enquanto os demais ajustes sob os estimadores robustos parecem mais adequados para a

maioria dos dados. Ao retirar esse *outlier* do conjunto de dados (gráfico à direita), a curva do MLE fica muito próxima das curvas do SMLE e LSMLE para os dados completos. Em contrapartida, as curvas para os estimadores robustos nos dados sem a observação 46 são equivalentes à do MLE, uma vez que o processo de estimação retornou $q = 1$, resultando no próprio MLE. A constante q selecionada para o modelo sob o LSMLE foi de 0,96, ou seja, a mesma selecionada para o modelo de precisão constante, enquanto que para o SMLE, o q ótimo foi de 0,76. Ressalta-se que o valor baixo para a constante de afinação no modelo sob o SMLE pode significar instabilidade das estimativas, o que indica uma provável inadequação do modelo especificado sob esse estimador.

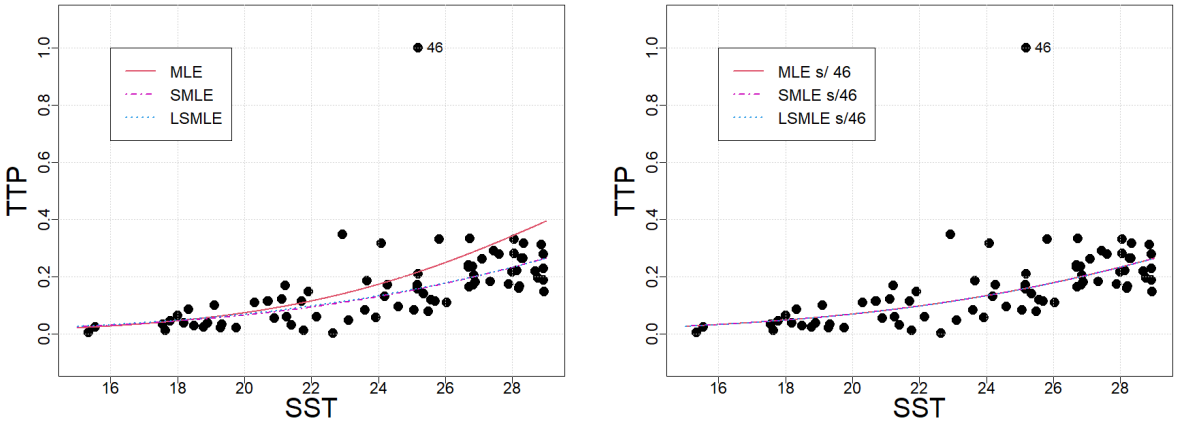

Figura 17: Gráficos de dispersão entre a resposta TTP e a covariável SST juntamente com as curvas ajustadas com os modelos com precisão variável sob os três estimadores para os dados completos (à esquerda) e os dados após exclusão da observação discrepante (à direita).

Na Tabela 5 são elencadas as estimativas dos parâmetros, erros padrão, estatísticas tipo-Wald e respectivos p-valores *bootstrap* para os modelos com precisão variável sob o MLE, SMLE e LSMLE para os dados completos, e também para o mesmo modelo sob o MLE para os dados reduzidos. Para esse caso também se observa uma influência desproporcional da observação atípica no modelo sob o MLE, resultando em estimativas distantes das obtidas sob os estimadores robustos, e tambem em relação às verificadas sob o MLE nos dados reduzidos, causando mudanças relevantes na conclusão inferencial. Em relação aos testes de hipóteses, os p-valores calculados indicaram significância dos parâmetros associados à covariável no submodelo da média em todos os casos. Para os submodelos da precisão, à exceção do MLE nos dados completos, todos os demais modelos apresentaram p-valores altos que conduziram à não rejeição da hipótese nula, de não significância dos respectivos parâmetros. Comparando esses resultados com os dos modelos com precisão contante mostrados na Tabela 4, observa-se que as estimativas referentes aos submodelos da média ficaram bastante próximas nos casos dos estimadores robustos e do MLE nos dados reduzidos, e muito diferentes no caso do MLE para os dados completos,

indicando que a observação de número 46, apesar de ter influência desproporcional em ambos os casos, afetou de forma diferente os ajustes com e sem precisão variável.

Tabela 5: Estimativas, erros-padrão *bootstrap*, estatísticas w e valores-p *bootstrap* para os modelos com precisão variável ajustados nos dados completos e nos dados sem a observação 46.

	MLE - Dados completos				MLE - Dados sem a obs. 46			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−9,945	0,668	221,906	-	−8,775	0,554	251,039	-
SST	0,669	0,026	652,913	$< 0{,}002$	0,608	0,024	626,329	$< 0{,}002$
submodelo da precisão								
Intercepto	19,909	3,247	37,601	-	1,667	3,448	0,234	-
SST	−5,671	1,018	31,020	$< 0{,}002$	0,516	1,085	0,227	0.620

	SMLE - Dados completos				LSMLE - Dados completos			
	Estimativa	SE	estat w	valor-p	Estimativa	SE	estat w	valor-p
submodelo da média								
Intercepto	−8,962	0,553	262,510	-	−8,773	0,593	219,228	-
SST	0,615	0,023	721,825	$< 0{,}002$	0,608	0,025	594,706	$< 0{,}002$
submodelo da precisão								
Intercepto	4,924	3,699	1,772	-	1,865	3,253	0,329	-
SST	−0,456	1,173	0,151	0,632	0,461	1,026	0,202	0,660

A Figura 18 ilustra o nível de qualidade do ajuste dos modelos por meio dos gráficos de probabilidade normal dos seus resíduos juntamente com envelope simulado a um nível de 90% de confiança, para os dados completos (MLE, SMLE e LSMLE), e para os dados após exclusão da observação 46 (somente MLE). Os resultados são muito parecidos com os observados para os modelos com precisão constante, onde se verifica uma inadequação muito evidente do ajuste sob o MLE para os dados completos e ajustes mais adequados dos estimadores robustos para esses mesmos dados, exceto em relação à observação discrepante, cujo resíduo ficou muito acima dos demais, o que é esperado. Além disso, conclui-se que o ajuste para o modelo com MLE nos dados reduzidos também produziu um resultado que pode-se considerar adequado.

Na Figura 19 são mostrados os gráficos de dispersão das ponderações estimadas versus os resíduos dos ajustes sob o MLE para os dados completos e reduzidos, e sob o SMLE e LSMLE somente para os dados completos. Inicialmente percebe-se uma diferença significativa no gráfico referente ao SMLE quando comparado a este mesmo gráfico gerado para o modelo com precisão constante mostrado na Figura 16. Nesse caso do modelo com precisão variável sob o SMLE, nota-se que foram atribuídos pesos relativamente baixos a diversas outras observações que não foram apontadas como discrepantes nos demais ajustes. Isso corrobora a percepção de inadequação do modelo sob o MLE com a especificação em (5.2.3) para os dados completos utilizados. Nos demais casos, os resíduos apresentaram comportamento análogo ao observado para o modelo com precisão constante. Portanto, para o modelo sob o LSMLE os pesos atribuídos foram próximos a 1 e diferentes para cada observação, exceto quanto à observação 46, que teve atribuição de peso próximo a zero, e pesos constantes e iguais a 1 para o MLE.

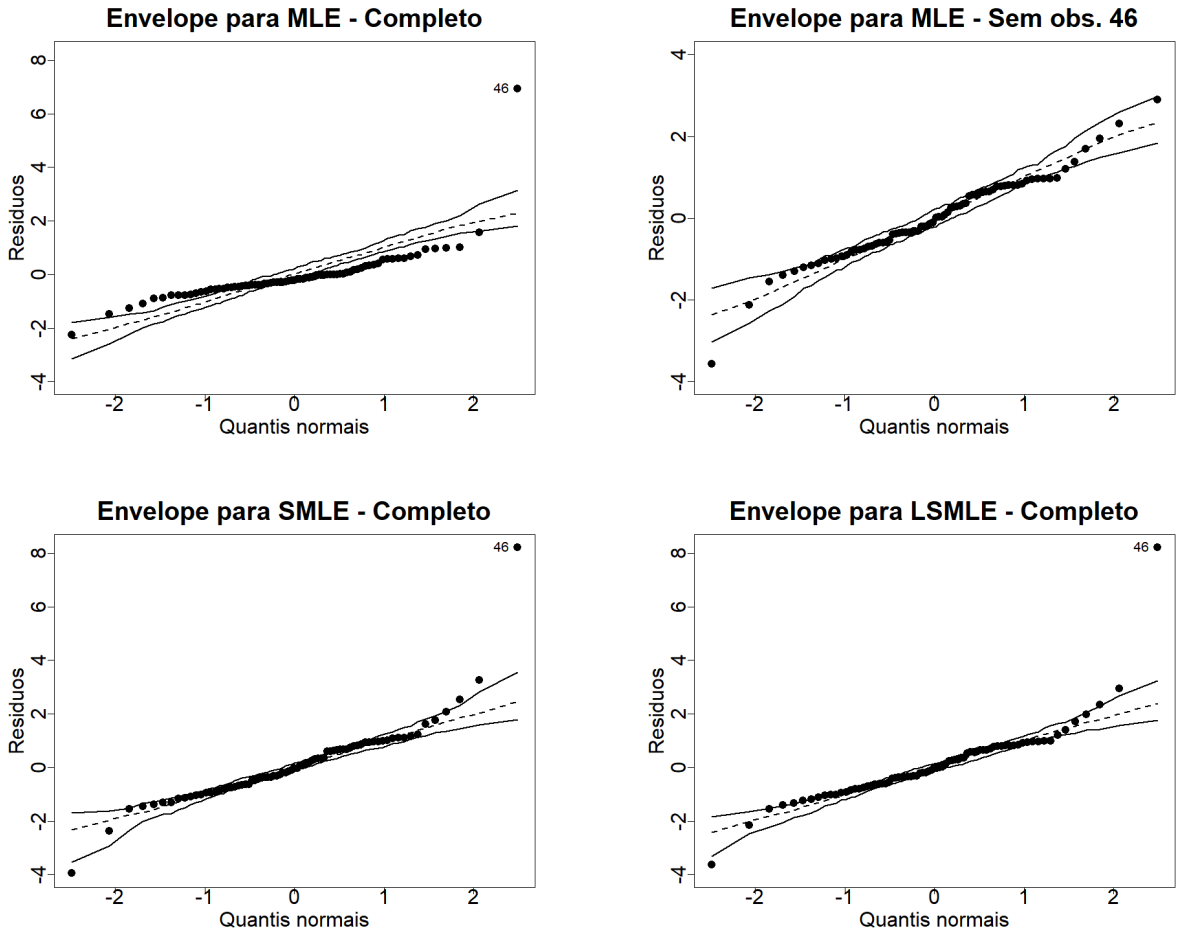

Figura 18: Gráficos de probabilidade normal e envelope simulado dos resíduos dos modelos com precisão variável ajustados para os dados completos e para os dados após a exclusão da observação atípica.

Finalizada a análise, pode-se concluir que, apesar do ajuste sob o MLE com os dados completos indicar a necessidade de modelar também a precisão, verificou-se que a modelagem da precisão com a covariável SST não melhorou o ajuste do MLE e nem os ajustes com o SMLE e LSMLE. Portanto, dentre os modelos experimentados nesta aplicação, a regressão beta não linear com precisão constante especificada em (5.2.2) utilizando os estimadores robustos aqui estudados conduziram a ajustes robustos e mais adequados aos dados utilizados.

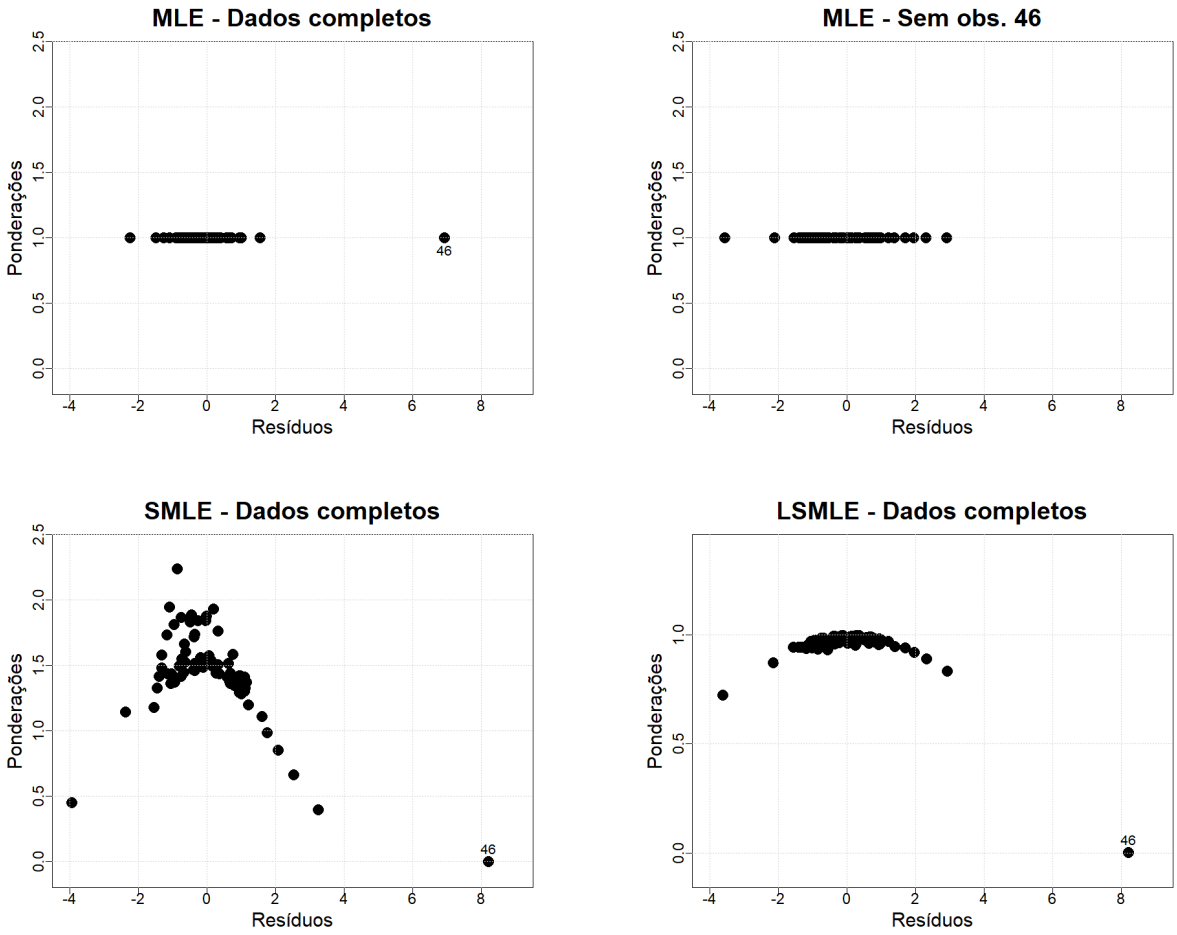

Figura 19: Gráficos das ponderações estimadas versus resíduos correspondentes aos modelos com precisão variável ajustados para os dados completos e para os dados após a exclusão da observação atípica.

6 Considerações finais

Neste trabalho, foi efetuada uma revisão de literatura contemplando alguns dos principais métodos de estimação robustos desenvolvidos recentemente para modelos de regressão beta lineares, com foco nos estimadores SMLE e LSMLE, que são casos particulares do procedimento geral de M-estimação. Este processo é conhecido por produzir estimadores robustos, além de garantir que estes possuam boas propriedades, a exemplo da normalidade assintótica.

A partir da revisão de literatura, foi proposta a regressão beta não linear robusta como uma generalização dos métodos de estimação robustos estudados, mais precisamente o SMLE e o LSMLE, para modelos de regressão beta nos quais a estrutura de regressão é flexibilizada para contemplar formas não lineares. Ao longo do texto foram apresentadas propriedades e também expressões de algumas medidas robustez para essa classe não linear de modelos. Por meio das propriedades teóricas apresentadas e dos resultados observados nas aplicações e estudos de simulação, foi evidenciado que, sob a regressão beta não linear, o procedimento de estimação por máxima verossimilhança é sensível a observações atípicas na variável resposta, podendo ser demasiadamente influenciado por estas e conduzir a conclusões inferenciais errôneas sobre os dados de interesse.

Os métodos robustos citados dependem de uma constante de afinação que tem a importante função de controlar o balanceamento entre robustez e eficiência assintótica do estimador. Neste trabalho foi proposta uma adaptação ao processo de seleção da contante de afinação desenvolvido por Ribeiro e Ferrari (2023). Ao longo dos estudos foi identificada uma instabilidade do processo original em situações onde eram utilizados modelos de regressão beta robustos não lineares cujas estimativas para o erro padrão dos estimadores eram obtidas via *bootstrap*. Além disso, a adaptação deixou o processo computacionalmente mais eficiente para esses casos. Referido algoritmo para seleção da constante foi utilizado em todos as simulações e aplicações efetuadas neste trabalho, obtendo, em todos os casos, bons resultados. Tal conclusão é evidenciada pela escolha de valores que se mostraram adequados para a constante, inclusive quando os modelos foram ajustados a amostras e dados sem a presença de contaminação, situação na qual foi retornado o valor de 1 para a constante, o que resulta no próprio MLE.

Foi mostrado por meio de estudos de simulação de Monte Carlo e de aplicações a dados simulados que a utilização do método de máxima verossimilhança para os modelos de regressão beta não lineares robustos em dados sob contaminação conduziram a estimativas enviesadas e, consequentemente, a conclusões inferenciais incorretas sobre os dados. Nesse mesmo sentido, verificou-se que a utilização dos métodos de estimação robustos para estes mesmos dados resultaram em estimativas muito próximas dos verdadeiros valores dos

parâmetros. Por outro lado, os ajustes efetuados em dados não contaminados resultaram, para todos os casos, em igualdade entre os estimadores robustos e o não robusto, o que reforça o bom funcionamento do algoritmo de seleção da contante de afinação. Também foi mostrado que a falta de robustez do procedimento sob máxima verossimilhança, diferentemente do que foi observado para os estimadores robustos, conduziu à conclusão de não significância dos parâmetros associados às covariáveis, especialmente em amostras menores.

Nas aplicações com dados reais foi mostrado que os modelos de regressão beta não lineares robustos também se ajustaram melhor aos dados quando comparados ao modelo sob o MLE. Também foi evidenciado que uma única observação discrepante na variável resposta foi suficiente para causar distorções consideráveis nos valores das estimativas do modelo sob o MLE, e que ao retirar essa observação atípica os modelos sob o SMLE e LSMLE se igualam ao modelo sob o MLE.

Adicionalmente, cabe salientar que os métodos de estimação robustos discutidos nesta dissertação são particularmente recomendados para cenários em que se deseja reduzir a influência de observações atípicas sobre os resultados inferenciais. Entretanto, em situações nas quais o interesse reside precisamente na identificação de ocorrências raras ou incomuns, tais métodos não são apropriados, uma vez que sua natureza é atenuar o impacto dessas observações em vez de evidenciá-las.

Por fim, é importante ressaltar que os resultados apresentados nesta pesquisa possuem algumas limitações, sendo a maioria delas decorrentes da restrição de tempo para o desenvolvimento da dissertação. A primeira delas se refere à utilização de estimativas para os erros padrão obtidas somente por meio de *bootstrap*, não sendo calculadas e utilizadas essas estimativas a partir da distribuição assintótica dos estimadores robustos sob os modelos não lineares. Não obstante os bons resultados alcançados com o erro padrão *bootstrap*, a utilização dessa medida calculada a partir do método analítico seria importante como complemento para a pesquisa. Outra limitação está relacionada ao não aprofundamento da avaliação do desempenho do teste tipo-Wald. Como o erro padrão de cada réplica ou conjunto de dados foi obtido por meio de *bootstrap*, ficou inviabilizada a realização de simulações de Monte Carlo para quantificar os níveis empíricos do tamanho e poder do teste, uma vez que levaria um longo tempo para a realização de todo o processamento computacional necessário. Por fim, as simulações e aplicações efetuadas utilizaram modelos com um número limitados de formas para o preditor não linear. Desse modo, os resultados aqui discutidos podem variar a depender da forma não linear utilizada nas estruturas de regressão.

A partir das citadas limitações e considerando outros achados observados durante o desenvolvimento deste trabalho, listamos alguns pontos que podem ser melhor explorados em trabalhos futuros:

1. Estudar de forma mais aprofundada o desempenho do teste de hipóteses tipo-Wald avaliando os níveis empíricos do tamanho e poder dos testes sob a hipótese nula H_0 ao se obter o p-valor via *bootstrap*.

2. Desenvolver e obter as expressões analíticas referentes aos erros padrão assintóticos para o SMLE e LSMLE sob a regressão beta não linear robusta conforme (4.1.8) e (4.2.4), respectivamente, e utilizar esses valores em estudos de simulação e aplicações, a exemplo dos aqui efetuados. Além disso, efetuar uma comparação dos desempenhos do método original de seleção da contante de afinação com a adaptação proposta neste trabalho.

3. Estudar o desempenho dos modelos de regressão beta não lineares robustos contemplando formas diferentes das utilizadas neste trabalho para as as estruturas de regressão, além de outras opções para as funções de ligação associadas aos submodelos da média e da precisão.

4. Avaliar o desempenho dos modelos de regressão beta não lineares robustos em relação aos resultados obtidos por modelos de regressão beta lineares robustos, para identificar situações e cenários onde a forma não linear melhor se adequa.

5. Avaliar o desempenho dos modelos de regressão beta não lineares robustos utilizando outros métodos inferenciais para obtenção das estimativas dos parâmetros, a exemplo do MDPDE e LMDPDE.

6. Desenvolver os modelos de regressão beta inflacionados não lineares robustos, incluindo o processo de estimação robusta para os submodelos correspondentes aos componentes discretos das distribuições beta inflacionadas (OSPINA; FERRARI, 2010).

Referências

ATKINSON, A. C. *Plots, Transformations and Regression; an Introduction to Graphical Methods of Diagnostic Regression Analysis*. [S.l.]: New York: Oxford University, 1985.

BASU, A.; HARRIS, I. R.; HJORT, N. L.; JONES, M. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, Oxford University Press, v. 85, n. 3, p. 549–559, 1998.

BAYES, C. L.; BAZÁN, J. L.; GARCIA, C. A new robust regression model for proportions. *Bayesian Analysis*, v. 7, n. 4, p. 841–866, 2012.

BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society B*, n. 26, p. 211–252, 1964.

BRISCO, A. M. D.; MIGLIORATI, S.; ONGARO, A. Robustness against outliers: A new variance inflated regression model for proportions. *Statistical Modelling*, SAGE Publications Sage India: New Delhi, India, v. 20, n. 3, p. 274–309, 2020.

CASELLA, G.; BERGER, R. L. *Inferência Estatística*. [S.l.]: CENGAGE Learning, 2011. 95–96 p.

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.

EFRON, B. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, p. 7:1–25, 1979.

EFRON, B. Bootstrap methods: another look at the jackknife. In: *Breakthroughs in Statistics: Methodology and Distribution*. [S.l.]: Springer, 1992. p. 569–593.

EFRON, B.; TIBSHIRANI, R. J. *An Introduction to the Bootstrap*. [S.l.]: Chapman and Hall/CRC, 1994.

ESPINHEIRA, P. L.; FERRARI, S. L.; CRIBARI-NETO, F. Influence diagnostics in beta regression. *Computational Statistics & Data Analysis*, Elsevier, v. 52, n. 9, p. 4417–4431, 2008.

ESPINHEIRA, P. L.; SANTOS, E. G.; CRIBARI-NETO, F. On nonlinear beta regression residuals. *Biometrical Journal*, v. 59, n. 3, p. 445–461, 2017.

FERRARI, D.; La Vecchia, D. On robust estimation via pseudo-additive information. *Biometrika*, Oxford University Press, v. 99, n. 1, p. 238–244, 2012.

FERRARI, D.; YANG, Y. Maximum lq-likelihood estimation. *The Annals of Statistics*, v. 2, n. 38, p. 753–783, 2010.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, Taylor & Francis, v. 31, n. 7, p. 799–815, 2004.

GHOSH, A. Robust inference under the beta regression model with application to health care studies. *Statistical Methods in Medical Research*, SAGE Publications Sage UK: London, England, v. 28, n. 3, p. 871–888, 2019.

GHOSH, A.; BASU, A. Robust estimation in generalized linear models: the density power divergence approach. *Test*, Springer, v. 25, p. 269–290, 2016.

GÓMEZ-DÉNIZ, E.; SORDO, M. A.; CALDERÍN-OJEDA, E. The log–lindley distribution as an alternative to the beta regression model with applications in insurance. *Insurance: mathematics and Economics*, v. 54, p. 49–57, 2014.

GUMBEL, E. Statistical theory of extreme values and some practical applications: a series of lectures. *NBS Applied Mathematics Series*, US Department of Commerce, v. 33, 1954.

HAMPEL, F. R.; M, R. E.; ROUSSEEUW, P. J.; STAHEL, W. A. Robust statistics: The approach based on influence functions. John Wiley and Sons, v. 196, 2011.

HERITIER, S.; RONCHETTI, E. Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*, Taylor & Francis, v. 89, n. 427, p. 897–904, 1994.

HUBER, P. J. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, JSTOR,, p. 73–101, 1964.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions, volume 2*. [S.l.]: John Wiley & Sons, 1995. v. 289.

KALLIANPUR, G.; RAO, C. R. On fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, JSTOR, v. 15, n. 4, p. 331–342, 1955.

KIESCHNICK, R.; MCCULLOUGH, B. D. Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. *Statistical modelling*, Sage Publications Sage CA: Thousand Oaks, CA, v. 3, n. 3, p. 193–213, 2003.

LA VECCHIA, D.; CAMPONOVO, L.; FERRARI, D. Robust heart rate variability analysis by generalized entropy minimization. *Computational Statistics & Data Analysis*, Elsevier, v. 82, p. 137–151, 2015.

LEMONTE, A. J.; BAZÁN, J. L. New class of johnson distributions and its associated regression model for rates and proportions. *Biometrical Journal*, v. 58, n. 4, p. 727–746, 2016.

LIMA, F. P. *Inferência bootstrap em modelos de regressão beta*. Tese (Doutorado) — Universidade Ferederal de Pernambuco (UFPE), 2017.

MALUF, Y. S.; FERRARI, S. L. P.; QUEIROZ, F. F. Robust beta regression through the logit transformation. *Metrika*, Springer, v. 88, n. 1, p. 61–81, 2025.

MARONNA, R. A.; MARTIN, R. D.; YOHAI, V. J.; SALIBIÁN-BARRERA, M. *Robust statistics: theory and methods (with R)*. [S.l.]: John Wiley & Sons, 2019.

MCDONALD, J. B.; XU, Y. J. A generalization of the beta distribution with applications. *Journal of Econometrics*, Elsevier, v. 66, n. 1-2, p. 133–152, 1995.

MIGLIORATI, S.; BRISCO, A. M. D.; ONGARO, A. A new regression model for bounded responses. *Bayesian Analysis*, Carnegie Mellon University, v. 13, n. 3, p. 845–872, 2018.

MONLLOR-HURTADO, A.; PENNINO, M. G.; SANCHEZ-LIZASO, J. L. Shift in tuna catches due to ocean warming. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 6, p. e0178196, 2017.

NELDER, J. A.; LEE, Y. Generalized linear models for the analysis of taguchi-type experiments. *Applied stochastic models and data analysis*, v. 7, n. 1, p. 107–120, 1991.

NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, v. 135, n. 3, p. 370–384, 1972.

OSPINA, R. *Estimação Pontual e Intervalar em um Modelo de Regressão Beta*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2004.

OSPINA, R.; FERRARI, S. L. P. Inflated beta distributions. *Statistical Papers*, Springer, v. 51, n. 1, p. 111, 2010.

OSPINA, R.; FERRARI, S. L. P. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, Elsevier, v. 56, n. 6, p. 1609–1623, 2012.

PEREIRA, T. L. *Regressão Beta Inflacionada: Inferência e Aplicações*. Tese (Doutorado) — Universidade Federal de Pernambuco (UFPE), 2010.

PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D.; HEISTERKAMP, S.; WILLIGEN, B. V.; MAINTAINER, R. Package 'nlme'. *Linear and Nonlinear Mixed Effects Models*, v. 3, n. 1, p. 274, 2017.

PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T.; FLANNERY, B. P. *Numerical Recipes in C*. [S.l.]: Cambridge university press New York, NY, 1992.

QUEIROZ, F.; MALUF, Y. *robustbetareg: Robust Beta Regression*. [S.l.], 2022. R package version 0.3.0. Disponível em: ⟨https://CRAN.R-project.org/package=robustbetareg⟩.

QUEIROZ, F. F.; FERRARI, S. L. P. Power logit regression for modeling bounded data. *Statistical Modelling*, v. 24, n. 5, p. 395–421, 2024.

QUEIROZ, F. F. d. *Análise de Dados com Suporte Limitado: Modelos Power Logit e Contribuições à Inferência Robusta*. Tese (Doutorado) — Universidade de São Paulo, 2022.

R Core Team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Austria: Vienna*. 2024.

RIBEIRO, T. K.; FERRARI, S. L. P. Robust estimation in beta regression via maximum lq-likelihood. *Statistical Papers*, Springer Nature BV, v. 64, n. 1, p. 321–353, 2023.

RIBEIRO, T. K. d. A. *Regressão Beta Robusta*. Tese (Doutorado) — Universidade de São Paulo, 2020.

ROSS, S. *Probabilidade: Um Curso Moderno com Aplicações*. [S.l.]: Bookman Editora, 2009.

ROUSSEEUW, P. J. A new infinitesimal approach to robust estimation. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Springer, v. 56, n. 1, p. 127–132, 1981.

SCHMIT, J. T.; ROTH, K. Cost effectiveness of risk management practices. *Journal of Risk and Insurance*, JSTOR, p. 455–470, 1990.

SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, Elsevier, v. 54, n. 2, p. 348–366, 2010.

SMITHSON, M.; SHOU, Y. Cdf-quantile distributions for modelling random variables on the unit interval. *British Journal of Mathematical and Statistical Psychology*, v. 70, n. 3, p. 412–438, 2017.

SMITHSON, M.; VERKUILEN, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, American Psychological Association, v. 11, n. 1, p. 54–71, 2006.

SMYTH, G. K.; VERBYLA, A. P. Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics: The official journal of the International Environmetrics Society*, v. 10, n. 6, p. 695–709, 1999.

WIKIPEDIA. *Palangre — Wikipedia, The Free Encyclopedia*. 2025. [Online; acesso em 01-junho-2025]. Disponível em: ⟨https://pt.wikipedia.org/wiki/Palangre⟩.

WRIGHT, S.; NOCEDAL, J. Numerical optimization. *Springer Science*, v. 35, n. 67-68, p. 7, 1999.